

MG3010

Methods in Business

Research

Spring 2026

Dr. Sadia Nadeem, Mr. Muhammad Qaiser Shafi,
Dr. Usama Arshad & Mr. Sajjad Hanif
FAST School of Management, Islamabad
National University of Computer & Emerging Sciences

Course Pack

Table of Contents

| Chapter | Book |
|--|---|
| 1 The Nature & Process of Business Research | <i>Bell et al.</i> (2019; 3-16) |
| 2 Critically Reviewing the Literature | <i>Saunders et al.</i> (2023; 72-122) |
| 3 Generating a Research Idea & Developing Your Research Proposal | <i>Saunders et al.</i> (2023; 28-66) |
| 4 Formulating the Research Design | <i>Saunders et al.</i> (2023; 176-228) |
| 5 Theoretical Framework & Hypothesis Development | <i>Sekaran & Bougie</i> (2016; 71-94) |
| 6 Selecting Samples | <i>Saunders et al.</i> (2023; 288-334) |
| 7 Collecting Primary Data Using Questionnaires | <i>Saunders et al.</i> (2023; 506-563) |
| 8 Analyzing Data Quantitatively | <i>Saunders et al.</i> (2023; 572-642) |
| 9 Collecting Primary Data Using Interviews & Diaries | <i>Saunders et al.</i> (2023; 440-500) |
| 10 Analyzing Data Qualitatively | <i>Saunders et al.</i> (2023; 650-712) |
| 11 Big Data & Data Mining | <i>Sharda et al.</i> (2023) <i>Big Data</i> <i>Data Mining</i> |
| 12 Negotiating Access & Research Ethics | <i>Saunders et al.</i> (2023; 220-280) |

The Nature & Process of Research

Bell et al. (2019; 3-16)

CHAPTER 1

THE NATURE AND PROCESS OF BUSINESS RESEARCH

| | | | |
|---|---|---|----|
| Introduction | 4 | Concepts and theories | 8 |
| What is 'business research'? | 4 | Research questions | 9 |
| Why do business research? | 4 | Sampling | 11 |
| Business research methods in context | 5 | Data collection | 11 |
| Relevance to practice | 6 | Data analysis | 12 |
| The process of business research | 8 | Writing up | 12 |
| Literature review | 8 | The messiness of business research | 13 |
| | | Key points | 15 |
| | | Questions for review | 15 |

CHAPTER OUTLINE

This chapter introduces some fundamental considerations in conducting business research. It begins by outlining what we mean by business research and the reasons why we conduct it. The chapter considers three main areas:

- *Business research methods in context.* This introduces issues such as the role of theory; values and ethical considerations; debates about relevance versus rigour; and how political considerations affect business research.
- *The elements of the research process.* The whole book is dedicated to the elements of business research, but here an overview of the essential stages is given. The elements are: a **literature review**; formulating **concepts** and theories; devising **research questions**; sampling; data collection; data analysis; and writing up.
- *The messiness of business research.* This section acknowledges that business research often does not conform to a neat, linear process and that researchers may find themselves facing unexpected contingencies and difficulties. At the same time, it is suggested that a familiarity with the nature of the research process and its principles is crucial to navigating unexpected issues.

All of the issues presented here are addressed in greater detail in later chapters, but they are introduced at this stage to provide you with an early opportunity to think about them.

Introduction

This book is concerned with the ways that business researchers go about their craft. It covers the research process in all its phases—formulating research objectives; choosing research methods; securing research participants; collecting, analysing, and interpreting data; and disseminating findings. Understanding business research methods is important for several reasons, but three stand out. First, such an understanding will help you to avoid the many pitfalls that are common when relatively inexperienced people try to do business research, such as failing to match research questions to appropriate research methods, asking ambiguous questions in **questionnaires**, or engaging in practices that are ethically dubious. If you are expected to complete a research project, an education in research methods is important, not just to ensure that correct procedures are

followed but also to gain an appreciation of the choices that are available to you. Second, an understanding of business research methods is important from the point of view of consuming published research. If you are doing a degree in a business subject, you will probably read a lot of published research in the substantive areas you are studying. A good grounding in the research process and a knowledge of potential pitfalls provides an invaluable critical edge when reading about research done by others. Finally, an understanding of research methods will enable you to satisfy your curiosity about topics that interest you by undertaking your own research project, either for a dissertation or in a work-related context. Such a project may generate insight into important business issues and allow you, in a small, incremental way, to contribute to business knowledge.

What is ‘business research’?

The term ‘business research’ as used here refers to the academic study of topics related to questions relevant to business, including management and organizations. Business research is situated in the context of social science disciplines, such as sociology, psychology, anthropology, and economics. These inform the study of business and its specific fields, which include marketing, human resource management (HRM), strategy, organizational behaviour, accounting and finance, industrial relations, and operational research. Business research may be motivated by developments and changes in

organizations and societies, such as concerns about rising levels of executive pay or a desire to improve the environmental sustainability of businesses, but social scientific ideas are key to illuminate and explain those changes. The social sciences also provide ideas about how to formulate research topics and how to interpret and draw implications from research findings. In other words, what distinguishes business research as discussed in this book is that it is deeply rooted in the ideas and intellectual traditions of the social sciences.

Why do business research?

Academics conduct research because in the course of reading the literature on a topic or reflecting on what is going on in organizations, questions occur to them. They may notice a gap in the literature, or an inconsistency between a number of studies, or an unresolved issue in the literature. A societal development may also provide a point of departure for the development of a research question. For example, one of the authors of this book, Emma Bell, was interested in observing the behaviours

of consumers who prefer a particular technology brand, Apple. She therefore decided to study how consumer groups develop and express strong loyalty to a particular brand through exploring their relationship to its co-founder and former CEO, Steve Jobs, following his death in 2011 (Bell and Taylor 2016). In exploring this issue, the researchers drew on organizational studies literatures on leadership and culture, and on sociological and philosophical literatures related to cultural practices of

mourning and loss, to generate insights. As we note in Chapter 2, there is no single reason why people do business research, but, at its core, it is done when there is an

aspect of business and management that is believed to be inadequately understood.

Business research methods in context

Business research and its associated methods do not exist in a vacuum. The following factors form aspects of the context within which business research takes place.

- The *theories* that social scientists develop to understand the social world influence what is researched and how research findings are interpreted. The topics of business research are deeply influenced by the theoretical position adopted. Drawing on our earlier example, Bell and Taylor (2016) were interested in understanding the behaviours of brand-loyal consumers in mourning the death of a leader, so they took into account existing theories concerning the collective cultural processes whereby groups of people develop deep attachments to organizational leaders and brands and use this to construct a shared sense of identity. This illustrates how current research is informed and influenced by existing theory. Research also contributes to theory because new research feeds into the stock of knowledge to which the theory relates.
- Existing knowledge about an area also forms an important part of the background in which business research takes place. This means that someone planning to conduct research must be familiar with the *literature* on the area of interest. You have to be familiar with what is already known so that you can build on it and avoid covering the same ground as others. Reviewing literature is the main focus of Chapter 5 and is also discussed in other chapters, such as Chapter 7.
- The researcher's views about the nature of the *relationship between theory and research* are also important. For some researchers, theory should be addressed at the beginning of a research project. The researcher engages in theoretical reflections from which a **hypothesis** or hypotheses are formulated and then tested. This was the approach taken in the study by Elsesser and Lever (2011: 1559), discussed in Research in focus 1.3, in which they proposed nine hypotheses based on their review of relevant theory. An alternative position is to view theory as an outcome of the research process—that is, as something that is arrived at after the research has been carried out. The first approach implies that a set of theoretical ideas drive data collection and analysis, whereas the second suggests a more open-ended strategy in which theoretical ideas emerge out of data. Of course, the choice is rarely as stark as this, but there are contrasting views about the role of theory in research. This issue will be a major focus of Chapter 2.
- Assumptions and views about *how research should be conducted* influence the research process. It is often assumed that a 'scientific' approach should be followed, in which a hypothesis is formulated and then tested using precise measurement techniques. Such research exists, but the view that this is how business research should be done is not universally accepted. These are **epistemological** considerations. They focus on how the social world should be studied. Some researchers argue that people and organizations are very different from the subject matter of the natural scientist and require an approach that is more sensitive to the special qualities of people and social life. This issue will also be a major focus of Chapter 2.
- Assumptions about the *nature of social phenomena* influence the research process too. It is sometimes suggested that the social world is external to social actors and that they have no control over it. It is simply there, acting upon and influencing their behaviour, beliefs, and values. The culture of an organization, for example, can be seen as a set of values and behavioural expectations that exert a powerful influence over people who work in it, and into which new recruits have to be socialized. But we could also view culture as something that is constantly being reformulated and reassessed, as members of the organization modify it through practices and through small innovations. Considerations of this kind are **ontological**. They focus on the nature of social phenomena—are they relatively inert and beyond our influence, or are they a product of social interaction? This issue will also be discussed in Chapter 2.
- The *quality criteria* used to evaluate research are a further important influence on the research process. How

do you do good research, and how do you know that a piece of research is good when you read it? Assessments of quality relate to all phases of the research process. As we shall see, assessing research quality has become a prominent issue among business researchers and policy-makers. There are several reasons for this, but the key point here is that debates have arisen about whether there are universal quality criteria that apply to all forms of research. As we will discuss in Chapter 17, some methodologists argue that a more sensitive approach is required whereby quality criteria need to take into account the kind of investigation to which they are being applied.

- The values of the research community have significant implications for research. *Ethical issues* have always been a point of discussion and controversy in research, but in recent times they have become even more prominent. Universities and funding bodies routinely scrutinize research proposals for ethical integrity to ensure that ethical principles are upheld. Ethical values and

the institutional arrangements that are used ensure them have implications for what and who can be researched and for how research can be conducted, to the point that certain research methods are rarely used. Ethical issues are addressed in Chapter 6 and touched on in several other chapters.

- So far, we have stressed the academic nature and purpose of business research. However, much business research has a practical purpose and seeks to make a positive difference to organizations and the people who work in them. This means that studies focus on issues that are likely to have *implications for practice*. Some business research, such as **action research**, discussed in Chapter 17, involves those being researched (such as managers, employees, and consumers, as well as policy-makers) participating in the research process, perhaps by helping to develop research questions. While opinions differ about the need for business research to be directly relevant to practice, this is an issue that researchers are expected to reflect upon.

Relevance to practice

The diverse nature of business research means there is considerable debate about its relationship to practice. Some suggest that management research can be understood as an applied field that is concerned with understanding organizations in order to solve problems of managerial practice. Gummesson (2000) sees academic researchers and management consultants as groups of knowledge workers who place a different emphasis on theory and practice. He writes: 'Backed by bits and pieces of theory, the consultant contributes to practice, whereas the scholar contributes to theory supported by fragments of practice' (2000: 9), but fundamentally their roles are closely related. Gummesson sees researchers and consultants as both involved in addressing problems that concern management, so the value of both groups is determined by their ability to convince the business community that findings are relevant and useful. Other writers, such as Tranfield and Starkey (1998), noted that business researchers have at times lost touch with the concerns and interests of practitioners and that researchers must be responsive to them in order for research to retain its value and purpose. Since the first decade of the twenty-first century there has been some debate around the concept of evidence-based management

(Key concept 1.1). Advocates of evidence-based management suggest that managers need to move their 'professional decisions away from personal preference and unsystematic experience toward those based on the best available scientific evidence' (Rousseau 2006: 256). However, others are more cautious, arguing that the changing and context-dependent nature of management and business makes it difficult to identify generally applicable best practices. Reay et al. (2009) carried out a review of articles that used evidence-based analysis and found that none of them demonstrated a link between adoption of evidence-based management and improved organizational performance. Other researchers (Morrell and Learmonth 2015) have expressed concern that evidence-based management privileges certain kinds of research evidence. They argue that evidence-based management restricts our ability to understand the diversity of problems in management and business studies. These writers further suggest that evidence-based management is characterized by a selective, narrow, and exclusionary view of what counts as evidence, a view that devalues narrative forms of knowledge. Specifically, what is considered to be 'trustworthy and relevant' (2015: 522) by those who favour evidence-based



1.1 KEY CONCEPT

What is evidence-based management?

Evidence-based management is ‘the systematic use of the best available evidence to improve management practice’ (Reay et al. 2009). The approach is proposed as a way of overcoming the ‘research–practice gap’ (Wright et al. 2016), which seeks to address the problem whereby, according to some commentators, business research is insufficiently relevant to practice. The concept developed during the 1990s to enhance patient care in medicine. It was subsequently applied in other fields such as education (Petticrew and Roberts 2006). There are four sources of information that contribute to evidence-based management:

1. practitioner expertise and judgement;
2. evidence from the local context;
3. critical evaluation of the best available research evidence;
4. perspectives of those who may be affected by a particular decision (Briner et al. 2009: 19).

Point 3 is based on the practice of **systematic review** of the literature (see Chapter 5), a cornerstone of evidence-based management practice. The value of evidence-based management depends on whether it enables research findings to be transferred or translated into practice. What is distinctive about evidence-based management is that managerial decision-making is based on explicit, systematic use of research evidence (Wright et al. 2016). Based on their study of an operations management problem in an Australian hospital emergency department, Wright et al. (2016) describe how an evidence-based approach was enacted by a physician manager. Their analysis highlights the importance of ‘fit’ between the personal characteristics of the decision-maker and the organizational context in ensuring the success of this approach. They argue that there is a need for a balanced view that takes into account the critiques of evidence-based management, recognizing the situated expertise of managers in addition to relying on scientific evidence.

management is research that privileges a **positivist** view of knowledge. Morrell and Learmonth conclude that despite claiming to be scientific and impartial, evidence-based management is managerialist, i.e. it privileges and supports management rather than critically analysing it. This builds on earlier arguments by Learmonth (2009), who argues that, because management research operates from within conflicting **paradigms** (see Key concept 2.14), it is not possible to develop a consensus-based notion of evidence that transcends these fundamental philosophical differences.

A further debate that influences understanding of the role of business research stems from the thesis developed by Gibbons et al. (1994). These writers suggest that the process of knowledge production in society falls into two contrasting categories or types, which they describe as ‘mode 1’ and ‘mode 2’.

- **Mode 1.** Within this traditional, university-based model, knowledge production is driven primarily by an academic agenda. Discoveries tend to build upon existing knowledge in a linear fashion. The model makes a distinction between theoretically pure knowledge and

applied knowledge, the latter being where theoretical insights are translated into practice. Limited emphasis is placed on the dissemination of knowledge, because the academic community is viewed as the most important audience or consumer of knowledge.

- **Mode 2.** This model draws attention to the role of *trans-disciplinarity* in research, which refers to a process that causes the boundaries of single contributing disciplines to be exceeded. Findings are closely related to context and not easily replicated, so knowledge production is less of a linear process. Moreover, the production of knowledge is not confined to academic institutions. Instead, it involves academics, policy-makers, and practitioners, who apply a broad set of skills and experiences to tackle a shared problem. Knowledge can be quickly disseminated and findings used to enable practical improvement.

Although mode 2 research is intended to exist alongside mode 1, rather than to replace it, Tranfield and Starkey (1998) argue that business research is more suited to mode 2 knowledge production. Recently, attention has turned towards the challenges of complexity that arise

from the relationship of business research to business practice. This includes the study of complex social issues which are referred to as ‘wicked problems’ (Conklin 2006) or ‘grand challenges’ (Ferraro et al. 2015). The study of wicked problems involves the acknowledgement of large-scale social changes, such as climate change, poverty, and migration, which transcend national boundaries and involve multiple stakeholders with diverse interests. The complex nature of these issues makes it

very difficult to work out exactly what the problem is, or how to study it, and the interlocking nature of **variables** makes it hard to establish patterns of attribution. Hence it is likely that it will be impossible to accurately trace their cause or identify solutions to the problem (Ferraro et al. 2015). The pervasiveness of business in society means business researchers are increasingly likely to be involved in the study of wicked problems, sometimes as members of interdisciplinary research teams.

The process of business research

In the rest of this chapter, we introduce the main elements of a typical research project. It is common for writers of textbooks on business research methods to compile flow charts of the research process, as you will see from, for example, Figures 2.1, 8.1, and 17.1. However, at this stage we do not present the stages or elements of the research process in a sequence, because the order in which they are carried out will vary according to **research strategy** and design. We therefore introduce the main elements that are common to most types of business research and which will be addressed in more detail in later chapters.

Literature review

Existing literature represents an important element in all research. When we have a topic or issue that interests us, we must read further to determine

- what is already known about the topic;
- what concepts and theories have been applied to it;
- what research methods have been used to study it;
- what controversies exist about the topic and the ways in which it is studied;
- what clashes of evidence (if any) exist;
- who are the key contributors to research on the topic.

Many topics have a rich tradition of research, so it is unlikely that you will be able to conduct an exhaustive review of the literature. What is crucial is to identify and read key books and articles by some of the main figures who have written in the field. As we suggest in Chapter 5, you must know what is known, so that you cannot be accused of naively going over old ground. Linking your research questions, findings, and discussion to existing literature is an important and useful way of demonstrating the credibility of your research and the contribution

to knowledge that you are making. However, as will become clear in Chapter 5, a literature review is not simply a summary: it is expected to be critical. This does not necessarily mean that you must be highly critical of existing work, but it does mean that you should assess its significance and how each published item fits into the overall narrative you construct about the literature.

Concepts and theories

In the social sciences, concepts are how we make sense of the social world. They are the labels we give to aspects of the social world that have significant common features. As will be outlined in Chapter 2, the social sciences have a strong tradition of using concepts, many of which have become part of the language of everyday life. Concepts such as bureaucracy, power, social control, status, charisma, labour process, McDonaldization, and alienation are all part of the body of theory that generations of social scientists have constructed. Concepts are a key ingredient of theories.

Concepts serve several purposes in business research. They are important to how we organize and signal our research interests. They help us to think and be more disciplined about what we want to find out about, and help with the organization of research findings. The relationship between theory and research is often depicted as involving a choice between theories driving the research process in all its phases, or theories as a product of the research process. This is invariably depicted as a choice between **deductive** and **inductive** approaches, which will be expanded upon in Chapter 2. Unsurprisingly, this choice has implications for concepts. Concepts may be something we start out with that represent key areas around which we collect data in an investigation. In other words, we might collect data in order to shed light on a concept (or more likely several concepts and how they

are connected). This is the approach taken in the investigation reported in Research in focus 1.3. The alternative view is that concepts are outcomes of research. According to this second view, concepts help us to reflect on and organize the data that we collect.

One of the reasons why familiarity with existing literature is so important is that it alerts us to the main concepts already in use in an area of research and enables us to assess how useful or limited those concepts have been in helping to unravel the main issues. Research in focus 1.3 provides an example of this. Even when we are reading the literature solely as consumers of research—for example, when writing an essay—it is crucial that we know what the main concepts are, who is responsible for them, and what controversies (if any) surround them.

Research questions

Research questions are important in the research process because they force you to consider that most basic of issues—what is it that you want to know? Most people

begin research with a general idea of what they are interested in. Research questions require you to consider much more precisely what you want to find out about (see Key concept 1.2).

Researchers vary in terms of how specific their research questions are, as discussed in Key concept 1.2. In business research it is commonly expected that you will have a clear research question. However, not all business research starts with a clear question. Some researchers start with a general topic or phenomenon they want to study. They choose not to constrain the process of discovery by imposing a predetermined question. These researchers tend to work inductively and build knowledge and theory after they have collected their data. It is important, though, to be aware of the danger that without a clear research question your research may become unfocused. The value of a research question is that it can

- guide your literature search;
- guide your decisions about the kind of **research design** to employ;

1.2 KEY CONCEPT

What are research questions?

A research question provides an explicit statement of what it is the researcher wants to know about. A research purpose can be presented as a statement (for example, 'I want to find out whether [or why] ...'), but a question forces the researcher to be more explicit about what is to be investigated. A research question is interrogatory, i.e. it has a question mark at the end of it. However, research questions are sometimes worded as an overall aim or objective of study and hence not in question format. Research questions can be quite broad, for example 'How do graduates experience job searching?'; or very specific: 'Does frequency of use of social media influence how quickly or slowly graduates obtain employment?'

Denscombe (2010, first edn 2002) provides a list of types of research question:

1. Predicting an outcome: does *y* happen under circumstances *a* and *b*?
2. Explaining causes and consequences of a phenomenon: is *y* affected by *x* or is *y* a consequence of *x*?
3. Evaluating a phenomenon: does *y* exhibit the benefits that it is claimed to have?
4. Describing a phenomenon: what is *y* like or what forms does *y* assume?
5. Developing good practice: how can we improve *y*?
6. Empowerment: how can we enhance the lives of those we research?

White (2009) is critical of Denscombe's last category, arguing that an emphasis on political motives of this kind can impede the conduct of good quality research. To some extent, this difference of opinion can be attributed to differences in viewpoint about the purposes of research. White proposes an alternative:

7. Comparison: do *a* and *b* differ in respect of *x*?

There are many ways that research questions can be categorized, but these seven types provide a rough indication of the possibilities as well as drawing attention to a controversy about the wider goals of research.

1.3 RESEARCH IN FOCUS

A research question about gender bias in attitudes towards leaders

The research question posed in the title of the article by Elsesser and Lever (2011) is 'Does gender bias against female leaders persist?' They begin by reviewing the literature, which suggests that negative attitudes towards female leaders still persist. However, they question whether prior research can be generalized to 'real world scenarios' because much of it is based on 'student samples surveyed on vignettes of hypothetical leaders, attitudes about ideal leaders, or ratings of task leaders in laboratory settings' (Elsesser and Lever 2011: 1556). The aim of their study was therefore to examine whether biases exist towards actual female leaders and, if so, the conditions and management styles that cause such biases to emerge. At one level, this research addresses a practical issue by identifying factors which prevent or discourage women from assuming leadership positions. As noted earlier, it is generally viewed as a good thing that researchers address relevant problems with a view to improving practice. But the authors also draw on theory, in this case role congruity theory, to help explain the processes whereby gender bias against female leaders persists. This theory states that individuals who behave in ways that are incongruent with stereotypically defined sex roles are likely to be viewed negatively. Based on their literature review, the authors present nine hypotheses which they aim to test through data collection and statistical analysis. Hypotheses are typically based on questions but are very specific questions framed as propositions which can be tested.

Their method of data collection was a US-based national survey, titled 'Rate Your Boss', posted on the popular news website msnbc.com in 2007 for ten days. A total of 60,470 people responded to the survey. The majority (68 per cent) had a male boss, but most (89 per cent of women and 78 per cent of men) had experience of both male and female management. Measures included 'relationship quality', 'competence', 'competitiveness', 'sensitivity and directness', and 'preference for a male or female boss'. Preliminary findings were reported in *Elle* magazine before being written up academically. In addition to statistical analysis, the researchers identified a random stratified subsample from 12,440 responses to an optional open-ended follow-up question which asked participants who had expressed a preference for the gender of their boss to explain why. The researchers identified common themes that emerged from these narratives using a grounded theory approach (see Key concept 24.3), providing quotes that were supportive of the themes.

The researchers found a cross-sex bias in how respondents rated their bosses: men judged female bosses more favourably, and women judged male bosses more favourably. A further finding from the qualitative results of the study was that while respondents who said they preferred female bosses cited such things as their compassion or understanding, those who said they preferred male bosses justified this by referring to the negative attributes of female leaders, describing them as too 'emotional', 'moody', 'gossipy', and 'bitchy'. The authors conclude that the answer to their research question is 'yes' and 'no', as participants were less likely to show gender bias when evaluating their own boss, indicating minimal 'bias against women for violating their sex role by adopting a leadership position', but a high level of descriptive bias—'where women are seen as having less potential for management' (Elsesser and Lever 2011: 1571). We return to this example in Chapter 7, where we use it to illustrate the process of writing up **quantitative research**.

- guide your decisions about what data to collect and from whom;
- guide your analysis of data;
- guide your writing-up of findings;
- stop you going off in unnecessary directions; and
- provide your readers with a clear sense of what your research is about.

It is possible that reading the literature may prompt you to revise your research questions and may even suggest some new ones. Therefore, at an early stage of a research study, research questions and the literature relating to them are likely to be intertwined. At the beginning of a research project your initial reading of the literature may generate one or two research questions; further reading guided by the initial research questions may lead you to

revise them and possibly generate new ones. In Chapter 4, there will be more discussion of research questions and how they can be developed.

Sampling

As will be discussed in later chapters, there are a number of principles behind the idea of sampling. Many people associate sampling with surveys and the quest for **representative samples**. Such sampling is usually based on constructing a **sample** that can represent (and therefore act as a microcosm of) a wider **population**. The principles that lie behind the quest for the representative sample will be explained in Chapter 9. These principles often apply to questionnaire **survey research** of the kind described in Research in focus 1.3. In that research, Elsesser and Lever (2011) point out that their sample was unusually large and broadly representative in terms of gender (51 per cent men and 49 per cent women). Respondents covered a wide range in terms of their educational experience and the sector of employment they worked in. However, the sample ‘was not nationally representative, and the survey did not include information on race or ethnicity’ (Elsesser and Lever 2011: 1574). This sample may have been affected by the fact that participation relied on internet access, although in the USA where the study was carried out, internet samples are relatively diverse with respect to gender, age, and socio-economic status. A final limitation relates to the self-selecting nature of the sample, which means that it may be skewed in ways that do not reflect overall patterns of employment. For example, 94 per cent of respondents in Elsesser and Lever’s study were employed full-time and 36 per cent of female and 51 per cent of male participants described themselves as managers; these proportions are not representative of the overall population.

As this example illustrates, even business research, which is traditionally seen as prioritizing representative samples, involves convenience, i.e. making use of the data collection opportunities that are available. In Part Three we encounter sampling principles based not on the idea of representativeness but on the notion that samples should be selected on the basis of their appropriateness to the purposes of the investigation. This is common in **case study** research, where there may be just one or two units of analysis. Here, the goal is to understand the selected case or cases in depth. Sampling issues are still relevant to such research because cases have to be chosen according to criteria relevant to the research, and individuals who are members of the case study context have to be sampled according to criteria too. The key

issue is that sampling is an inevitable feature of most kinds of business research and therefore constitutes an important stage of any investigation.

It is also important to remember that business research is not always carried out on people. For example, we may want to examine mass-media content and employ a technique such as **content analysis**, covered in Chapter 13. In such a situation, we are collecting our data from newspapers or television programmes rather than from people. Because of this, it is common for writers on business research methods to use the term ‘case’ to cover the wide variety of objects on whom, or from whom, data will be collected. Much if not most of the time, ‘cases’ will be people. In business research we are rarely in a position in which we can interview, observe, or send questionnaires to all possible individuals who are appropriate to our research; equally, we are unlikely to be able to read and analyse the content of all articles in all newspapers relating to an area of media content that interests us. Time and cost issues will always constrain the number of cases we can include in our research, so we almost always have to sample.

Data collection

Data collection is the key point of any research project, and therefore this book gives more space to this stage of the research process than any other. Some methods of data collection, such as interviewing and questionnaires, are likely to be more familiar to readers than others. Some methods require a structured approach—that is, the researcher establishes in advance the broad contours of what he or she wants to find out about and designs the research accordingly. The questionnaire is an example of a structured method; the researcher designs questions that will allow data to be collected to answer specific research questions. Similarly, a **structured interview**—the kind of interview used in survey investigations—includes questions designed for exactly the same purpose.

Many methods of data collection are less structured than this. In Part Three we concentrate on research methods that emphasize an open-ended view of the research process, so that there is less restriction on the topics and issues being studied. Research methods such as **participant observation** and **semi-structured interviewing** allow the researcher to keep an open mind about what he or she needs to know about, so that concepts and theories can emerge out of the data. This is the inductive approach to theorizing and conceptualization referred to above. Such research is usually still geared to answering research questions, but these are expressed less explicitly

than in more structured research. This can be seen by comparing the specificity of the hypotheses developed by Elsesser and Lever (2011, Research in focus 1.3) to address their overarching research question, with the question that guided the study of academics in business schools in UK universities by Clarke et al. (2012):

Our objective is to understand how the historical, cultural, economic, political and institutional relations in higher education (and in our case specifically UK business schools) shape or reshape the conditions of identity work and how academic subjectivities are sustained or transformed. In particular, we examine how the cultural, institutional and managerial changes of the last decade or so have affected academic identities.

(Clarke et al. 2012: 6)

This research question, which the authors describe as an ‘objective’, is derived from and illuminated by concepts of managerialism, audit, and performativity. To address the research question, semi-structured interviews were conducted with 48 academics in business schools in UK universities. The interviews were ‘conversations with a purpose’ (Burman 1994)—that of elucidating the impact of new public management on academic identities (Clarke et al. 2012: 8). This is a noticeably less structured approach to data collection, and it reflects the open-ended nature of the research question. Data collection, then, can entail different approaches in terms of how structured or open-ended the methods are.

Data analysis

Data analysis is a stage that incorporates several elements. At the most obvious level, this might mean the application of statistical techniques to data. However, even when data is amenable to quantitative data analysis, there are other things going on when it is analysed. For a start, the raw data has to be *managed*. This means the researcher has to check the data to establish whether there are any obvious flaws. For example, in the research by Clarke et al. (2012), the interviews were audio-recorded and transcribed. **Transcription** enables the researcher to upload the transcripts into a computer software program of the kind discussed in Chapter 25. In the research by Clarke et al., once the transcripts had been uploaded into the software, the authors began by **coding** each transcript. This is a process whereby the data are broken down into component parts which are then given labels. The analyst searches for recurrences of sequences of coded **text** within and across cases and for links between different codes. Clarke et al. began

by identifying ‘descriptive first order’ categories such as ‘emotion’ and ‘changes in the higher education system’ (2012: 8), which they later expanded or collapsed as the analysis progressed, refining them into more analytic categories such as ‘professionalism’, eventually arriving at core themes which they concentrated on. This approach is referred to as **thematic analysis**. There is a lot going on here: data are being made more manageable than they would be if the researcher just kept listening and relistening to the recordings; the researcher is making sense of data through coding; and data are being interpreted—that is, the researcher is linking the process of making sense of the data with the research question, as well as with the literature and theoretical concepts.

The data analysis stage is fundamentally about *data reduction*—that is, reducing the large corpus of information gathered to make sense of it. Unless the researcher reduces the data collected—for example, in the case of quantitative data by producing tables or averages, and in the case of qualitative data by grouping textual material into categories such as themes—it is more or less impossible to interpret the material.

Data analysis can also refer to interpretation of secondary data. Primary data analysis means that the researcher who collected the data conducts the analysis, as was the case with Elsesser and Lever (2011) and Clarke et al. (2012). Secondary data analysis occurs when someone else analyses such data. Researchers in universities are encouraged to deposit their data in archives, which allows others to analyse it. Given the time and money involved in business research, this is a sensible thing to do: it increases the value of an investigation, and a researcher conducting **secondary analysis** can explore the research questions in which he or she is interested without having to go through the time-consuming and lengthy process of collecting primary data. Secondary analysis is discussed in Chapters 14 and 24. As Thinking deeply 1.4 illustrates, the possibilities for analysis of secondary data have grown exponentially because of the internet as a context for business activity such as shopping and career networking. Much information about people’s internet usage is digitally recorded, creating vast potential sources of data.

Writing up

The finest piece of research is useless if it is not disseminated so that others can benefit from it. We do research so that others can read about what we have done and about our findings. Writing up is often neglected, so Chapter 7 is devoted to this topic.

1.4 THINKING DEEPLY

What is big data?

The term 'big data' refers to the vast quantities of digital information generated, stored, and circulated, including via the internet. While understandings of this term vary, Ruppert et al. (2013: 41) provide the following definition:

'big data' refers to large volumes of digital content that is generated either online or offline in social, commercial, scientific and governmental databases. But the term does not simply signify an increase in the volume but also the velocity of data collection and the increasing variety of data sources and formats.

Hence data can be understood as 'big' in several different ways: in addition to volume, we may consider its *velocity* (it can be produced in real time), its *variety*, and its *scope* (it can cover vast populations) (Hand 2014). Big data research has tended to be commercial, often involving big technology corporations such as Google, Facebook, and Amazon. However, social science researchers are becoming increasingly attuned to the potential of big data. One of the challenges they face is that much algorithmically-produced data generated through people's internet usage, including via social media, is relatively unstructured and unformatted. Big data research therefore additionally involves 'the innovation of data structures, computational capacities, and processing tools and analytics to capture, curate, store, search, trace, link, share, visualize and analyse big datasets' (Ruppert et al. 2013: 41). A final issue raised by big data research concerns ethical questions about who owns the data and how it can be accessed (see Chapter 6).

There are different ways to write up research. More structured research, like that presented in Research in focus 1.3, is sometimes written up differently from more open-ended research of the sort represented by Clarke et al. (2012). However, there are core ingredients that dissertations, theses, research articles, and books usually include.

- *Introduction.* The research area and its significance are outlined. The research questions will also probably be introduced.
- *Literature review.* What is already known about the research area is examined critically. This section often relates to theoretical concepts that are the focus of the research, as shown in Table 1.1.

- *Research methods.* The research methods (sampling, methods of data collection, methods of data analysis) are presented and justified.
- *Results.* The findings are presented.
- *Discussion.* The findings are discussed in relation to the literature and the research questions.
- *Conclusion.* The significance of the research is reinforced.

These elements are discussed in detail in Chapter 7. This is not an exhaustive list, because writing conventions differ, but these are recurring elements of the completed research.

Table 1.1 summarizes the seven elements of the research process.

The messiness of business research

There is one final point we want to make before you read on. Business research is often a lot less smooth than you might assume based on accounts of the process that you read in books such as this. Our purpose is to provide an overview of the research process and to give advice on how it should ideally be done. In reality, research is full of false starts, blind alleys, mistakes, and enforced

changes. We know that research is messy from the confessional accounts that have been written over the years (e.g. Hammond 1964; Bell and Newby 1977; Bryman 1988b; Townsend and Burgess 2009; Streiner and Sidani 2010). It is therefore important for business researchers to remain flexible and to modify and adapt their research plans in response to opportunities and problems

TABLE 1.1

Stages of the research process in two studies

| Stage | Description of stage | Example: Elsesser and Lever 2011 * | Example: Clarke et al. 2012 |
|-----------------------|---|--|--|
| Literature review | Critical examination of existing research relating to the phenomena of interest and relevant theoretical ideas. | Literature concerning gender bias in organizational leadership, focusing on role congruity theory. | Literature concerning identity and new public management (including managerialism, audit, and performativity) related to business school academics. |
| Concepts and theories | Ideas that drive the research process and shed light on the interpretation of resulting findings. | Stereotypes; role congruity; female managers; female bosses. | Audit and performative culture; new public management; identity; managerialism. |
| Research question | Question that provides an explicit statement of what the researcher wants to know about. | 'Does gender bias against female leaders persist?' (Elsesser and Lever 2011: 1555)—followed by nine hypotheses (see Research in focus 1.3). | 'Our objective is to understand how the historical, cultural, economic, political and institutional relations in higher education (and in our case specifically UK business schools) shape or reshape the conditions of identity work and how academic subjectivities are sustained or transformed.' (Clarke et al. 2012: 6). |
| Sampling | Selection of sample relevant to the research question(s). | 60,470 men and women who responded to a US-based national survey, 'Rate Your Boss', posted on the msnbc.com website for 10 days in 2007. Subsample of 1000 narratives from 12,440 responses to an open-ended follow-up question asking respondents to explain their gender preferences for a boss. | Sample of 48 lecturers, senior lecturers, readers, and professors working in 8 different UK business schools in the field of organization studies who were interviewed in 2009/10. The sampling approach was both purposeful (i.e. participants worked in a range of different universities) and self-selecting 'because the onus was on participants to respond to a detailed invitation to take part in this study' (Clarke et al. 2012: 8). |
| Data collection | Gathering data from sample so that research question(s) can be answered. | Large-scale web survey involving an online questionnaire (as described in Chapter 11). | Semi-structured interviews between 45 and 70 minutes in length. |
| Data analysis | Management, analysis, and interpretation of data. | Statistical analysis of the questionnaire data. | Thematic analysis of interview transcripts. |
| Writing up | Dissemination of research and findings. | Initial descriptive findings published in <i>Elle</i> magazine and in msnbc.com's financial section in 2007, before being written up as an academic article (see also Chapter 7 on writing up). | Research written up as an article in Clarke et al. (2012) and also in Clarke and Knights (2014) (see also Chapter 7 on writing up). |

*Clarke and Knights (2014).

that arise. Of course, research often does go relatively smoothly and, in spite of minor difficulties, may proceed roughly according to plan. However, what we read in reports of research are often quite sanitized accounts of how the research was carried out, without a sense of the difficult problems the researcher faced. This is not to say that business researchers seek to deceive us, but rather that research when written up tends to follow an implicit template that emphasizes some aspects of the research but not others. This tendency is not unique to *business*

research: in Chapter 7 a study of how natural scientists present and discuss their work shows that here too certain core aspects of the production of 'findings' tend to be omitted from the written account (Gilbert and Mulkay 1984). Bell and Thorpe (2013) draw attention to the role of research communities, including the relationship between research student and supervisor, in passing on research skills: for example, by telling stories about their research experience and by doing research in groups. As we explain in Chapter 4, developing a good working

relationship with your supervisor is crucial in ensuring the success of your research project because this relationship will support you in dealing with unexpected events.

It is also the case that, regardless of the various ways in which research happens, this book can deal only with generalities. It is quite possible that, when doing your research, you will find that these generalities do not fit

perfectly. We therefore urge a degree of caution in the way that you read this book, and we encourage you to adapt the advice we provide to suit your individual circumstances. However, it is also crucial to have an appreciation of businesss research methods as they are explained in the rest of this book because this will provide you with a road map for the journey ahead.



KEY POINTS

- Business research is embedded in a wider context that requires ethical and practical considerations to be taken into account.
- Business research comprises some common elements that are nearly always present. These include: a literature review; concepts and theories; research questions; sampling of cases; data collection; data analysis; and writing-up of research findings.
- Rigorous engagement with these steps is what distinguishes academic business research from practitioner research, such as market research done by private companies.
- Although we can attempt to formulate general principles for doing business research, it is important to recognize that things do not always go entirely to plan.



QUESTIONS FOR REVIEW

What is meant by ‘business research’?

- What is distinctive about academic business research?

Why do business research?

- If you were to start a research project now or in the near future, what would you study and why?

Business research methods in context

- What contextual factors affect the practice of business research and researchers’ choice of methods?

Relevance to practice

- What are the differences between mode 1 and mode 2 forms of knowledge production and why is this distinction important?
- What are ‘wicked problems’ and how might business researchers be involved in addressing them?

The process of business research

- Why is a literature review important when conducting business research?
- What role do concepts and theories play in the process of business research?
- Why are researchers encouraged to specify their research questions? What are the different kinds of research question?
- Why do researchers need to sample? Why is it important for them to outline the principles that underpin their sampling choices?

- Outline one or two factors that might affect a researcher's choice of data-collection method.
- What are the main differences between the kinds of data analysed by Elsesser and Lever (2011) and by Clarke et al. (2012)?
- How might you structure a report of the findings of a research project that you conducted?

The messiness of business research

- If research does not always go according to plan, why should we bother to learn business research methods at all?



ONLINE RESOURCES

www.oup.com/uk/brm5e/

The Interactive Research Guide that accompanies this book contains exercises relevant to each chapter of the book.

Critically Reviewing the Literature

Saunders et al. (2023; 72-122)

Chapter 3



Critically reviewing the literature

Learning outcomes

By the end of this chapter you should be able to:

- understand what is meant by being critical when reviewing the literature;
- recognise the purpose of the critical literature review and its different forms;
- discuss the content of a critical literature review and possible ways to structure it;
- identify the types of literature available;
- plan your literature search strategy and undertake searches using a variety of databases;
- evaluate the relevance, value and sufficiency of potentially relevant literature;
- reference the literature accurately;
- outline the process of systematic review;
- draft a critical literature review;
- understand why you must acknowledge others' work or ideas and avoid plagiarism;
- apply the knowledge, skills and understanding gained to your own research project.

3.1 Introduction

As part of your studies you have almost certainly been asked by your tutors to 'review the literature', 'produce a literature review' or 'critically review the literature' on a given topic. You may, like many students, have grown to fear the literature review. This is not because of the process of searching for and obtaining and reading the literature. It is because of the requirement both to make reasoned judgements about the value of each piece of work and to summarise and synthesise as you organise those ideas and findings that are value into a written product



known as the **critical (literature) review**. It is this process of obtaining and synthesising previous research, making reasoned judgements and organising your thoughts into a written review, that many find difficult and time-consuming.

There are three ways in which you are likely to use literature in your research project (Creswell and Poth 2017):

- as the preliminary search that helps you to generate and refine your research ideas and draft your research proposal (Sections 2.3 and 2.5);
- to provide the context and theoretical framework for your research (the focus of this chapter);
- to place your research findings within the wider body of knowledge and form part of your discussion chapter (Section 14.3).

Most research textbooks, as well as your project tutor, will argue that a critical review of the literature is necessary. Although you may feel that you already have a reasonable knowledge of your research area, we believe that the process of critically reviewing and writing this review is essential. Project assessment criteria usually require you to demonstrate awareness of the current state of knowledge in your subject, its limitations and how your research fits in this wider context. As Synder (2019: 333) notes: ‘building your research on and relating it to existing knowledge is the building block of all academic research activities, regardless of discipline’. This means you have to bring together and discuss what has been published and is relevant to your research topic critically.

The significance of your research and what you find out will inevitably be judged in relation to other people’s research and their findings. Your written review needs to be structured and written to show you understand your field and its key theories, concepts and ideas, as well as the major issues and debates about your topic (Denyer and Tranfield 2009). You therefore need to show you have established what relevant research has been published in your chosen area and, if possible, identified any other research that might currently be in progress. Although the literature you read as part of the reviewing process will enhance your subject knowledge and help you to clarify your research question(s) further, only that which is relevant to your research will be included in your finished critical review.

Unlike some academic disciplines, business and management research makes use of a wide range of literature. While your review is likely to include specific business disciplines such as accounting, finance, operations, strategy, marketing and human resource management, it is also likely to include other disciplines such as economics, psychology, sociology, education and geography. Given this, and the importance of the review to your research, it is vital for you to be aware of what a critical literature review is and the range of literature available before you start the reviewing process. We therefore start this chapter by outlining what is meant by being critical, and the various purposes and forms a critical review of the literature can have (Section 3.2). Subsequently, we consider the structure of the literature review (Section 3.3).

Once you have a good knowledge of the literature sources available (Section 3.4), you can start the process by planning your literature search (Section 3.5) and conducting your search (Section 3.6). Potentially relevant literature obtained can then be read and evaluated (Section 3.7),

The critical review is more than an online retailer's web pages . . .

Students often have difficulties writing their literature reviews for their research projects. . . Mark summarises these:

So, what happens sometimes is . . . a student comes to see me having obviously done a great deal of work. The student has usually already emailed me what they say is the finished critical literature review. Yet the purpose of their review is unclear. It is little more than a summary of the articles and books read, each article or book being given one or two paragraphs. Some students arrange these paragraphs alphabetically in author order; others arrange them in chronological order. None link or juxtapose the ideas. Their literature reviews often look more like an online retailer's web pages than a critical review. Just like the items on these pages, each article or book has some similarities in terms of subject matter and so are grouped together. However, unlike the retailer's web pages, the reasons for these groupings are not made explicit. In addition, while it makes sense to provide



Andrey_Popov/Shutterstock

similar length summary descriptions of items on the retailer's web pages to help the prospective purchaser come to a decision about whether or not to purchase, this is not the case in a literature review. For each article or book in a literature review, the amount written should reflect its value to the research project.

Mark concludes:

While such an approach obviously makes good sense for online retailers and prospective purchasers, it does not work for the critical review of the literature. We obviously need to explain better what we mean by a critical review of the literature to our students.

those which are relevant being noted and referenced (Section 3.8). Alternatively, you may decide that rather than undertaking a traditional literature review, yours will be a self-contained research project to explore a clearly defined research question. In such situations, particularly where questions are derived from organisational practice or policy problems, business and management researchers often adopt the systematic review methodology to critically review the literature. We discuss this in more detail in Section 3.9. You are then ready to start drafting your review (Section 3.10), fully acknowledging your sources and avoiding plagiarism (Section 3.11).

For most research projects, critically reviewing the literature will be an early activity. Despite this early start, it is usually necessary to continue refining your review throughout your project's life. The process can be likened to an upward spiral, culminating in the finished product, a written critical literature review of the literature (Figure 3.1).

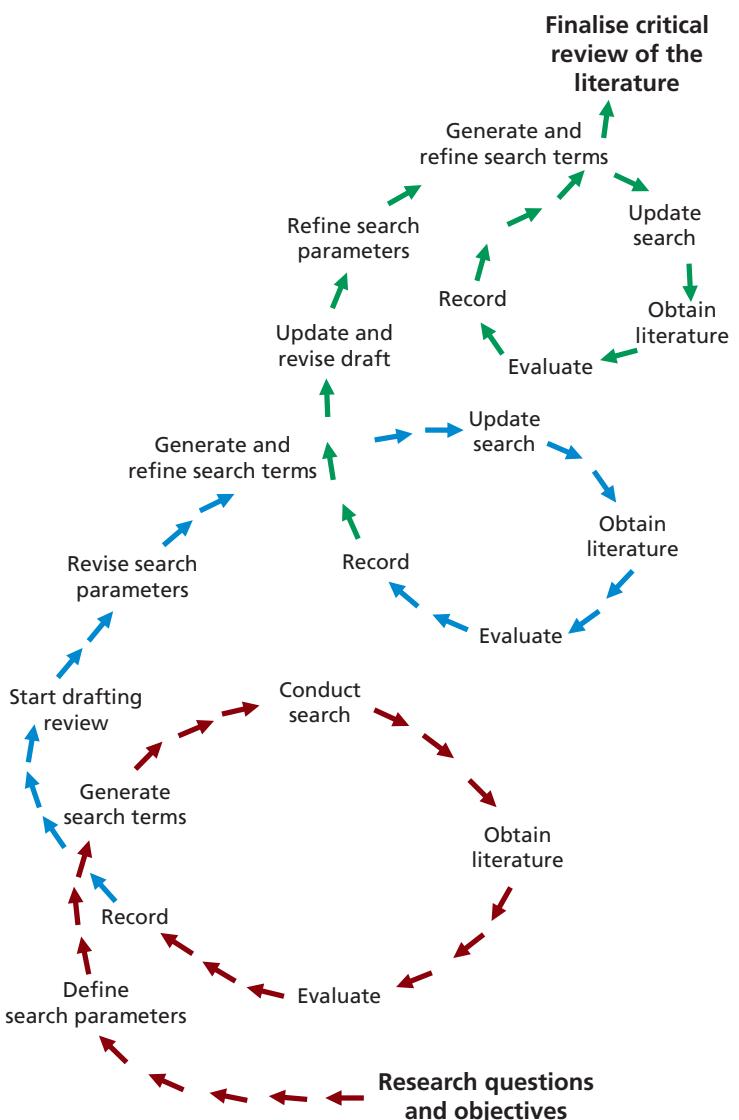


Figure 3.1 The literature review process

3.2 Being 'critical' and the purposes and forms of review

Your critical literature review should be a constructively critical analysis that develops a clear argument about what the published literature indicates is known and not known about your research question (Wallace and Wray 2016). This means, as highlighted in the opening vignette, your literature review is not just a series of paragraphs, each describing and summarising a book or journal article. Rather, you will need to assess what is significant to your research and, on this basis, decide whether or not to include it. If you think

the concepts, theories, arguments or empirical research findings reported and discussed in an article are unclear, biased or inconsistent with other work and need to be researched further, you will need to state this and justify why. This is not easy and requires careful thought. However, by doing this you will be able to produce a reasonably detailed, analytical, constructively critical analysis of the key literature that relates to your research question. Within this you will need to discuss both theories and research findings that support and oppose your ideas.

Being 'critical'

Within your degree programme you have probably already been asked to take a critical approach for previous assignments. However, it is worth considering what we mean by critical within the context of your literature review. Mingers (2000: 225–6) argues that there are four aspects of a critical approach that should be fostered by management education:

- critique of rhetoric;
- critique of tradition;
- critique of authority;
- critique of objectivity.

The first of these, the critique of rhetoric, means appraising or evaluating a problem with effective use of language. In the context of your critical literature review, this emphasises the need for you, as the reviewer, to use your skills to make reasoned judgements and to argue effectively in writing. The other three aspects Mingers identifies also have implications for being critical when reading and writing about the work of others. This includes using other literature sources to question, where justification exists, the conventional wisdom (critique of tradition) and the dominant view portrayed in the literature you are reading (critique of authority). Finally, it is likely also to include recognising in your review that the knowledge and information you are discussing are not value-free (critique of objectivity).

Being critical in reviewing the literature is, therefore, a combination of your skills and the attitude with which you read and your ability to write cogently. In critically reviewing the literature, you need to read the literature about your research topic with some scepticism and be willing to question what you read; the term critical referring to the judgement you exercise. This means as you write your review you need to be constantly considering and justifying your own critical stance with clear arguments and references to the literature rather than just giving your own opinion. As you review the literature, your existing views and opinions are likely to be challenged by what you read. You should welcome these challenges and recognise that through thinking critically about what you are reading, your views and opinions may alter. Critically reviewing the literature for your research project, therefore, requires you to have gained topic-based background knowledge, understanding, the ability to reflect upon and to analyse the literature and, based on this, to make reasoned judgements that are argued effectively in writing. Your written review provides a detailed and justified analysis of, and commentary on, the merits and faults of the key literature within your chosen area.

Part of your critical judgement will involve you in identifying those theories and research findings that are most relevant to your research aims and objectives. This is not as easy as it seems and will invariably involve you in reading and evaluating literature that you subsequently judge is not relevant to your review. For some research topics, as you begin to review the literature, you will observe that a certain theory and set of ideas provide the

theoretical base for much of the research reported. This theory is likely to be considered by researchers as **seminal**. In other words, it has been of great importance or had great influence. Seminal theories will often also be discussed in textbooks on your research area, the associated articles being frequently cited. At the same time, you will begin to recognise those researchers whose work is seminal and has been most influential in relation to your topic. The work of these researchers is likely to be discussed more widely in journal articles, and they may also be referred to by name in textbooks. These researchers are likely to be recognised as the experts in your research area. However, although others consider a particular theory seminal and recognise particular researchers as experts for your topic, this does not mean that you should ignore alternative theories and other researchers. These will also need to be considered in relation to your own research in your critical review.

For other research topics, you are likely to need to integrate a number of different theoretical strands to develop your understanding and provide a firm foundation. You may through your reading discover that there are contrasting theoretical perspectives on the same topic, or it may be that your research needs to integrate two or more theories from different subject areas. Dees (2003) suggests that this means you should:

- refer to and assess research by those recognised as experts in your chosen area accurately;
- consider, discuss and evaluate research that offers both similar and differing perspectives on your chosen area;
- explain your evaluation regarding the value of this research, showing clearly how it relates to your research and acknowledging key work;
- develop a clear, persuasive, logical, balanced and justified argument;
- distinguish clearly between research findings and researchers' opinions;
- ensure your references are completely accurate.

When you draft your critical review (Section 3.10), the extent to which your literature review is critical can be evaluated using the checklist in Box 3.1. The more questions that you can answer 'yes' to, the more likely your review is critical.

Box 3.1 Checklist

Evaluating whether your literature review is critical

Have you:

- ✓ contextualised your own research showing how your research question relates to previous research reviewed, acknowledging seminal work?
- ✓ assessed the strengths and weaknesses of the previous research reviewed in relation to your research topic?
- ✓ been rigorous in your discussion and assessment of previous research?

- ✓ been balanced, referring to research that is counter to, as well as supports, your views and opinions, having now reviewed the literature?
- ✓ distinguished clearly between research findings and researcher's opinions?
- ✓ made reasoned judgements about the value and relevance of others' research to your own?
- ✓ justified clearly your own ideas?
- ✓ highlighted those areas where new research (yours!) is needed to provide fresh insights and taken these into account in your arguments? In particular where:
 - there are inconsistencies in current knowledge and understanding
 - you have identified potential bias in previous research



- there are omissions in published research
 - research findings need to be tested in alternative contexts
 - evidence is lacking, inconclusive, contradictory or limited.
- ✓ developed a clear, persuasive, logical argument?
 - ✓ accurately referenced all research to which you refer?

Purposes of a critical review

The critical review provides the foundation on which your research is built. As you will have gathered from the introduction, a critical review will help you to develop a good understanding and insight into relevant previous research and the trends that have emerged. Likewise, you should not expect to start your research without first reading what other researchers in your area have already found out.

Most critical reviews fulfil a series of related purposes. These can be summarised as providing (Ridley 2018; University of Southern California 2021):

- the historical background to your research;
- an overview of your research's context by locating it in the associated contemporary debates, issues and questions provided by existing literature;
- resolution to conflicts among apparently contradictory previous research;
- a discussion of the relevant theories and concepts that underpin your research;
- definitions and clarifications regarding how relevant terms are being used in your research;
- insights into related research that your own work is designed to extend or challenge;
- supporting evidence that your research questions and aims are worth researching, in other words their significance.

Your literature review therefore contextualises your research in relation to previous research. You are, in effect, providing the background to and justification for your own research project. However, the way you do this in your critical review will depend on the approach you are intending to use in your research. For some research projects you will use the literature to help you to identify theories and ideas that you will subsequently test with data. This is known as a **deductive approach** (Section 4.5) in which you use the literature to develop a theoretical or conceptual framework for subsequent testing. For other research projects the literature review, while outlining what is known, will reveal an aspect about which very little is known or for which there is no clear theoretical explanation. This can be likened to the literature revealing an unopened box within which you do not know what is happening. In such instances, the literature review will provide the context and justification for finding out what is going on inside the box. You will use data to explore what is going on inside the box and from these insights develop a theory or conceptual framework. These will subsequently be related to the literature in your following discussion. This is known as an **inductive approach** (Section 4.5) and, although your research still has a clearly defined aim with research question(s) and objectives, you need to first use data to either get a clearer feeling of what is going on, or better understand the nature of the problem in order to create a conceptual framework or develop a theory. It may also be that you wish to explore the phenomenon in a particular context without being over sensitised to existing theoretical constructs. We believe such an approach cannot be taken without a competent knowledge of the literature in your subject area.

Forms of critical review

The way you organise your critical review depends on your research question and aim. Forms of review have been classified and grouped in a wide variety of ways. One our students have found useful, categorises critical literature reviews into six forms (University of Southern California 2021):

Integrative, critically analysing and examining the main ideas and relationships in representative literature on a topic in an integrative way. The purpose is to provide an overview, and either generate new frameworks and perspectives on a topic for testing or, alternatively, reveal an area where it is unclear what is happening.

Theoretical, examining the body of theory that has accumulated in regard to an issue, concept, theory or phenomenon. Theoretical reviews are often used to establish what theories exist and the relationships between them. They are also used to reveal a lack of appropriate theories or that current theories are inadequate for explaining new or emerging research problems. They can therefore be used as the basis for developing new theory to be tested, or revealing an area where it is unclear what is happening (Box 3.3).

Historical, examining the evolution of research on a particular topic over a period of time to place it in an historical context. They are used to place research in an historical context and identify directions for future work.

Methodological, focussing on research approaches (Section 4.5), strategies (Section 5.5), data-collection techniques or analysis procedures, rather than the research findings. Methodological reviews are often used to provide a framework for understanding a method or methodology and to enable researchers to draw on a wide body of methodological knowledge and can help highlight potential ethical issues.

Argumentative, examining literature selectively to either support or refute well established positions or assumptions. They are used to establish an alternative viewpoint, although care must be taken not to introduce bias if they are used to develop summary conclusions about what is known.

Systematic, which uses a comprehensive pre-planned strategy for locating, critically appraising, analysing and synthesising existing research that is pertinent to a clearly formulated research question to allow conclusions to be reached about what is known (Section 3.9). They are used to summarise all relevant research about the topic.

The most common of these forms for student research projects is the integrative review, although systematic and theoretical reviews are also popular. It is also worth noting that, depending upon the precise focus of your research project, your review may be a combination of these types. For example, a theoretical review may be supplemented with an integrative review, or a historical review may focus on the development of a particular body of theory. Alternatively, following an integrative or theoretical review, a methodological review may be incorporated into the methodology.

It is impossible to review every single piece of the literature before collecting your data. Consequently, your literature review should review the most relevant and significant research on your topic. When you write your critical review, you will need to show how your findings and the theories you have developed, or are using, relate to the research that has gone before. This will help you demonstrate that you are familiar with what is already known about your research topic.

3.3 The content and structure of a critical review

The content of a critical review

As you begin to find, read and evaluate the literature, you will need to think how to combine the academic theories and research findings about which you are reading to form the critical review that will appear in your project report. Your review will need to evaluate the research that has already been undertaken in the area of your research project, show and explain the relationships between published research findings and reference the literature in which they were reported (Appendix 1). It will draw out the key points and trends (recognising any omissions and bias) and present them in a logical way that also shows the relationship to your own research. In doing this, you will provide readers of your project report with the necessary background knowledge to your research question(s) and objectives, and establish the boundaries of your own research. Your review will also enable the readers to see your ideas against the background of previous published research in the area. This does not necessarily mean that your ideas must extend, follow or approve those set out in the literature. You may be highly critical of the earlier research reported in the literature and seek to question or revise it through your own research. However, if you wish to do this, you must still review this literature, explain clearly why you consider it may require revision and justify your own ideas through clear argument and with reference to the literature.

In considering the content of your critical review, you will therefore need:

- to include the key academic theories within your chosen area of research that are pertinent to, or contextualise, your research question;
- to demonstrate that your knowledge of your chosen area is up to date;
- to enable those reading your project report to find the original publications that you cite through clear complete referencing.

When you draft your critical review (Section 3.10) its content can be evaluated using the checklist in Box 3.2.

Possible structures for a critical review

The precise structure of the critical review is usually your choice, although you should check, as it may be specified in the assessment criteria. Three common structures are:



Box 3.2 Checklist

Evaluating the content of your critical literature review

- ✓ Have you ensured that the literature covered relates clearly to your research question and objectives?

- ✓ Have you covered the most relevant and significant theories of recognised experts in the area?
- ✓ Have you covered the most relevant and significant literature or at least a representative sample?
- ✓ Have you included up-to-date relevant literature?
- ✓ Have you referenced all the literature used in the format prescribed in the assessment criteria?

- a single chapter;
- a series of chapters (for example, in a larger research project);
- occurring throughout the project report as you tackle various issues (for example, where your research project is conducted inductively).

In all project reports, you should return to the key issues you raise in your literature review in your discussion and conclusions (Section 14.3).

In the opening vignette we highlighted a common problem with literature reviews: they just describe what each author has written, one author after another (horizontal arrows in Figure 3.2), each item being selected subjectively by the researcher (Hart 2018). It is much easier to be critical (and more interesting to read) if you take a thematic approach comparing and, where necessary, contrasting those authors who discuss each theme (vertical arrows in Figure 3.2). Although there is no single structure that your critical review should take, our students have found it useful to think of the review as a funnel in which you:

- 1 start at a more general level before narrowing down to your specific research question(s) and objectives;
- 2 provide a brief overview of key ideas and themes;
- 3 summarise, compare and contrast the research of the key authors for each theme;
- 4 narrow down to highlight that which is most relevant to your own research question(s) and objectives;
- 5 provide a detailed account of these theories and findings and explain how they are related;
- 6 highlight those aspects where your own research will provide fresh insights linking explicitly to your research question(s) and objectives;
- 7 outline briefly how subsequent sections of your project report will address these aspects.

Whichever way you structure your review, it must demonstrate that you have read, understood and evaluated the literature you have located and know the themes (and their authors) which are key to your own research. The key to structuring a critical literature review is therefore to link the different ideas you find in the literature to form a coherent and cohesive argument, which sets in context and justifies your research. Obviously, it should relate to your research question and objectives. It should show an explicit link

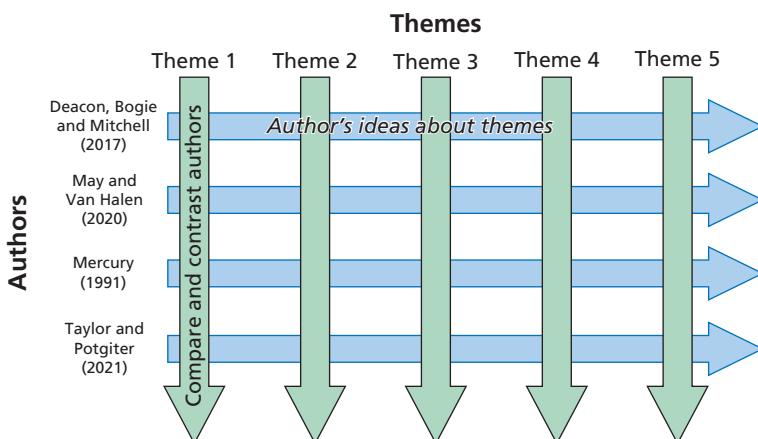


Figure 3.2 Literature review structure

from these as well as a clear link to empirical work that will follow. Subsequent parts of your project report (Section 14.3) must follow on from this. Box 3.3 provides an extract from the literature review in a recently published paper.



Box 3.3 Focus on management research

Structure of the literature review

A refereed academic journal article by Neve Isaeva, Kira Gruenewald and Mark Saunders, published in *The Service Industries Journal* (Isaeva et al. 2020) offers a theoretical review and synthesis of the trust theory and customer service research literature. The following extract is taken from the first three paragraphs of the introduction to their article; although your review will be longer than this extract (pp. 1031–2). It illustrates a structure that:

- in the first paragraph, starts at a more general level looking at the nature and importance of trust for organisational success;
- then, in the second paragraph, narrows the focus to trust research in the service industries, highlighting how scholars have adopted different conceptualisations of trust;
- and, in the third paragraph, builds on these points to justify the need for the review arguing that service industries' trust researchers need to engage more fully with the theoretical foundations of trust, outlining the contribution of the article and justifying their use of a theoretical review.

The remainder of the introduction outlines the structure of their review article.

Trust is a complex, multidisciplinary concept that is recognised as a strategic asset (Castaldo et al. 2010) and vital element for organisational success (Dietz and Gillespie 2011; Kramer and Cook 2004). In particular, service firms have recognised increasingly that the building and maintenance of trusting relationships with customers is the key success factor (Bozic 2017; Váquez Casielles et al.

2005). Correspondingly, fostering customer trust is considered a viable way for tackling challenges surrounding attracting new and retaining existing customers in highly competitive contexts (Sun and Lin 2010). Furthermore, trust is shown to contribute to achieving greater competitive advantage (Barney and Hansen 1994; Berry 1996; Warrington et al. 2000), sales effectiveness (Hu et al. 2003; Johnson and Grayson 2005), customer loyalty (Chen and Mau 2009; La and Choi 2012; Nguyen 2016; Sirdeshmukh et al. 2002), customer commitment (Johnson and Grayson 2005; Nguyen 2016; Pereira et al. 2016; van Tonder and Petzer 2018), improved perception of value (Sirdeshmukh et al. 2002; van Tonder and Petzer 2018), and collaborative, cooperative, and successful exchange relationships (Doney and Cannon 1997; Grayson et al. 2008; Morgan and Hunt 1994).

Trust has experienced a considerable increase of scholarly interest across different service industries (Agariya and Singh 2011; Bachmann and Zaheer 2006) such as finance (Chiao et al. 2008; Gillespie and Hurley 2013; Hansen 2017; Kosiba et al. 2018; Nguyen 2016; Wang et al. 2015), tourism and hospitality (Altinay and Taheri 2019; DeWitt et al. 2008; Gregori et al. 2014; Nunkoo and Smith 2015; Pereira et al. 2016; Rather et al. 2019; Tussyadiah and Park 2018; Wang et al. 2014), healthcare (Calnan and Rowe 2006; Dean et al. 2017; Murray and McCrone 2014; Nienaber and Schewe 2014; Peters and Bilton 2018; Skinner et al. 2004; Şengün and Wasti 2011), media (Filieri et al. 2015; Jackob 2010; Lee 2010), technology (Barua et al. 2018; Jeon et al. 2019; McKnight 2005) and the like. This surge of interest has paralleled the shift of focus from transactional- to relationship-based service orientation, scholars adopting differing conceptualisations of trust.

Applauding this increase in scholarly interest, we contend advancement of service industries' trust research could benefit from engaging more fully with the theoretical foundations of trust, exploring cross-disciplinary debates, and considering

associated questions. Our contribution is therefore to review the fundamental trust theories while synthesising service industries and management literatures on trust; the purpose being to provide a foundation for service industries scholars alongside service firms and their representatives focusing on customer trust. Recognising that a systematic review is only suitable for some research purposes (Petticrew and Roberts 2006), adherence to pre-set criteria restricting the extent of the diversity among included publications (Cassell 2011) and invariably necessitating constraining the focus perhaps to a single industry, we have sought more flexibility and broader focus when evaluating the extant trust theory. Consequently, we have carried out a theoretical review (Saunders et al. 2019), focusing on those trust theories frequently identified and utilised in key trust literature, in particular, the Journal of Trust Research, widely cited research

volumes such as Handbook of Trust Research (Bachmann and Zaheer 2006), Landmark Papers on Trust (Bachmann and Zaheer 2008), Handbook of Advances in Trust Research (Bachmann and Zaheer 2013), and the Handbook of Research Methods on Trust (Lyon et al. 2015), alongside those utilised in 167 most highly cited trust articles considered by Isaeva et al. (2015). These theories are explored in the service industries context using empirical work such as published in The Services Industries Journal. Although we acknowledge the possibility of overlooking potentially relevant and crucial work through such an approach, we believe we provide a clear and comprehensive representation of key trust theory that constitutes a base for those service industries scholars undertaking trust research.

Source: Neve, I., Gruenewald, K. and Saunders, M.N.K. (2020). Copyright © Informa UK Limited, trading as Taylor & Francis Group, Reproduced by permission of the publisher

3.4 Literature sources

The amount of literature available to help you to develop a good understanding of, and insight into, previous research is expanding rapidly as new resources are developed and made available online. The literature sources you are likely to make most use of are often referred to as:

- **white literature** sources, these being formally published scholarly items, particularly journals, that have been peer reviewed;
- **grey literature** sources, these being items that have not been through the peer-review process and have been published in formats such as conference proceedings, dissertations or theses, government reports and other institutions reports.

The process of **peer review** is important as it means that prior to publication the work has been scrutinised and evaluated by others who are experts in the same field and deemed worthy of publication.

Historically, grey literature has been difficult to access as it has not been included in academic databases (Adams et al. 2017). However, while digitisation means grey literature are far easier to find and retrieve, it is crucial to evaluate their credibility (Section 3.7). Your university's librarians are likely to be aware of a wide range of these business and management literature sources that can be accessed, principally from your university library's web pages, and will keep themselves up to date with new resources. Those resources that are most likely to be of use to your literature review are usually print and electronic materials, and those subscription databases that include information from journals, electronic books, magazines and newspapers. In addition, as discussed in Sections 8.2 and 8.4, library subscription databases also offer access to secondary data. Many universities now offer a single search interface, which allows integrated access to search their collections of print, electronic and audio-visual materials including journals, books and all their subscription databases.

The white and grey literature sources you are most likely to use are outlined in Table 3.1. When placing your ideas in the context of earlier research, white literature is likely to be the most important, particularly academic journals. Books authored by academics are likely to be more important than professional and trade journals in this context.

Journals

Journals are also known as ‘periodicals’, ‘serials’ and ‘magazines’, and are published on a regular basis. Journals are a vital literature source for any research. The articles are accessed using full-text databases such as *Business Source Complete* and *Emerald Management Plus*, including through your university’s *single search interface*, this usually being restricted to members of the university (Tables 3.1 and 3.2). Subject to copyright restrictions, many academics also make pre-publication versions of their articles available at no charge on platforms such as *Academia.edu* and *ResearchGate*. While the articles uploaded are not facsimiles of published versions, usually being an earlier draft, they are still extremely useful. Most universities make copies of research outputs authored by their staff available online free of cost or other barriers (**open access**) through their institutional repositories or research archives. In addition, a growing number of national governments, including the UK, have through ‘access to research’ initiatives provided free, walk-in access to academic articles and research in public libraries (Access to Research 2021). Trade and some professional journals may be covered only partially by online databases (Table 3.2) but can usually be accessed through search engines such as Google. You may therefore need to browse these journals’ webpages regularly to be sure of finding useful literature. Beware, although available online, they are often only available to subscribers. For many academic journals you can receive email ‘alerts’ of the table of contents (TOC). TOCs can also be browsed online and downloaded through tertiary literature sources such as *Journal TOCs* and the British Library’s *ZETOC* database (Table 3.2). Similarly, social network platforms allow you to follow particular academics and receive automatic updates regarding their work.

Articles in **peer-reviewed (refereed) academic journals** (such as the *Journal of Management Studies* and the *Academy of Management Review*) are evaluated by academic peers prior to publication to assess their quality and suitability. They are usually written by those considered to be experts in the field, pay rigorous attention to detail and verification of information, and contain an extensive list of references. Such articles are written for a narrower audience of scholars with a particular interest in the field. The language used may be technical or highly specialised as a prior knowledge of the topic will be assumed. Prior to being accepted for publication, articles usually undergo several serious revisions, based on the peer reviewers’ comments, before they are published.

These are usually the most useful for research projects as they will contain detailed reviews of relevant earlier research. Not all academic journals are refereed. Most *non-refereed academic journals* will have an editor and possibly an editorial board with subject knowledge to select articles. The relevance and usefulness of such journals varies considerably and, occasionally, you may need to be wary of possible bias.

Professional journals (such as *People Management*) are produced for their members by organisations such as the Chartered Institute of Personnel and Development (CIPD), the Association of Chartered Certified Accountants (ACCA) and the American Marketing Association (AMA). They contain a mix of news-related items and articles that are more detailed. However, you need to exercise caution, as articles can be biased towards their author’s or the organisation’s views. Articles are often of a more practical nature and more closely related to professional needs than those in academic journals. Some organisations

will also produce newsletters or current awareness publications that you may find useful for up-to-date information. Some professional organisations now give access to selected articles in their journals via their web pages, although these may be only accessible to members. **Trade journals** fulfil a similar function to professional journals. They are published by trade organisations or aimed at particular industries or trades such as catering or mining. Often, they focus on new products or services and news items. They rarely contain articles based on empirical research, although some provide summaries of research. You therefore need to evaluate these particularly carefully if you wish to use them in your research project.

Table 3.1 Sources of white and grey literature

| Source | Content | Use for the literature review | Coverage by online databases | Likely availability |
|---|--|--|--|---|
| Peer-reviewed (refereed) academic journal | Detailed research articles. Written by experts and evaluated by other experts to assess quality and suitability for publication. Rigorous attention paid to detail and verification. | Most useful of all. | Well covered. In addition, content pages often available for searching via publishers' websites. | Online through various subscription services. Increasingly available via institutional repositories, national 'access to research' initiatives or social networking platforms. Those not available may be obtained using inter-library loans. |
| Non-refereed academic journal | May contain detailed reports of research. Selected by editor or editorial board with subject knowledge. | Varies considerably. Beware of bias. | Reasonably well covered. In addition, content pages often available for searching via publishers' websites. | Online through various subscription services. Increasingly available via institutional repositories, national 'access to research' initiatives or social networking platforms. Those not available may be obtained using inter-library loans. |
| Professional journals | Mix of news items and practical detailed accounts. Sometimes include summaries of research. | Insights into practice but use with caution. | Reasonably well covered by online databases. In addition, content pages often available for searching via professional associations' websites. | Online through various subscription services. Those not available may be obtained using inter-library loans. Professional associations may also provide access to their journals via their own web pages. |
| Trade journals/magazines | Mix of news items and practical detailed accounts. | Insights into practice but use with caution. | Content pages often available for searching via professional associations' websites. | Not as widely available in university libraries as peer-reviewed academic journals. Try the trade association's associated website. |

(continued)

Chapter 3 Critically reviewing the literature

Table 3.1 (Continued)

| Source | Content | Use for the literature review | Coverage by online databases | Likely availability |
|------------------------|--|--|---|---|
| Books and e-books | Written for specific audiences. Usually in an ordered and relatively accessible format. Often draw on wide range of sources. | Particularly useful for an overview and to find recognised experts. | Searches can be undertaken on university OPACs.* | Increasingly online through university libraries. Those not available locally may be obtained using inter-library loans. |
| Newsmedia | Written for a particular market segment. Filtered dependent on events. May be written from particular viewpoint. | Good for topical developments. Beware of possible bias in reporting and coverage. | National newspapers reasonably well covered by specialised databases. | Online access to stories, often with additional information for most national and international 'quality' newspapers via university libraries or subscription services. |
| Conference proceedings | Selected papers presented at a conference. | Can be very useful if on same theme as research. | Depends on conference, although often limited. Specialist indexes sometimes available, such as 'Index to conference proceedings'. | Not widely held by university libraries. Can be difficult to find even using search engines. Increasingly only contain abstracts. |
| Reports | Topic specific. Written by academics and organisations. Those from established organisations often of high quality. | Very useful, when matches your topic. | Poor, although some specialised indexes exist. | Not widely held by university libraries. Often available online. May be possible to obtain others using inter-library loans. |
| Theses | Often most up-to-date research but very specific. | Good for doctorate level (and to a lesser extent MPhil) research degrees, otherwise less useful. | Covered by indices of theses. | Increasingly available online, although can also be obtained using inter-library loans. |

*OPAC, Online Public Access Catalogue.

Source: © 2021 Mark Saunders

Table 3.2 Databases, interfaces, platforms and search engines and their coverage

| Name | Coverage |
|---|---|
| Academia.edu | Platform for sharing academic research containing over 22 million articles across all disciplines |
| Access to Research | Database for locating walk-in access to over 30 million research articles (including business and management) via participating UK public libraries |
| British National Bibliography (BNB) | Catalogue of books and serials (journals) deposited at the British Library by UK and Irish publishers since 1950 including electronic publications |
| British Library Integrated Catalogue | Catalogue of print and electronic resources held by the British Library. Includes reference collections and document supply collections |
| British Library Management and Business Studies Portal | Interface to digital full-text research reports, summaries, working papers, consultancy reports, think pieces as well as details of journal articles, sound recordings, video and other resources relevant to business and management |
| British Newspapers 1600–1900 | Cross-searchable interface to full-text British newspapers |
| Business Source Complete (also referred to as EBSCO) | Database including full-text articles from over 3,500 management, business, economics and information technology journals. Contains a wide range of trade and professional titles |
| (The) Conference Index | British Library database containing proceedings of all significant conferences held worldwide (over 400,000 at time of writing) |
| Datassential Reports | Reports, including keynote and market, covering a range of business sectors |
| Emerald Management Plus (EMXP) | Database providing access to over 235,000 articles from over 300 journals in management and complementary subjects |
| EThOS (E Thesis Online Service) | Aggregated database of all doctoral theses awarded by UK HEIs dating back to 1800. Approximately 500,000 records with free access to c. 260,000 digitised theses. |
| Google Scholar | Search engine for scholarly literature across disciplines. Includes citation data and some direct links to downloadable articles |
| Hospitality and Tourism Index | Database of articles in hospitality and tourism journals and trade magazines since 1924 |
| IngentaConnect | Database including full text for articles in business, management, banking, finance and marketing journals |
| ISI Web of Knowledge | Interface to multiple databases including citation indexes for social sciences and for arts and humanities |

(continued)

Chapter 3 Critically reviewing the literature

Table 3.2 (Continued)

| Name | Coverage |
|--|---|
| Journal TOCs | Database of tables of contents (TOCs) for over 35,000 journals. Current awareness service allows journals from which there is a wish to receive future TOCs alerts to be specified |
| JSTOR | Database containing full-text journals, most going back to first issue (in some cases going back to the eighteenth or nineteenth century). Covers sciences, social sciences and arts and humanities. Most recent years usually not available |
| Mintel Reports | Database containing reports of detailed market analysis on wide range of sectors |
| Nexis | Database of full text of UK national and regional newspapers. Increasing international coverage and company profiles and industry reports |
| ProQuest One Business | Database including full text of business journals; company industry and country reports; <i>Wall Street Journal</i> , <i>The Economist</i> and <i>Financial Times</i> |
| Regional Business News | Database of full text regional business publications for the USA and Canada |
| Researchgate | Platform for professionals, researchers and academics to share, discover and discuss research, containing over 100 million publications |
| Sage Journals | Database of full text for peer-reviewed journals, textbooks and digital resources with business-related focus accessible through a business and management hub |
| Science Direct | Database of full text of Elsevier journals including social sciences |
| Social Science Citation Index | Access to current and retrospective bibliographic information, author abstracts and cited references found in over 3,400 social sciences journals covering more than 50 disciplines. Also includes items selected from approximately 3,500 of the world's leading science and technology journals |
| Times Digital Archive 1785–2014 | Database containing complete digital editions (including photographs, illustrations and advertisements) from <i>The Times</i> national newspaper (UK) |
| University library single search interface | A university library's 'one stop' interface to access and search all their print and electronic collections, including subscription databases |
| Wiley Online Library | Database of 1,600 full-text journals including business and law |
| ZETOC | Database giving access to articles and citations of articles. Allows setting up of email alerts of selected journal contents pages |

Books

Books and monographs are written for specific audiences. Some are aimed at the academic market, with a theoretical slant. Others, aimed at practising professionals, may be more applied in their content. The material in books is usually presented in a more ordered and accessible manner than in journals, pulling together a wider range of topics. They are, therefore, particularly useful as introductory sources to help clarify your research question(s) and objectives or the research methods you intend to use. Most academic textbooks can be accessed in electronic form using your university library's *single search interface*; and many, like this one, are also supported by websites providing additional information. However, not all books will be available electronically, some only being available as print copies. In addition, published books may contain out-of-date material even by the time they are published.

News media

News media including newspapers are a good source of topical events, developments within business and government, as well as recent statistical information such as share prices. They also sometimes discuss recent research reports (Box 3.4). Back copies of newspapers starting in the early 1990s are available online via a full-text subscription service, such as *ProQuest One Business* (Table 3.2). Current editions of newspapers are available online and in print, although there is often a charge for full online access. Items in earlier issues are more difficult to access and often only include text. An exception is the *Times Digital Archive 1785–2014* (Table 3.2) of *The Times* newspaper. You need to be careful as, notwithstanding accusations of 'fake news' news media may contain bias in their coverage, be it political, geographical or personal. Reporting can also be inaccurate, and you may not pick up any subsequent amendments or retractions. In addition, the news presented is filtered depending on events at the time, with priority given to more headline-grabbing stories.



Box 3.4 Focus on research in the news

Feeling the strain: stress and anxiety weigh on world's workers

By Emma Jacobs and Lucy Warwick-Ching

The intensity of work since the start of the pandemic pushed Sarah 'close to a breakdown'. The owner of a small UK-based business could not sleep or eat. 'The pressure to keep the business going was all-consuming, I couldn't take time off because I had hundreds of clients relying on me and looking to [me to] guide them through.'

Judging by a global survey by the *Financial Times* on work and mental health, to which more than 250 readers responded, Sarah's experience was not unique. The respondents, who came from all corners of the world, were predominantly white-collar from sectors including education, financial services and media. They spoke of the difficulties – and benefits – of new work practices and about the demands spurred by the pandemic that have affected their mental health.



The pandemic has illuminated the areas of respondents' lives – including career seniority, home environment and caring responsibilities – that have had an impact on people's ability to do their job. Surveys show that mental well-being varies across the world. Britons, according to research by YouGov, are the most likely to report that Covid has harmed their mental health (65 per cent) followed by those in Hong Kong (63 per cent), and Italy (62 per cent) – with Germans the least affected (44 per cent).

Tears, stress and feeling overwhelmed came up regularly. Some had to take time off due to burnout, others spoke of a lack of motivation, difficulty sleeping and increased drinking. '[I've] been ending the day by opening a bottle of wine [or] beer, which has quickly become a daily habit,' says John. In the US, Rachel says her runs are an outlet not just for exercise but so that her kids do not see her cry.

Yet there was also liberation for many workers who had swapped offices for their homes. They could set their own timetables, no longer tethered by the grind of the commute, eat meals with their families and exercise throughout the day. Some praised naps and the joys of watching Netflix in downtime.

Employers' responses varied. Some proved empathetic, others definitely did not. As one respondent put it, the 'workload is insurmountable and [there is] denial about the issues', so that management deem the inability to 'achieve the unachievable objectives' as a personal failing. Here is what readers told the FT in confidence about working during the pandemic. We have used first names only, where we have been given permission, and anonymised some replies ...



Source of abridged extract: Jacobs, E. and Warwick-Ching, L. (2021). 'Feeling the strain: stress and anxiety weigh on the world's workers', FT.com, 8 February. Available at <https://www.ft.com/content/02d39d97-23ed-45ff-b982-7335770ae512> [Accessed 11 March 2021]. Copyright © 2021 Financial Times Limited

Reports

Reports include market analyses and research reports such as those produced by *Mintel* and *Datassential*, government reports and academic reports. Even if you are able to locate these, you may find it difficult to gain access to them because they are often not available free of charge (Section 8.3). Reports are not well indexed in the databases, and you will need to rely on specific search tools such as the *British Library Management and Business Studies Portal* (Table 3.2).

Freedom of information legislation by many governments now means a vast number of reports are now available online; for example, through the European Union's EUROPA website and the Commission's statistics website Eurostat. These and other governmental gateways and archives are listed in Table 8.2.

Conference proceedings

Conference proceedings, sometimes referred to as symposia, are often published as unique titles within journals or as books. Most conferences will have a theme that is very specific, but some have a wide-ranging overview. Proceedings are not well indexed by tertiary literature, so, as with reports, you may have to rely on specific search tools such as *The Conference Index* (Table 3.2) as well as general search engines such as Google. If you do locate and are able to obtain the proceedings for a conference on the theme of your research, you will have a wealth of relevant information. Many conferences have associated web pages providing abstracts and occasionally the full papers presented at the conference.

Theses

Theses are unique and so for a major research project can be a good source of detailed information; they will also be a good source of further references. Unfortunately, they can be difficult to locate and, when found, difficult to access as there may be only one copy at the awarding institution. Specific search tools such as *EThOS*, the *E Thesis Online Service* (Table 3.2) are offering access to an increasing number of digitised theses. Only research degrees, in particular PhDs, are covered well by these tertiary resources. Research undertaken as part of a taught master's degree (usually called a dissertation) is not covered as systematically.

3.5 Planning your literature search

It is important that you plan this search carefully to ensure that you locate relevant and up-to-date literature. This will enable you to establish what research has previously been published in your area and to relate your own research to it. All our students have found their literature search a time-consuming process, which takes far longer than expected. Fortunately, time spent planning will be repaid in time saved when searching for relevant literature. As you start to plan your search, you need to beware of information overload! One of the easiest ways to avoid this is to start the main search for your critical review with clearly defined research question(s), objectives and outline proposal (Sections 2.4 and 2.5). Before commencing your literature search, we suggest that you undertake further planning by writing down your search strategy and, if possible, discussing it with your project tutor. This should include:

- the parameters of your search;
- the search terms and phrases you intend to use;
- the online databases and search engines you intend to use;
- the criteria you intend to use to select the relevant and useful studies from all the items you find.

While it is inevitable that your search strategy will be refined as your literature search progresses, we believe that such a planned approach is important as it forces you to think carefully about your research strategy and justify, at least to yourself, why you are doing what you are doing.

Defining the parameters of your search

For most research questions and objectives, you will have a good idea of which subject matter is going to be relevant. You will, however, be less clear about the parameters within which you need to search. In particular, you need to be clear about the following (derived from Bell and Waters 2018):

- language of publication (e.g. English);
- subject area (e.g. accounting);
- business sector (e.g. manufacturing);
- geographical area (e.g. Europe);
- publication period (e.g. the last 10 years);
- literature type (e.g. peer-reviewed journals and books).

One way of starting to firm up these parameters is to re-examine your lecture notes and course textbooks in the area of your research question. While re-examining these, we

suggest you make a note of subjects that appear most pertinent to your research question and the names of relevant authors. These will be helpful when generating possible search terms and phrases later.

For example, if your research was on the adoption of the United Nations sustainable development goals by public sector organisations, you might identify the subject area as sustainable development and public sector. Implicit in this is the need to think broadly. A frequent comment we hear from students who have attempted a literature search is ‘there’s nothing written on my research topic’. This is usually because they have identified one or more of their parameters too narrowly or chosen their search terms poorly. We therefore recommend that if you encounter this problem you broaden one or more of your parameters to include material that your narrower search would not have located (see Box 3.7).

Generating your search terms

It is important at this stage to read both articles by key authors and recent review articles in the area of your research. This will help you to define your subject matter and to suggest appropriate search terms and phrases. Recent review articles in your research area are often helpful here, as they discuss the current state of research for a particular topic and can help you to refine your search terms. In addition, they will probably contain references to other work that is pertinent to your research question(s) and objectives (Box 3.5). If you are unsure about review articles, your project tutor should be able to point you in the right direction. Another potentially useful source of references are dissertations and theses in your university’s library.

After re-reading your lecture notes and textbooks and undertaking this limited reading, you will have a list of subjects that appear relevant to your research project. You now need to define precisely which search terms are relevant to your research.

The identification of search terms is the most important part of planning your search for relevant literature (Bell and Waters 2018). **Search terms** are the basic terms that describe your research question(s) and objectives, and will be used to search the tertiary literature.



Box 3.5 Focus on student research

Generating search terms

Han’s research question was, ‘How do the actual management requirements of a school pupil record administration system differ from those suggested by the literature?’ She brainstormed this question with her peer group, all of whom were teachers in Singapore. The resulting list included the following search terms and phrases:

schools, pupil records, administration, user requirements, computer, management information system, access, legislation, information, database, security, UK, Singapore, theories

The group evaluated these and others. As a result, the following search terms (and phrases) were selected:

pupil records, management information system, computer, database, user requirement

Online dictionaries and encyclopaedias were used subsequently to add to the choice of search terms:

student record, MIS, security

Han made a note of these prior to using them in various combinations to search her university library’s single search interface.

Search terms (which can include authors' family names identified in the examination of your lecture notes and course textbooks) can be identified using one or a number of different techniques in combination. Those found most useful by our students include:

Discussion

We believe you should be taking every opportunity to discuss your research. In discussing your work with others, whether online or face-to-face, you will be sharing your ideas, getting feedback and obtaining new ideas and approaches. This process will help you to refine and clarify your topic.

Brainstorming

Brainstorming has already been outlined as a technique for helping you to develop your research question (Section 2.3). However, it is also helpful for generating search terms. Either individually or as part of a group, you note all the words and short phrases that come to mind on your research topic (Box 3.5). These are then evaluated and search terms (and phrases) selected.

Initial reading, dictionaries, encyclopaedias, handbooks and thesauruses

To produce the most relevant search terms you may need to build on your brainstorming session with support materials such as dictionaries, encyclopaedias, handbooks and thesauruses, both general and subject specific. These are also good starting points for new topics with which you may be unfamiliar and for related subject areas. Initial reading, particularly of recent review articles, may also be of help here. Project tutors, colleagues and librarians can also be useful sources of ideas.

It is also possible to obtain definitions via the Internet. Google offers a 'define' search option (by typing 'Define:[enter term]') that provides links to websites providing definitions. Definitions are also offered in online encyclopaedias such as Wikipedia. These are often available in multiple languages and, although registered users are allowed to edit the entries, inappropriate changes are usually removed quickly. While entries tend to become more comprehensive and balanced as contributors add to and revise them, Wikipedia (2021) 'makes no guarantee of validity'. However, while online encyclopaedias such as Wikipedia may be useful for a quick reference or in helping to define keywords, your university will almost certainly expect you to justify the definitions in your research project using refereed journal articles or textbooks.

Relevance trees

Relevance trees provide a useful method of bringing some form of structure to your literature search and of guiding your search process (Sharp et al. 2002). They look similar to an organisation chart and are a hierarchical 'graph-like' arrangement of headings and subheadings (Box 3.6). These headings and subheadings describe your research question(s) and objectives and may be terms (including authors' names) with which you can search. Relevance trees are often constructed after brainstorming and can help you decide:

- those search terms that are most relevant to your research question(s) and objectives;
- those areas you will search first and which your search will use later;
- the areas that are more important – these tend to have more branches.



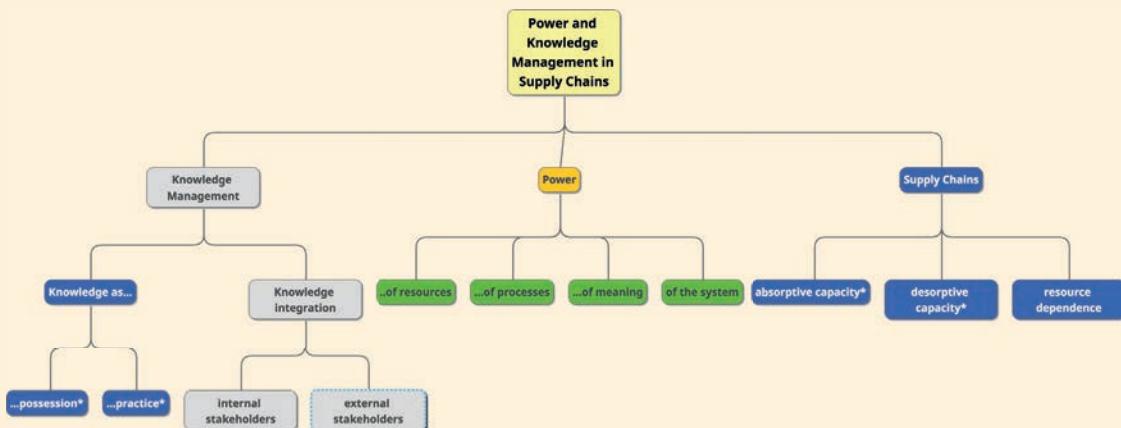
Box 3.6 Focus on student research

Using a relevance tree

Simone's research question was 'How does power facilitate knowledge integration in supply chains?'

After brainstorming her question, she decided to construct a relevance tree on her tablet using the search terms and phrases that had been generated.

Using her relevance tree, Simone identified those areas that she needed to search immediately (in blue) and those that she particularly needed to focus on (starred*).



To construct a relevance tree:

- 1 Start with your research question or an objective at the top level.
- 2 Identify two or more subject areas that you think are important.
- 3 Further subdivide each major subject area into sub-areas that you think are of relevance.
- 4 Further divide the sub-areas into more precise sub-areas that you think are of relevance.
- 5 Identify those areas that you need to search immediately and those that you particularly need to focus on. Your project tutor will be of particular help here.
- 6 As your reading and reviewing progress, add new areas to your relevance tree.

Mind mapping tools such as SimpleMind (2021) and software such as MindGenius (2021) can be used to help generate relevance trees. Many also allow you to attach notes to your relevance tree and can help generate an initial structure for your literature review.

3.6 Conducting your literature search

Your literature search will probably be conducted using a variety of approaches:

- searching online databases;
- obtaining relevant literature referenced in books and journal articles you have already read;
- browsing and scanning secondary literature in your library;
- general online searching;
- institutional repositories and social networking platforms.

Eventually, it is likely you will be using a variety of these in combination. However, we suggest that you start your search by obtaining relevant literature that has been referenced in books and articles you have already read. Although books are unlikely to give adequate up-to-date coverage of your research question, they provide a useful starting point and usually contain some references to further reading. Reading these will enable you to refine your research question(s), objectives and the associated search terms prior to searching using tertiary literature sources. It will also help you to see more clearly how your research relates to previous research and will provide fresh insights.

Searching using databases

It is very tempting with easy access to the Internet to start your literature search with a general search engine such as Google or Bing. While these will retrieve some useful information, they must be treated with care. Your project report is expected to be an academic piece of work and hence must use academic sources. Therefore, it is essential that you use online literature sources that provide access to academic literature. These consist of three types of online databases and are listed in order of likely importance to your search:

- 1 Full-text (online) databases** that index and provide abstracts and full-text of articles from a range of journals (and sometimes books, chapters from books, reports, theses and conferences).
- 2 Abstracts** that only include an index of the abstracts of articles from a range of journals (and sometimes books, chapters from books, reports, theses and conferences), hence the name abstract.
- 3 Indexes** that, as the name suggests, only index articles from a range of journals (and sometimes books, chapters from books, reports, theses and conferences).

Within all of these, the information provided will be sufficient to locate the item – for example, for journal articles:

- author or authors of the article;
- date of publication;
- title of the article;
- title of the journal;
- volume (and part number) of the journal issue;
- page numbers of the article.

Most searches will be undertaken to find articles using user defined search terms or an author's name. Occasionally, you may wish to search by finding those authors who have referenced (cited) an article after it has been published. A citation index enables you to do this as it lists by author the other authors who have cited that author's works subsequent to their publication. Alternatively, if you are using the specialised search engine Google Scholar you can find out who has cited a particular article by clicking on 'cited by'. The article's abstract will be useful in helping you to assess the content and relevance of an article to your research before obtaining a copy. You should beware of using abstracts as a substitute for the full article, as a source of information for your research. They contain only a summary of the article and are likely to exclude much of relevance. Full-text databases usually allow both the searching and retrieval of the full text, principally for journal articles; the articles being retrieved in portable document file (PDF) format. These are read using software such as Adobe Reader, which can be downloaded free of charge.

Your access to the majority of full-text databases will be paid for by a subscription from your university (Table 3.2). Universities provide direct access to these databases

through their library web pages. Many also provide database access through their **library single search interface** allowing you to search both their collections of full text articles, and online public access catalogue (OPAC) of print and digital books and other materials together. There are, however, some pay-as-you-use databases, where the cost of the search is passed on to the user. Specialist search engines (such as Google Scholar) are often free but offer only limited access to the full text (Table 3.4). While many databases are intuitive to use, it is still advisable to undertake your university's online training prior to your search to find out about the specific features available. It is also vital that you plan and prepare your search in advance so your time is not wasted.

Virtually all universities' library OPACs are accessible online. These provide a very useful means of locating resources. If you identify useful collections of books and journals, it is possible to make use of other university libraries during your vacations. Within the UK, the SCONUL Vacation Access Scheme allows students to use books and journals at the 170 institutions which participate in the scheme.¹ In addition, over 70 research libraries in the UK and Ireland (including the British Library, Oxford and Cambridge Universities and the National Libraries of Scotland and Wales) have also made their catalogues available online. These can be accessed through COPAC, the National Academic and Specialist Library Catalogue.²

To ensure maximum coverage in your search, you need to use all appropriate online databases. One mistake many people make is to restrict their searches to their university library's single search interface or Google Scholar rather than using a variety of full text databases. Full text databases' search facilities are usually far more sophisticated and, as each differs both geographically and types of journal included, using a selection ensures a wide coverage of available literature. Some of those more frequently used are outlined in Table 3.2. However, new databases and interfaces are being developed all the time, so it is worth establishing which are available at your university.

Once your search terms have been identified, searching using databases is a relatively straightforward process. You need to:

- 1 Make a list of the search terms that describe your research question(s) and objectives.
- 2 Search appropriate online databases.
- 3 Note precise details, including the search strings used, of the actual searches you have undertaken for each database.
- 4 Note the full reference of each item found; this can normally be done by importing the references into software for managing bibliographies, such as Endnote™ or Reference Manager™ or research tools such as 'Mendeley' or 'Zotero'.
- 5 Wherever possible, import the article into your bibliography or research tool or, alternatively, download it in PDF format and save it on your USB storage device or smartphone using the author, date and a brief description as a filename. This will help you locate it later. For example, an article by Mark and colleagues on gaining access to undertake multi-organisation surveys might be saved using the filename: Saunders[2017] access_surveys.pdf.

Tranfield et al. (2003) emphasise the importance of reporting your literature search strategy in sufficient detail to ensure that your search could be replicated (see Box 3.8). Your review will be based on the subset of those items found which you consider are relevant.

Unlike university library single search interfaces, most full-text databases allow both full-text and abstract only searches using natural language where you decide on the word or phrase combinations for search terms. This means you can search the complete text of an article using your search terms. All relevant results are returned, usually after applying

¹Details of these can be found at: <https://www.sconul.ac.uk/>

²The Internet address for COPAC is <https://copac.ac.uk/>

a process of **lemmatisation** to your search query. This removes all inflectional endings and takes categories and inflections into account to reduce each word used as a search term to its base or ‘lemma’. However, some databases rely on or also offer the option to search using **stemming**. This cuts off a word’s ending in order to determine the word stem. Despite using these tools, your searches may still be unsuccessful. The most frequent causes of failure are summarised in Box 3.7 as a checklist.



Box 3.7 Checklist

Minimising problems with your database search

- ✓ Is the spelling incorrect? Behaviour is spelt with a ‘u’ in the UK but without in the USA.
- ✓ Is the language incorrect? Chemists in the UK but drug stores in the USA.
- ✓ Are you using incorrect terminology? In recent years some terms have been replaced by others, such as ‘redundancy’ being replaced by ‘downsizing’.

- ✓ Are you using recognised acronyms and abbreviations? For example, UK for United Kingdom or BA instead of British Airways.
- ✓ Are you avoiding jargon and using accepted terminology? For example, out-of-the-box thinking rather than creativity.
- ✓ Are you searching over a sensible publication period? For example, the last 15 years rather than the last five years.
- ✓ Are you searching the most suitable type of literature for your research project? For example, peer-reviewed (refereed) journal articles rather than all articles.



Box 3.8 Focus on management research

Framing the system that sustains toxic leadership through followers' continued support

In their recent article in the *International Journal of Management Reviews*, Mergen and Ozbilgin (2021) use a narrative literature review to address the allure of toxic leaders and illustrate the processes and mechanisms that motivate these leaders' followers to remain. Drawing on Bourdieu's concept of *illusio*, whereby individuals believe that the game they engage in is significant and the benefits desirable, they argue that a better understanding of these followers is needed to address the negative consequences of toxic leadership.

Before starting their narrative review, the authors were aware of key texts and theories of both toxic leadership and Bourdieusian *illusio*. They also had a well-defined research purpose: to explain why individuals may become and remain followers of toxic leaders. Their narrative review process started using a small number of articles and books to identify key authors and other articles as starting points in each of the different literatures on which they were drawing. Alongside combing through recent articles in relevant journals they also searched Web of Science and Google Scholar for combinations of search terms. These included ‘toxic, destructive leaders’, ‘pseudo-transformational leaders’, ‘followers’, ‘illusion’, ‘personal uncertainty’ and ‘ethical decision making’. For each search they read abstracts and reviewed those papers considered relevant in detail.

Noting the literature on followership is well-developed they found relevant articles on toxic followers necessitating their use of toxic leadership articles to understand how followers were framed in the literature. Bourdieusian *illusio* was therefore used as the main ‘umbrella’ construct in their building of a conceptual model.

Table 3.3 Search connectors

| Connector | Purpose | Example | Outcome |
|-----------|----------------------------|---|---|
| AND | Narrows search | Recruitment AND interviewing AND skills | Only articles containing all three search terms selected |
| OR | Widens search | Recruitment OR selection | Articles with at least one search term selected |
| NOT | Excludes terms from search | Recruitment NOT selection | Selects articles containing the search term 'recruitment' that do not contain the search term 'selection' |

Searches normally use a combination of search terms linked using **Boolean logic**. These are known as **search strings** and enable you to combine, limit or widen the variety of items found using 'link terms' (Table 3.3). Initially, it may be useful also to limit your search to peer-reviewed journal articles for which the full text is available. It may also be valuable to narrow your search to specific years, especially if you are finding a wealth of items and need to concentrate on the most up-to-date. By contrast, searching by author allows you to broaden your search to find other work by known researchers in your area.

There are, however, problems with searching the full text. In particular, the context of a search term may be inappropriate, leading to retrieval of numerous irrelevant articles and information overload. Fortunately, for most databases you can also search one or more specified fields such as the abstract, author or title. Usually, searching the abstract results in fewer irrelevant articles, although, inevitably, you may not find some relevant ones either! Specifying other fields, for example the author, will be useful if you wish to find articles by a key author in your subject area.

Browsing and scanning

Any search will find only some of the relevant literature. You will therefore also need to browse and scan the literature. New publications such as journal articles are unlikely to be indexed immediately in databases, so you will need to be **browsing** the relevant journals' web pages to gain an idea of their most recent and 'advance online' content. These 'early view' articles are the final version (except for their volume numbering and pagination) and are usually made available through the specific journal's website some months ahead of being included in a specific volume/issue. In contrast, **scanning**, also known as **snowballing** (Hiebl 2021), will involve you going through the reference list of already identified publications to identify references to additional potentially relevant publications (Box 3.8). Scanning is sometimes known as going backwards as it only identifies items that were published earlier. It is particularly important that you browse and scan trade and professional journals, as these are less likely to be covered by the online databases.

To make browsing and scanning easier you should:

- identify when those journals that are the most relevant are published and, where possible, ensure you receive email 'alerts' of their tables of contents (TOCs);
- identify those professional journals that are most relevant and regularly browse them;
- scan new book reviews in journals and newspapers;
- scan publishers' new book catalogues where available;
- discuss your research with your project tutor and librarians, who may be aware of other relevant literature.

Websites of bookshops such as Hive and Amazon can be searched by author, title and subject, and may have reviews attached. Some bookseller websites (and Google Books) have a facility whereby you can view selected pages from the book. However, as when using electronic indexes and abstracts, it is important that you keep full details of the literature you have scanned and browsed (Box 3.9). As well as enabling you to outline the method you used for your literature review, it will also help prevent you repeating searches you have already undertaken.

General searching

When using other search tools, we recommend you keep full details of the searches you have undertaken, making a note of:

- the search tool used;
- the precise search undertaken;
- the date when the search was undertaken;
- the total number of items retrieved.

Search tools, often referred to as **search engines**, are probably the most important method of online searching for your literature review as they will enable you to locate most current and up-to-date items. Although normally accessed through home pages, each search tool will have its own address (Table 3.4). Search tools can be divided into four distinct categories (Table 3.4):

- general search engines;
- metasearch engines;
- specialised search engines and information gateways;
- subject directories.

Most search engines index every separate document. In contrast, subject directories index only the ‘most important’ online documents. Therefore, if you are using a clear term to search for an unknown vaguely described document, use a search engine. If you are looking for a document about a particular topic, use a subject directory.



Box 3.9 Focus on student research

Searching using databases

Matthew described his research project using the search terms ‘marketing’ and ‘non profit’. Unfortunately, he encountered problems when carrying out his search using one of the online databases of full text and abstracts for business, management and economics journals to which his university subscribed.

When he entered the search term ‘marketing’, he retrieved references to over 1,288,000 items, many

of which were in trade magazines. Entering the term ‘non profit’ on its own retrieved fewer references, only 48,000! He was unsure how to combine his search terms into search strings to make his search more specific. Full-text versions were not available for many of the most recent items retrieved.

After discussing the problem, the librarian showed Matthew how to use the advanced search option of the online database. Using this, Matthew first searched using the terms ‘marketing’ AND ‘non profit’ combined as a search string. This still resulted in over 3,700 items being highlighted. He then refined his search further by limiting it to the collection of scholarly (peer-reviewed) journals. This resulted in over 1,300 items being



Box 3.9 Focus on student research (continued)

Searching using databases

retrieved. He therefore decided to limit his search to the abstract field rather than the full text. This resulted in 263 items being retrieved, some of which (including the first) did not appear relevant to his research.

He then copied the references for those items (articles) that appeared relevant onto his smartphone. As Matthew scrolled through these, he noted that some of them had direct links to copies

of the full text stored as a PDF file. For many of the others, the librarian informed him that he could access the full text using different online databases. However, he still needed to assess each article's relevance to his research before obtaining full copies.

Matthew made a note of the details of his search:

| | |
|------------------------|--------------------------|
| Database: | Business Source Premier |
| Collection: | Peer-reviewed journals |
| Dates: | 1951 to 2021 |
| Search: | marketing AND non profit |
| Fields searched: | Abstract |
| Date of search: | 25 March 2021 |
| Total items retrieved: | 263 |

The screenshot shows the EBSCOhost Business Source Premier search interface. At the top, there are search fields for 'Search' and 'Select a Field (optional)'. Below these are three rows for Boolean operators: 'AND', 'OR', and another 'AND'. The search term entered is 'marketing AND non profit'. On the right, the 'UNIVERSITY OF BIRMINGHAM' logo is visible. The main area displays the search history and alerts section, followed by a table of search results. The table has columns for 'Actions', 'Search Options', and 'Search Terms'. Two results are shown: S2 (marketing AND non profit) and S1 (marketing AND non profit). Both results have 'View Results' and 'Edit' buttons. The bottom section shows a detailed view of the first result, titled '1. DIFFERENCES IN ATTITUDES AND BEHAVIORS OF RELIGIOUS TOURISTS: COMPARATIVE RESEARCH BETWEEN BRAZIL AND PORTUGAL.' It includes a thumbnail for 'DIFERENÇAS DE ATITUDES E COMPORTAMENTOS DE TURISTAS RELIGIOSOS: INVESTIGAÇÃO COMPARATIVA ENTRE BRASIL E PORTUGAL.', authors (Yamakawa Zavatti Campos, Waleska Yone; Martins Rodrigues, Maria Carolina; Barbieri da Rosa, Luciana Aparecida; Pommer Barbosa, Raul Alfonso; Jose Sousa, Maria; Cohen, Marcos), journal ('Brazilian Journal of Management / Revista de Administração da UFSM'), volume ('2020 Special Issue, Vol. 13 Issue 5'), pages ('p.1093-1113'), and DOI ('10.5902/1983465956939').

Table 3.4 Selected online search tools and their coverage

| Name | Internet address | Comment |
|-----------------------------------|------------------------|---|
| General search engines | | |
| Bing | www.bing.com | Access to billions of web pages, can link to Facebook |
| Google | www.google.com | Access to billions of web pages |
| Google UK | www.google.co.uk | UK (Country)-based Google – optimised to show country results |
| Specialised search engines | | |
| Google Scholar | www.scholar.google.com | Searches scholarly literature allowing you to locate and sometimes download the complete document, often from an institutional repository |
| UK government | www.gov.uk | Searches central and local government websites and government agencies |
| Information gateways | | |
| Publishers' catalogues homepage | www.lights.ca | Searchable links to major publishers' websites, listed alphabetically by country |
| Subject directories | | |
| Dotdash | www.dotdash.com | Organised by subjects ('brands'), offers numerous guides |

General search engines such as Google and Bing normally use search terms and Boolean logic (Table 3.3) or a phrase. Each search engine indexes and searches automatically, usually finding a very large number of sites (Box 3.10). As people have not evaluated these sites, many are inappropriate or unreliable. As no two general search engines search in precisely the same way, it is advisable (and often necessary) to use more than one. In contrast, metasearch engines allow you to search using a selection of search engines at the same time, using the same interface. This makes searching easier, and the search can be faster. Unfortunately, it is less easy to control the sites that are retrieved. Consequently, metasearch engines often generate more inappropriate or unreliable sites than general search engines.

Specialised search engines cater for specific subject areas. For example, Google Scholar searches scholarly literature across many disciplines using sources such as articles, theses, books and abstracts from academic publishers, professional bodies, universities and websites, allowing you to locate the complete document. Documents are subsequently ranked on a combination of factors including how often it has been cited, where it was published and by whom it was written. Of particular use are two links: the 'cited by' link to more recently published articles which have referenced this document; and the direct link to open access articles stored on institutional repositories and on social networking

sites (discussed next). To use specialised search engines, it is necessary to define your general subject area prior to your search. Information gateways also require you to define your subject area. Information gateways are often compiled by staff from departments in academic institutions. Although the number of websites obtained is fewer, they can be far more relevant, as each site is evaluated prior to being added to the gateway.

Subject directories are searchable catalogues of sites collected and organised by humans. The sites are categorised into subject areas and are useful for searching for



Box 3.10 Focus on student research

Undertaking an online search using a specialist search engine

Kay's research question was reasonably defined, if somewhat broad. She wanted to look at the impact of pandemics, in particular Covid-19, on tourism. As part of her search strategy she decided, alongside the academic databases of business and management journals, to search the Internet using a specialised search engine – Google Scholar. Her first search term 'pandemics tourism' revealed that there were over 18,700

scholarly publications and displayed the first few. Of these, the first appeared to be potentially useful as it focussed on the disruption caused by Covid-19, comparing this to previous epidemics and pandemics.

Kay clicked on the PDF link and, as it was an 'open access' article was able to download it without charge from the publisher's website. She then returned to Google Scholar and clicked on 'Cited by 822'. The first screen revealed a few of the 822 publications that had cited that paper since it had been published. As many could be downloaded as PDF files, she downloaded, and saved those publications that seemed relevant on her smartphone. Kay then made a note of the authors listed on the page, so she could search for them, using her university's online databases.

The screenshot shows the Google Scholar interface with the search term 'pandemics tourism' entered. The results are filtered to show 'Articles' (About 18,700 results). The first result is a paper titled 'Pandemics, tourism and global change: a rapid assessment of COVID-19' by S Gössling, D Scott, CM Hall, published in the Journal of Sustainable Tourism, 2020. It has 99 citations and is available as a PDF. The second result is 'Pandemics, transformations and tourism: be careful what you wish for' by CM Hall, D Scott, S Gössling, published in Tourism Geographies, 2020. It has 99 citations and is available as a PDF. The third result is '[HTML] Tourism in a world with pandemics: local-global responsibility and action' by T Jamal, C Budke, published in the Journal of Tourism Futures, 2020. It has 120 citations and is available as an HTML document. The fourth result is '[HTML] How pandemics affect tourism: International evidence' by G Karabulut, MH Bilgin, E Demir, published in Annals of Tourism, 2020. It has 19 citations and is available as an HTML document. The results are presented in a grid format with title, author, journal, year, citation count, and download link.

broad topics. As people normally compile them, their content has been partly censored and evaluated. Consequently, the number of sites retrieved is fewer, but they usually provide material that is more appropriate (Table 3.4).

Institutional repositories and social networking platforms

Many universities now expect their academics to deposit digital full-text copies of their publications, particularly journal articles, in their **institutional repository**. This is an open access collection of the university staff's research outputs from which full-text items can be downloaded. Increasingly academics (including Mark!) are also uploading pre-publication copies of their journal articles, book chapters and conference papers to social networking platforms such as Academia.edu and ResearchGate. Providing you know the author's name (and their university), and the publication has been uploaded, you can often access pre-publication versions through these resources free of charge. Such institutional repositories and social networking sites are useful if your university does not subscribe to the online database of full-text articles in which their publications are stored, particularly as uploaded copies can often be found using specialised search engines such as Google Scholar (Table 3.4, Box 3.10).

Obtaining the literature

As outlined earlier, searches using online databases (Table 3.2) and search tools (Table 3.4) will provide you with details of what literature is available and where to locate it, in many cases providing a hyperlink to an electronic copy. We emphasise again, you should, whenever possible, download the electronic copy in PDF format and save it on your USB storage device or smartphone. However, where there is no hyperlink, the next stage (Figure 3.1) is to obtain the remaining items. To do this you need to:

- 1 Check your library online catalogue or single search interface to find out whether you can access the appropriate publications.
- 2 For those publications that can be accessed through your library, note their location and: find the publication and scan it to discover whether it is likely to be worth reading thoroughly – for articles it is often possible to make a reasonable assessment of their utility using the abstract (Box 3.11).
- 3 For those items not accessible through your library, it may still be possible to obtain them online, either through institutional repositories or, for books that are no longer copyright, through Google Books.
- 4 Alternatively you may be able:
 - i to borrow the item from another library using the **inter-library loan** service. This is not a free service so make sure it is not available from any other source and you really need it; or
 - ii visit a library where they are held as 'reference only' copies. The British Library in London, for example, has one of the most extensive collection of books, journals, market research reports, trade literature, company annual reports, research reports, doctoral theses and conference proceedings in the world.



Box 3.11 Focus on student research

Assessing the utility of an article using the abstract

Jana's research project was about how companies collaborate with each other. In a search using the Emerald Insight online database she had found a peer-reviewed article in the *European Journal of Training and Development* by Darabi, Saunders and Clark (2020) that she considered might be useful. She decided to read the abstract online to check.

The abstract revealed that the **Purpose** of the article was to explore trust initiation and development

in collaborations between universities and small and medium sized enterprises (SMEs). More details regarding this were given in the **findings** section of the abstract emphasising the focus of the paper on trust and engaged scholarship. The **design/methodology/approach** indicated that the research had been undertaken with 14 SMEs and 12 university stakeholders, qualitative data being collected using semi-structured interviews. Subsequent sections emphasised further the focus on engaged scholarship and trust. Jana wondered if engaged scholarship might be a good aspect to focus on in her research.

Based on this information, Jana decided the article was potentially useful for her research project, so she downloaded it and saved an electronic copy in PDF format.

The screenshot shows a detailed view of an academic article abstract page. At the top, there is a title 'Article' and the full title 'Trust initiation and development in SME-university collaborations: implications for enabling engaged scholarship'. Below the title, the authors are listed as Fariba Darabi, Mark N.K. Saunders and Murray Clark. A summary of the purpose of the study is provided: 'The purpose of this study is to explore trust initiation and development in collaborations between universities and small- and medium-sized enterprises (SMEs) and the...'. On the right side, there are download options for 'HTML' and 'PDF (744 KB)', along with a 'Cite this article' button and a 'Check for updates' link. The main content area is divided into several sections: 'Abstract', 'Details', 'Keywords', 'Findings', and 'Research limitations/implications'. The 'Abstract' section contains a brief summary of the purpose and methodology. The 'Details' section provides publication information: 'European Journal of Training and Development, vol. ahead-of-print no. ahead-of-print', 'Type: Research Article', 'DOI: <https://doi.org/10.1108/EJTD-04-2020-0068>', and 'ISSN: 2046-9012'. The 'Keywords' section lists 'Trust', 'Engaged scholarship', 'SME', 'University', 'Collaboration', and 'HRD'. The 'Findings' section discusses the role of calculus-based trust in initiating collaborations. The 'Research limitations/implications' section notes that the study is based on a collaborative research between eight SMEs and one university business school and does not reflect ES fully as conceptualised.

Source: Darabi et al. (2020). Copyright © 2020 Emerald Group publishing (<https://www.emerald.com/insight/content/doi/10.1108/EJTD-04-2020-0068/full/html?skipTracking=true>). Reproduced by permission of the publisher.

3.7 Reading critically and evaluating the literature

Adopting a critical perspective in your reading

Harvard College Library (2020) provides its students with a useful list of strategies thinking intensive reading. These include:

Previewing: developing a set of expectations about the scope and aim of the item before you start reading in detail to establish how it may inform your literature search.

Annotating: conducting a dialogue with yourself, the author and the issues and ideas at stake and marking those aspects that seem important in the margins of the text using words and phrases rather than just highlighting or underlining. This is crucial as a reason for highlighting in a bright colour is often difficult to remember later without notes as to why.

Outlining, summarising and analysing: the best way to determine that you've really got the point is to be able to state it in your own words. Outlining the argument of an item is a version of annotating, and can be done quite informally in the margins of the text. Summarising does the same thing, making the connections between ideas explicit; analysing incorporates an evaluation to your summary. Analysing comprises your reflections on the effectiveness or otherwise of the arguments made.

Looking for repetitions and patterns: using these, alongside the way authors use language, to indicate what is important.

Contextualising: looking at what you have read and acknowledging how it is framed by other work.

Comparing and contrasting: asking yourself how your thinking has been altered by this reading and how it has affected your response to the issues and themes you have already considered.

Wallace and Wray (2016) recommend the use of **review questions**. These are specific questions you ask when reading, which will be linked either directly or indirectly to your research question. So you may, for example, address a piece of reading with the view to it answering the question: 'What does research suggest are the main reasons why customers are likely to change car insurance provider?'

The word 'critical' has appeared in this chapter a number of times so far. It is vital in your reading of the literature that a critical stance should be taken. So what is meant by critical reading? Wallace and Wray (2016: 8) sum this up rather succinctly by saying that critical skills 'can be boiled down to the capacity to evaluate what you read and the capacity to relate what you read to other information'.

More specifically, Wallace and Wray (2016) advocate the use of five critical questions to employ in critical reading. These are:

- 1 Why am I reading this? (The authors argue that this is where the review question is particularly valuable. It acts as a focusing device and ensures that you stick to the purpose of the reading and do not get sidetracked too much by the author's agenda.)
- 2 What is the author trying to do in writing this? (The answer to this may assist you in deciding how valuable the writing may be for your purposes.)
- 3 What is the writer saying that is relevant to what I want to find out?
- 4 How convincing is what the author is saying? (In particular, is the argument based on a conclusion which is justified by the evidence?)
- 5 What use can I make of the reading?

Evaluating the literature

A question frequently asked by our students is, ‘How do I know what I’m reading is relevant?’ Two further questions often asked by our students are, ‘How do I assess the credibility of what I read?’ and ‘How do I know when I’ve read enough?’ All of these are concerned with the process of evaluation. They involve defining the scope of your review and assessing the relevance and credibility of the items that you have obtained in helping you to answer your research question(s) and meet your objectives.

Assessing relevance

Assessing the relevance of the literature you have collected to your research depends on your research question(s) and objectives. Remember that you are looking for relevance, not critically assessing the ideas contained within. When doing this, it helps to have thought about and made a note of the criteria for inclusion and exclusion prior to assessing each item of literature. Box 3.11 also provides some help here.

You should, of course, try to read all the literature that is most closely related to your research question(s) and objectives. For some research questions, particularly for new research areas, there is unlikely to be much closely related literature and so you will have to review more broadly. For research questions where research has been going on for some years, you may be able to focus on more closely related literature.

Remember to make notes about the relevance (and credibility) of each item as you read it and the reasons why you came to your conclusion. You may need to include your evaluation as part of your critical review.

Assessing credibility

Assessing the credibility of the literature you have collected is concerned with the quality of the research that has been undertaken. As such it is concerned with issues such as methodological rigour, theory robustness and the quality of the reasoning or arguments. For example, you need to beware of managerial autobiographies, where a successful entrepreneur’s or managing director’s work experiences are presented as the way to achieve business success (Fisher 2010), and articles in trade magazines. The knowledge presented in such books and articles may well be subjective rather than based upon systematic research.

For refereed journal articles (and some book chapters), the review process means that peers have assessed the quality of research and suggested amendments before they are published. This means the research is likely to have been undertaken with methodological rigour, have used theory appropriately and been argued cogently. However, it is still important to assess the value yourself in terms of possible bias, methodological omissions and precision (Box 3.12).

It is worth noting that, within business and management and other subjects, lists exist that rank peer-reviewed journals according to their quality; higher rankings indicating better quality journals. The fortunes of academics and their business schools depend on publishing in such highly ranked journals. Harzing (2020) provides a regularly updated Journal Quality List for business and management, which includes lists from 12 different sources. While there is little doubt that journals ranked highly on lists are quality journals and are more likely to contain quality articles, this does not mean that every single paper within them will be of the same high quality. It also does not mean that articles in lower-ranked journals are of little value. Consequently, although journal ranking lists can provide a broad indicator of the quality of research, they are not a substitute for reading the article and making your own assessment of the quality of the research in relation to your research question(s) and objectives. The checklist in Box 3.12 will help in this assessment.



Box 3.12 Checklist

Evaluating the relevance, value and sufficiency of literature to your research

Relevance

- ✓ How recent is the item?
- ✓ Is the item likely to have been superseded?
- ✓ Are the research questions or objectives sufficiently close to your own to make it relevant to your own research?
- ✓ Is the context sufficiently different to make it marginal to your research question(s) and objectives?
- ✓ Have you seen references to this item (or its author) in other items that were useful?
- ✓ Does the item support or contradict your arguments? For either it will probably be worth reading!

Credibility

- ✓ Has the item been subject to a reviewing process prior to publication?

- ✓ Does the item appear to be biased? For example, does it use an illogical argument, emotionally toned words or appear to choose only those cases that support the point being made? Even if it is, it may still be relevant to your critical review.
- ✓ What are the methodological omissions within the work (e.g. sample selection, data collection, data analysis)? Even if there are many it still may be of relevance.
- ✓ Is the precision sufficient? Even if it is imprecise it may be the only item you can find and so still of relevance!
- ✓ Does the item provide guidance for future research?

Sufficiency

- ✓ As I read new items, do I recognise the authors and the ideas from other items I have already read?
- ✓ Have I read the work by those acknowledged by others as key researchers in my research area?
- ✓ Can I critically discuss the academic context of my research with confidence?
- ✓ Have I read sufficient items to satisfy the assessment criteria for my project report?

Sources: Authors' experience; Bell and Waters (2018); Colquitt (2013); Fisher (2010); Jankowicz (2005)

Assessing sufficiency

Your assessment of whether you have read a sufficient amount is even more complex. It is impossible to read everything, as you would never start to write your critical review, let alone your project report. Yet you need to be sure that your critical review discusses relevant research that has already been undertaken and that you have positioned your research project in the wider context, citing the main writers in the field (Section 3.2). One clue that you have achieved this is when further searches provide mainly references to items you have already read (Box 3.12). You also need to check what constitutes an acceptable amount of reading, in terms of both quality and quantity, with your project tutor.

3.8 Note-taking and referencing

The literature search, as you will now be aware, is a vital part of your research project, in which you will invest a great deal of time and effort. As you read each item, you need to ask yourself how it contributes to your research question(s) and objectives and to make notes with this focus (Bell and Waters 2018). When doing this, many students download

and save copies of articles or photocopy or scan pages from books to ensure that they have all the material. We believe that, even if you save, print or photocopy, you still need to make notes.

The process of note-making will help you to think through the ideas in the literature in relation to your research. When making your notes, make sure you always use quotation marks and note the page number if you are copying the text exactly. This will ensure you know it is a direct quotation when you begin to write your project report and so help you avoid committing plagiarism (Section 3.11). The Harvard College Library (2020) suggests that you should get into the habit of hearing yourself ask questions of your reading and makes notes as you read. Their advice is summarised in Box 3.13.

In addition to making notes, it is helpful to record the:

- bibliographic details;
- brief summary of content;
- supplementary information.

Bibliographic software such as Reference Manager™, EndNote™ or research tools such as ‘Mendeley’ or ‘Zotero’ provide a powerful and flexible method for recording the literature and automatically generating references in the required style and, for some, can send suggestions for further reading based on the references you have uploaded. In addition, there are online bibliography generators such as ‘Cite This For Me’ that can help you create a bibliography or reference list in the prescribed format. Many specialist search engines, such as Google Scholar, allow references (and in some case full text) to be exported directly into such software and tools. Where this is not the case, recording can seem very tedious, but it must be done. We have seen many students frantically repeating searches for items that are crucial to their research because they failed to record all the necessary details in their database of references.

Box 3.13 Checklist

Making notes when reading

- ✓ Don't use a highlighter pen. Highlighting can actually distract from the business of learning and dilute your comprehension. It only seems like an active reading strategy; in actual fact, it can lull you into a dangerous passivity.
- ✓ Mark up the margins of the text with words: ideas that occur to you, notes about things that seem important to you, reminders of how issues in a text may connect with your research questions and objectives. If you are reading a PDF copy on screen, use the ‘sticky notes’ feature of Adobe Reader®. This kind of interaction keeps you conscious of the reason you are reading. Throughout

your research these annotations will be useful memory triggers.

- ✓ Develop your own symbol system: asterisk a key idea, for example, or use an exclamation mark for anything that is surprising, absurd, bizarre etc. Like your margin words, your symbols can help your thoughts when you first read it. They will be indispensable when you return to a publication later in the term, in search of a particular passage that you may want to refer to in your project report.
- ✓ Get in the habit of hearing yourself ask questions – ‘what does this mean?’; ‘why is she or he drawing that conclusion?’ Write the questions down (in your margins, at the beginning or end of the reading, in a notebook or elsewhere). They are reminders of the unfinished business you still have with a text: to come to terms with on your own, once you've had a chance to digest the material further or have done further reading.

Bibliographic details

For some project reports you will be required to include a **bibliography**. Convention dictates that this should include all the relevant items you consulted for your project, including those not directly referred to in the text. For the majority, you will be asked to include only a list of **references** for those items referred to directly in the text. The **bibliographic details** contained in both need to be sufficient to enable readers to find the original items. These details are summarised in Table 3.5.

If you located the item online, you need to record the full Internet address of the resource and the date you accessed the information (Appendix 1). This address is the URL, the unique resource location or universal/uniform resource locator. For a journal article accessed online, and some other electronic documents, it is becoming more usual to also include that document's **digital object identifier** (DOI). The DOI provides a permanent and unique two-part identifier for the electronic document.

Most universities have a preferred referencing style that you must use in your project report. This will normally be prescribed in your assessment criteria. Three of the most common styles are the Harvard system (a version of which we have used in this book), the American Psychological Association (APA) system and the Vancouver or footnotes system. Guidelines on using these are given in Appendix 1.

Brief summary of content

A brief summary of the content of each item in your reference database will help you to locate the relevant items and facilitate reference to your notes and photocopies. This can be done by annotating each record with the search terms used, to help locate the item and the abstract. It will also help you to maintain consistency in your searches.

Supplementary information

As well as recording the details discussed earlier, other information may also be worth recording. These items can be anything you feel will be of value. In Table 3.6 we outline those that we have found most useful.

Table 3.5 Bibliographic details required

| Journal | Book | Chapter in an edited book |
|--|---|--|
| <ul style="list-style-type: none"> • Author(s) – family name, first name, initials • Year of publication (in parentheses) • Title of article • <i>Title of journal</i> (italicised) • Volume • Part/issue • Page numbers (preceded by 'p'. for page or 'pp.' for pages) | <ul style="list-style-type: none"> • Author(s) – family name, first name initials • Year of publication (in parentheses) • <i>Title and subtitle of book</i> (italicised) • Edition (unless first) • Place of publication • Publisher | <ul style="list-style-type: none"> • Author(s) – family name, first name initials • Year of publication (in parentheses) • Title of chapter • Author(s) of book – family name, first name initials • <i>Title and subtitle of book</i> (italicised) • Edition (unless first) • Place of publication • Publisher • Page numbers of chapter (preceded by 'pp.' for pages) |

Table 3.6 Supplementary information

| Information | Reason |
|---|---|
| ISBN | The identifier for any book, and useful if the book has to be requested on inter-library loan |
| DOI | The digital object identifier is both permanent and unique, meaning an electronic document can be found more easily |
| Class number (e.g. Dewey decimal) | Useful to locate print copies of books in your university's library and as a pointer to finding other books on the same subject |
| Quotations | Always note useful quotations in full and with the page number of the quote; if possible, also save entire document as a PDF file |
| Where it was found | Noting where you found the item is useful, especially if it is not in your university library and you could only take notes |
| The search engine, database or other resource used to locate it | Useful to help identify possible resources for follow-up searches |
| Evaluative comments | Your personal notes on the value of the item to your research in relation to your relevance and value criteria |
| When the item was consulted | Especially important for items found via the Internet as these may disappear without trace |
| Filename | Useful if you have saved the document as a PDF file |

3.9 Using systematic review

Systematic review is a replicable process for reviewing the literature using a comprehensive pre-planned strategy to locate existing literature, evaluate the contribution, analyse and synthesise the findings and report the evidence to allow conclusions to be reached about what is known and, also, what is not known (Denyer and Tranfield 2009). Conducting a systematic review can be likened to being a judge and jury, where you evaluate sceptically the evidence to come to the fairest judgement possible (Siddaway et al. 2019).

Originating in the medical sciences, systematic review has been used widely to evaluate specific medical treatments; in the past three decades its importance has been recognised in other disciplines. Within business and management, not all researchers have greeted systematic review with enthusiasm. While some argue that the systematic review method


**Box 3.14
Checklist**

Establishing whether a project may be suitable for systematic review

- ✓ Is there uncertainty about the effectiveness of the policy/service/intervention?
- ✓ Is there a need for evidence about the likely effects of a policy/service/intervention?

- ✓ Despite a large amount of research on the topic, do key questions remain unanswered?
- ✓ Is there a need for a general overall picture of the research evidence on the topic to direct future research?
- ✓ Is an accurate picture of past research and associated methods needed to help develop new methods?

(If the answer to one or more of these is 'yes' then the project may be suitable for systematic review.)

Source: Developed from Petticrew and Roberts (2006)

addresses potential research bias common in other forms of literature reviews, others argue that all evidence is inherently subjective and impacted by politics, values and interests (Rojon et al. 2021). Notwithstanding, Denyer and Tranfield (2009) have adapted the medical sciences guidance, ensuring that the process is transparent, inclusive, explanatory and enables learning. Systematic reviews usually, although not exclusively, focus on policy or practice questions such as the effectiveness of a particular intervention and the associated mechanisms with an emphasis on informing action. It is therefore not surprising that Petticrew and Roberts (2006) argue that systematic review is only suitable for some research projects (Box 3.14), emphasising that it is time-consuming and the need to involve others in the process.

Prior to conducting a systematic review, most writers suggest you undertake an exploratory **scoping study** to assess whether or not other systematic reviews have already been published and determine the focus of the literature search. Subsequent to this, a five-step process in which each stage is recorded precisely can be followed (Rojon et al. 2021):

1 *Review scope and question(s)* of systematic review, involving a broad range of expert stakeholders such as potential academic and practitioner users of the review as an advisory group and initial, non-systematic, scoping review. Resulting review questions, for example 'What are marketing professionals' understanding and definition of viral marketing?', can be developed using the CIMO acronym (Denyer and Tranfield 2009). This emphasises the need to include review questions which relate to the:

Context – the individuals, relationships or wider settings being researched;

Intervention – the effects of the events, actions or activities being researched;

Mechanisms – the mechanisms that explain how the intervention (within the context) results in the outcome;

Outcome – the effects of the intervention and how they are measured (Jones and Gatrell 2014).

2 *Search literature* to locate and generate a comprehensive list of potentially relevant research studies using online database searches, specialist bibliographies, tables of contents and other sources (Section 3.3).

- 3 Select and evaluate relevant research studies through:**
- a Initial review**, usually by title and abstract, to screen relevant research studies using predetermined explicit inclusion and exclusion (selection) checklists of criteria to assess the relevance of each in relation to the review question(s). These checklists can be developed by undertaking a small number of pilot searches and making a list of reasons for inclusion or exclusion of each article or adapting checklists developed for previous systematic reviews, by journals to assess general quality of research or to assess issues of relevance and value (Box 3.14). Common criteria include adequate methods, clear data analysis and conclusions derived from findings.
 - b Further review and data extraction**, usually evaluating by reading the full text of those not excluded in the initial review, breaking down each study into its constituent parts and recording the key points (research question/aim; study context – country, industry sector, organisational setting etc.; method(s) of data collection; sample size, frame and demographics; key findings; relevance to review questions) on a data extraction template.
- 4 Analysis and synthesis** of the relevant research studies using the data extraction forms to explore and integrate the studies and answer the specific review questions.
- 5 Reporting the results** including (Denyer and Tranfield 2009):
- a** an introductory section that states the problem and review questions;
 - b** a methodology section that provides precise details of how the review was conducted (search strategy, selection criteria, key points used for the analysis and synthesis) (Sections 3.3 and 3.4);
 - c** findings and discussion sections that review all the studies (Section 3.2), specifying precisely what is known and what is not known in relation to the review questions. Often researchers adopt the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta Analyses) checklist (Moher et al. 2009) and flow diagram for reporting and presenting their systematic review methodology. Using a flow diagram (Figure 3.3) allows the number of studies reviewed in steps two through four of the systematic review process to be reported clearly.

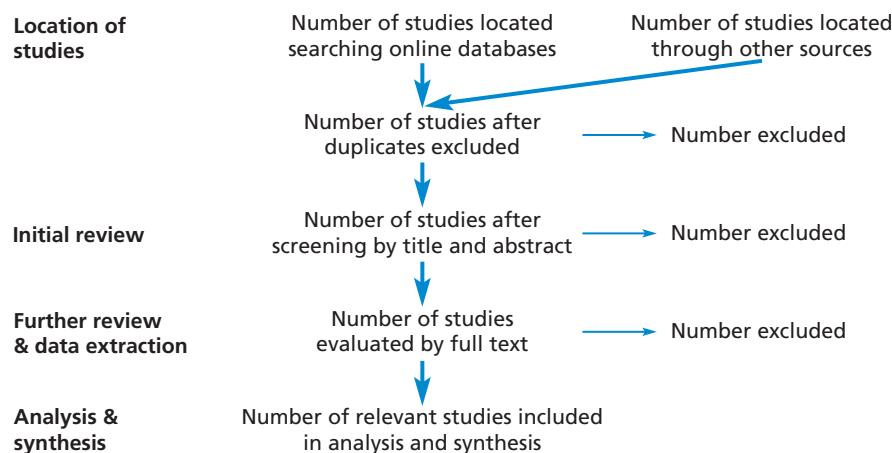


Figure 3.3 Reporting the systematic review process

Source: Developed from Moher et al. 2009; Rojon et al. 2021

3.10 Drafting your critical review

As we saw in Sections 3.2 and 3.3, the literature review that you write for your project report should be a description and critical analysis of what other authors have written, providing the background for your own research. When drafting your review, you therefore need to focus on using the literature to contextualise and justify your research question(s) and objectives. One way of helping you to focus is to think of your literature review as discussing how far existing published research goes in answering your research question(s). The shortfall in the literature will be addressed, at least partially, in the remainder of your project report – unless your entire research project is a literature review! Another way of helping you to focus is to ask yourself how your review relates to your objectives. If it does not, or does so only partially, there is a need for a clearer focus on your objectives.

In drafting your critical review, you will need to juxtapose different authors' ideas and form your own opinions and conclusions, comparing and contrasting these to form an evidence-based evaluation of the literature you have used. In doing this you will need to ensure the key themes are presented logically and that you highlight explicitly those areas where your own research will provide fresh insights, restating the research questions you will answer (Section 14.3). Subsequently, as part of your method, you will explain how you obtained the data to enable these questions to be answered before outlining and discussing your findings (Section 14.3). Although you will not be able to start writing until you have undertaken some reading, we recommend in addition to notetaking you start summarising your literature thematically early on. A way of doing this, which our students have found helpful, is to create a Thematic Analysis Grid (Anderson et al. 2015). This helps you to structure your notes for each article in a matrix with articles listed in rows in date order and each column representing a separate theme (Box 3.15, Figure 3.2). Summary notes for each article pertaining to a specific theme are then inserted in the appropriate cell.

To construct a Thematic Analysis Grid, you:

- 1** Identify potential themes from your initial annotated reading of the literature; these will form the grid's columns.
- 2** For each journal article, book or book chapter, insert a new row (keeping the date order) and make brief notes under the appropriate potential themes (columns); these should be in your own words, the page number being noted for quotations.
- 3** In a final column, 'methodology', note briefly the methodology used.
- 4** As your grid develops be prepared to:
 - a** add new themes;
 - b** remove themes that are no longer relevant to your research question;
 - c** introduce sub themes.
- 5** As your grid develops further, look for patterns emerging across the themes; look for where:
 - a** there is consensus;
 - b** there are contradictions;
 - c** the literature you have reviewed is most convincing (the methodology column will help here).

As your Thematic Analysis Grid develops you will be able to identify those themes that have been widely researched and those that are less well developed, in other words those that are most relevant and significant to your research and the associated recognised



Box 3.15 Focus on student research

Developing a thematic analysis grid*

Sam's research project required him to draw together academic research on data privacy in marketing.

Having undertaken his literature search and made brief notes on all those items he considered likely to be relevant, she began to compile her Thematic Analysis Grid. An extract of Sam's grid follows:

| Authors | Theme 1: Data usage outcomes | Theme 2: Privacy | Theme 3: Why do consumers share information in practice? | Theme 4: How to mitigate concerns | Comments on Meth- odology/ Context |
|--|--|---|---|---|--|
| Aguirre, E., Roggeveen, A., Grewal, D. and Wetzels, M. (2016) The personalization–privacy paradox: implications for new media. <i>Journal of Consumer Marketing</i> , 33(2), pp. 98–110. | Leveraging data can: a. create deep connections; b. increase attention to the ad; c. improve response rates Data can be collected overtly (e.g. fill in form) or covertly (click stream history) When firm uses covertly collected data, concerns raised re manipulative intentions and loss of control Leads to decrease in trust and willingness to engage and increases scepticism and ad avoidance Negative consumer reactions – providing false information, negative word of mouth, seeking stricter regulatory controls | Privacy is a commodity with a quantifiable value Authors identified conflict: one study shows consumers pay higher prices from sites offering more privacy Another study: 2 gift cards – more valuable if permission given to track BUT preferences reflected which card was offered first Difficult for consumers to know cost-benefit trade-off if info collected covertly Privacy concerns are situational, e.g. appearance of web site; no. of others who have given info | Consumers weigh up risks re loss of privacy against relevant offers and discounts | Grant consumers some control over privacy This is difficult to exert and time-consuming Increase consumer trust strategically, e.g. explicit privacy policies Offer third party privacy seals Transform the firm into a social entity | Literature review – paper seeks to identify factors that determine how consumers respond to personalised communications. No detailed methodology of how literature search conducted. Useful summary table in appendix showing summary findings from c 50 articles from 2003–2015 (NB given the topic, early papers could be out of date) |

*Abridged from Case 3 by Deborah Anderson in the 8th edition.

| Authors | Theme 1: Data usage outcomes | Theme 2: Privacy | Theme 3: Why do consum- ers share information in practice? | Theme 4: How to mitigate concerns | Comments on Meth- odology/ Context |
|---|---|---|---|---|--|
| Zhu, Y. and Chang, J. (2016) The key role of relevance in personalized advertisement: Examining its impact on perceptions of privacy invasion, self-awareness, and continuous use intentions. <i>Computers in Human Behavior</i> , 65, pp. 442–447. | Note that 'privacy' is investigated from a number of perspectives, e.g. perceived control, financial compensation, regulation, monetary rewards, convenience Privacy invasion perception is negatively related to continuous use intention Claim: privacy is a commodity that can be exchanged for perceived benefits | If better service, compensation or discounts are larger than perceived risks then consumers more willing to disclose personal information | Relevance of ad influences reaction: 'Relevant personalised advertising reduces privacy invasion perceptions and increases continuous use intentions' (p. 443) 'By accurately providing content that fits into users' interests and tastes, privacy invasion concerns are alleviated by the privacy calculus' p. 446 Relevant ads save time and resources | Relevance of ad influences reaction: 'Relevant personalised advertising reduces privacy invasion perceptions and increases continuous use intentions' (p. 443) 'By accurately providing content that fits into users' interests and tastes, privacy invasion concerns are alleviated by the privacy calculus' p. 446 Relevant ads save time and resources | On-line survey in Taiwan. 386 responses from 1,000 (incentivised) questionnaires. Investigated relevance, perceptions of privacy, self-awareness and usage intentions. Drew on rational-choice theory and self-awareness theory. Used factor analysis and Structural Equation Modelling. Used fake website (NB real world?) |

experts. You will also be able to see how themes have developed over time due to the articles being in date order. The Grid's final column will also allow you to see easily which methodologies have been used in the research. Because your reading is brought together in one place, you will have a clear overview about what is known about your research topic from which to begin drafting and re-drafting your review.

Remember to be critical as you draft your review (Box 3.1) and ensure that what you write relates clearly to your research question(s) and objectives (Box 3.2). In order to improve the transparency of your review process, you should also explain precisely how you selected the literature you have included in your review, outlining your choice of search terms and of databases used. This is usually done at the start of the review and is essential if you are using the systematic review methodology (Section 3.9). This can be thought of as 'Step 0' of the review funnel we outlined in Section 3.3. When you have completed your first draft you can use Box 3.16 to evaluate its suitability for your project report.



Box 3.16 Checklist

Evaluating your draft literature review

- ✓ Does your review have a clear title, which describes the focus of your research rather than just saying 'literature review'?
- ✓ Have you explained precisely how you searched the literature and criteria used to select those studies included?
- ✓ Does your review start at a more general level before narrowing down?
- ✓ Is your review organised thematically around the ideas contained in the research being reviewed rather than the researchers?
- ✓ Are your arguments coherent and cohesive – do your ideas link in a way that will be logical to your reader and tell a clear story?
- ✓ Have you used subheadings within the review to help guide your reader?
- ✓ Does the way you have structured your review draw your reader's attention to those issues that are the focus of your research, in particular your research questions and objectives?
- ✓ Does your review lead your reader into subsequent sections of your project report?

3.11 A note about plagiarism

There is no doubt that plagiarism has become an enormous concern in academic institutions in recent years, largely as a result of the ease with which material can be copied from the Internet and passed off as the work of the individual student. It is a serious topic because it is a breach of academic integrity when a person passes off another's work as their own. The consequences of being found guilty of plagiarism can be severe, including not being awarded your degree.

Neville (2016) argues that plagiarism is an issue that runs parallel to a debate with recurring questions about the purpose of higher education in the twenty-first century. He notes that, on the one hand, there is the argument that an insistence on 'correct' referencing is supporting a system and a process of learning that is a legacy of a different time and society. This argument holds that universities are enforcing upon you an arcane practice of referencing that you will probably never use again outside higher education. On the other hand, there is the argument that plagiarism is an attack upon values of ethical, proper, decent behaviour – values consistent with a respect for others. These are ageless societal values that need to be maintained.

So what precisely is plagiarism? Quite simply, it is presenting someone else's work or ideas as if they are your own, with or without their consent and failing to fully acknowledge the original source. The University of Oxford (2021) lists eight forms of plagiarism that are commonly found in universities. These are:

- 1 *Quoting* someone else's work, word for word, without clear acknowledgement.
- 2 *Cutting and pasting* text, diagrams or any other material from the Internet without acknowledgement.
- 3 *Paraphrasing* someone else's work by altering a few words or changing their order or closely following the structure of their argument without acknowledgement.
- 4 *Colluding* including unauthorised collaboration with others (unless expressly asked to do so such as in group work) and not attributing the assistance received.
- 5 *Inaccurately referencing (citing)*, within the text and list of references, of the source of a quoted passage. This often occurs when students pretend to have read an original source, when their knowledge is derived from a secondary source.

- 6 *Failing to acknowledge assistance* that leads to substantive changes in the content or approach.
- 7 *Using materials written by others* such as professional essay writing services, or friends, even with the consent of those who have written it.
- 8 *Auto or self-plagiarising*, that is submitting work that you have already submitted (either in part or fully) for another assessment. However, it is usually acceptable to cite earlier work you have had published.

It is tempting to think that all cases of plagiarism are a consequence of students either being too idle to pursue their research and write diligently or wishing to appear cleverer than they really are. But the fact is that plagiarism is an extremely complex issue and the reasons for it may owe as much to student confusion as wilful negligence. That said, there is little excuse for confusion. All universities have ample guidance for students on the topic of plagiarism and will emphasise that it is the responsibility of the individual student to become aware of the university's regulations surrounding its conduct. In addition, an increasing number of universities ask students to check their own work using plagiarism detection software such as Turnitin and submit the report alongside the electronic copy of their work.

3.12 Summary

- Critically reviewing the literature is necessary to help you to develop a thorough understanding of, and insight into, previous work that relates to your research question(s) and objectives.
- Your written review will set your research in context by discussing critically and referencing work that has already been undertaken, drawing out key points and presenting them in a logically argued way, and highlighting those areas where you will provide fresh insights. It will lead the reader into subsequent sections of your project report.
- There is no one correct structure for a critical review, although it is helpful to think of it as a funnel in which you start at a more general level prior to narrowing down to your specific research question(s) and objectives.
- You are most likely to make use of formally published items (white literature) and those which have not been through the peer review process (grey literature). Your use of these resources will depend on your research question(s) and objectives. Some may use only secondary literature. For others, you may need to locate grey literature as well.
- When planning your literature search you need to:
 - have a clearly defined research question(s) and research objectives;
 - define the parameters of your search;
 - generate search terms and phrases.
- Techniques to help you in this include discussion, brainstorming and relevance trees.
- Your literature search is likely to be undertaken using a variety of approaches in tandem. These will include:
 - searching online databases;
 - obtaining relevant literature referenced in books and articles you read;
 - browsing and scanning secondary in your university library;
 - general online searching;
 - searching institutional repositories and social networking platforms;

- Don't forget to make precise notes of the search processes you have used and their results.
- Once obtained, the literature must be evaluated for its relevance and value to your research question(s) and objectives. Each item must be read and noted. Bibliographic details, a brief description of the content and appropriate supplementary information should also be recorded.
- For literature reviews focusing on policy or practice questions in particular, you may decide to use a systematic review.
- Care should be taken when drafting and redrafting your literature review not to plagiarise the work of others.

Self-check questions

Help with these questions is available at the end of the chapter.

- 3.1** The following extract and associated references are taken from the first draft of a critical literature review. The research project was concerned with the impact of changes to UK legal aid legislation on motor insurance pricing policies.

List the problems with this extract in terms of its:

- a content;
- b structure.

The primary function of motor insurance is to provide financial protection against damage to vehicles and bodies resulting from traffic conditions and the liabilities that can arise (Wikipedia 2021). O'Brian (2014) suggests that motor insurers have been too eager to reap the benefits of legal aid. Papra-Servano (2013) notes that the average car insurance premium has reduced since changes in legislation brought about by the UK Legal Aid, Sentencing and Punishment of Offenders Act. This Act prohibits the payment and receipt of referral fees in relation to personal injury claims by solicitors, claims companies and other authorised persons. Motor insurance is particularly price sensitive because of its compulsory nature and its perception by many to have no real 'value' to themselves.

O'Brien, S. (2013). 'Motor insurance: Jumping the gun'. Post. 29 October. Available at: <https://www.postonline.co.uk/post/analysis/2301953/motor-insurance-jumping-the-gun>. [Accessed 28 April 2021]

Papra-Servano, C. (2013). 'Rates drop as motor insurers anticipate legal reform windfall'. Post. 17 July. Available at: <https://www.postonline.co.uk/post/news/2282883/rates-drop-as-motor-insurers-anticipate-legal-reform-windfall> [Accessed 28 April 2021]

Wikipedia (2018) *Vehicle Insurance*. Available at: http://en.wikipedia.org/wiki/Vehicle_insurance. [Accessed 28 March 2021].

- 3.2** Outline the advice you would give a colleague on:

- a how to plan their search;
- b which literature to search first.

- 3.3** Brainstorm at least one of the following research questions, either on your own or with a colleague, and list the search terms that you have generated.

- a How effective are share options as a motivator?
- b How do the opportunities available to a first-time house buyer through interpersonal discussion influence the process of selecting a financial institution for the purposes of applying for a house purchase loan?
- c To what extent do new methods of direct selling of financial services pose a threat to existing providers?

- 3.4** You are having considerable problems with finding relevant material for your research when searching databases. Suggest possible reasons why this might be so.
- 3.5** Rewrite the following passage as part of a critical literature review using the Harvard system of referencing:

Past research indicates important gender differences in the use of networks,¹ and suggests that male SME owners are more likely to successfully network and benefit from networks-driven performance in contrast to female SME owners.² In particular, as many women come to self-employment from domestic or non-management background,³ and thus are likely to have previously engaged in the relatively isolating domestic and childrearing work or lower-status support work, they can be expected to possess fewer, more personal, less formal and less powerful contacts, as well as less time for networking.^{1,4,5,6,7} However, empirical studies of SME owners challenge such expectations. For example, according to a 2012 survey of 2919 male- and 181 female-controlled SMEs, there is little gender difference, after controlling for education, experience, industry, age and size, in the SME owners' use of networking and its impact on business performance – in other words, SMEs owned by women and men enjoy similar performance benefits of networking.⁸

¹Hanson, S. and Blake, M. (2009) 'Gender and entrepreneurial networks', *Regional Studies*, Vol. 43, pp. 135–49.

²Watson, J. (2012) 'Networking: Gender differences and the association with firm performance', *International Small Business Journal*, Vol. 30, pp. 536–58.

³Cromie, S. and Birley, S. (1992) 'Networking by female business owners in Northern Ireland', *Journal of Business Venturing*, Vol. 7, pp. 237–51.

⁴Ardrich, H. (1989) 'Networking among women entrepreneurs', in O. Hagan, C.S. Rivchun and D. Sexton (eds) *Women-Owned Businesses*. New York: Praeger, pp. 103–32.

⁵Moore, G. (1990) 'Structural determinants of men's and women's personal networks', *American Sociological Review*, Vol. 55, pp. 726–35.

⁶Munch A., McPherson J.M. and Smith-Lovin L. (1997) 'Gender, children, and social contact: The effects of childrearing for men and women', *American Sociological Review*, Vol. 62, pp. 509–20.

⁷Orhan, M. (2001) 'Women business owners in France: The issue of financing discrimination', *Journal of Small Business Management*, Vol. 39, pp. 95–102.

⁸Watson, J. (2012) 'Networking: Gender differences and the association with firm performance', *International Small Business Journal*, Vol. 30, pp. 536–58.

Review and discussion questions

- 3.6** Use the specialised search services such as 'Google Scholar' and 'Google Finance' to search for articles on a topic which you are currently studying as part of your course.
- Make notes regarding the types of items that each of these services finds.
 - How do these services differ?
 - Which service do you think is likely to prove most useful to your research project?
- 3.7** Agree with a friend to each review the same article from a refereed academic journal which contains a clear literature review section. Evaluate independently the literature review in your chosen article with regard to its content, critical nature and structure using the checklists in Boxes 3.2 and 3.15 respectively. Do not forget to make notes regarding your answers to each of the points raised in the checklists. Discuss your answers with your friend.
- 3.8** With a friend select an article that you think will be of use to an assignment you are both currently working on. Use the checklist in Box 3.13 to assess the relevance and value of the article to your assignment.



Progressing your research project

Critically reviewing the literature

- Consider your research question(s) and objectives. Use your lecture notes, course textbooks and relevant review articles to define both narrow and broader parameters of your literature search, considering language, subject area, business sector, geographical area, publication period and literature type.
- Generate search terms using one or a variety of techniques such as reading, brainstorming and relevance trees. Discuss your ideas widely, including with your project tutor and colleagues.
- Use online databases to identify relevant secondary literature. At the same time, obtain

relevant literature that has been referenced in the books and articles you have already read. Do not forget to record your searches systematically and in detail.

- Obtain copies of items, evaluate them systematically and make notes. Remember also to record bibliographic details, a brief description of the content and supplementary information in your bibliographic software.
- Construct a Thematic Analysis Grid and start drafting your critical review as early as possible, keeping in mind its purpose and taking care to reference properly and avoid plagiarism.
- Continue to search the literature throughout your research project and redraft your review to ensure that your review remains up to date.
- Use the questions in Box 1.4 to guide your reflective diary entry.

References

- Access to Research (2021) *Access to research*. Available at <http://www.accesstoresearch.org.uk/> [Accessed 5 March 2021].
- Adams, R.J., Smart, P. and Huff, A.S. (2017) 'Shades of grey: Guidelines for working with the grey literature in systematic reviews for management and organizational studies', *International Journal of Management Reviews*, Vol. 19, pp. 432–54.
- Anderson, D., Lees, R. and Avery, B. (2015) Reviewing the literature using the Thematic Analysis Grid. *Poster presented at the European Conference on Research Methodology for Business and Management Studies*, Malta, June, 2015.
- Bell, J. and Waters, S. (2018) *Doing Your Research Project* (7th edn). Maidenhead: Open University Press.
- Creswell, J.W. and Poth, C.N. (2017) *Qualitative Inquiry and Research Design: Choosing Among Five Approaches* (4th edn). Thousand Oaks, CA: Sage.
- Darabi, F., Saunders M.N.K. and Clark, M. (2020, online first) 'Trust initiation and development in SME-University collaborations: implications for enabling engaged scholarship' *European Journal of Training and Development*, pp.1–26.
- Dees, R. (2003) *Writing the Modern Research Paper* (4th edn). Boston, MA: Allyn and Bacon.
- Denyer, D. and Tranfield, D. (2009) 'Producing a Systematic Review', in D.A. Buchanan and A. Bryman (eds) *The Sage Handbook of Organisational Research Methods*. London: Sage, pp. 671–89.
- Fisher, C. (2010) *Researching and Writing a Dissertation for Business Students* (3rd edn). Harlow: Financial Times Prentice Hall.
- Hart, C. (2018) *Doing a Literature Review* (2nd edn). London: Sage.
- Harvard College Library (2020) *Interrogating Texts: 6 Reading Habits to develop in your first year at Harvard*. Available at: <https://guides.library.harvard.edu/sixreadinghabits> [Accessed 11 April 2021].

- Harzing, A. W. (2020) *Journal Quality List*. Available at: <https://harzing.com/resources/journal-quality-list> [Accessed 11 April 2021].
- Hiebl, M.R.W. (2021, online first) 'Sample selection in systematic literature reviews of management research', *Organizational Research Methods*, pp. 1–33.
- Isaeva, N., Gruenewald, K. and Saunders M.N.K. (2020) 'Trust theory and customer services research: Theoretical review and synthesis', *The Service Industries Journal*, Vol. 40, pp. 1031–63.
- Jankowicz, A.D. (2005) *Business Research Projects* (4th edn). London: Thomson Learning.
- Jones, O. and Gatrell, C. (2014) 'Editorial: The future of writing and reviewing for IJMR', *International Journal of Management Reviews*, Vol. 16, pp. 249–64.
- Mergen, A. and Ozbilgin, M.F. (2021) 'Understanding followers of toxic leaders: Toxic illusion and personal uncertainty', *International Journal of Management Reviews*, Vol. 23, No. 1, pp. 45–63.
- MindGenius (2021) *MindGenius*. Available at <https://www.mindgenius.com/> [Accessed 21 March 2021].
- Mingers, J. (2000) 'What is it to be critical? Teaching a critical approach to management undergraduates', *Management Learning*, Vol. 31, No. 2, pp. 219–37.
- Moher, D., Liberati, A., Tetzlaff, J. and Altman, D.G. (2009) 'Preferred reporting for systematic reviews and meta-analyses: The PRISMA statement', *British Medical Journal (BMJ)*, No. 338, b2535. Available at: www.bmjjournals.org/content/339/bmj.b2535.full?view=long&pmid=19622551 [Accessed 27 March 2011].
- Neville, C. (2016) *The Complete Guide to Referencing and Plagiarism* (3rd edn). Maidenhead: Open University Press.
- Petticrew, M. and Roberts, H. (2006) *Systematic Review in the Social Sciences: A Practical Guide*. Malden, MA: Blackwell.
- Ridley, D. (2018) *The Literature Review: A Step-by-Step Guide for Students* (2nd edn). London: Sage.
- Rojon, C., Okupe, A. and McDowall, A. (2021) 'Utilization and development of systematic reviews in management research: What do we know and where do we go from here?' *International Journal of Management Reviews*, Vol. 23, No. 2, pp. 191–223.
- Sharp, J.A., Peters, J. and Howard, K. (2002) *The Management of a Student Research Project* (3rd edn). Aldershot: Gower.
- Siddaway, A.P., Wood, A.M. and Hedges, L.V. (2019) 'How to do a systematic review: A best practice guide for conducting and reporting narrative reviews, Meta-analyses, and meta-syntheses', *Annual Review of Psychology*, Vol. 70, pp. 747–70.
- SimpleMind (2021) *SimpleMind*. Available at: <https://simplemind.eu/> [Accessed 21 March 2021].
- Synder, H. (2019) 'Literature review as a research methodology: An overview and guidelines', *Journal of Business Research*, Vol. 104, pp. 333–9.
- Tranfield, D., Denyer, D. and Smart, P. (2003) 'Towards a methodology for developing evidence-informed management knowledge by means of systematic review', *British Journal of Management*, Vol. 14, No. 3, pp. 207–22.
- University of Oxford (2021) *Study Skills and Training: Plagiarism*. Available at <https://www.ox.ac.uk/students/academic/guidance/skills/plagiarism?wssl=1> [Accessed 20 April 2021].
- University of Southern California (2021) *Organising Your Social Sciences Research Paper: 5. The Literature Review*. Available at <https://libguides.usc.edu/writingguide/literaturereview> [Accessed 26 February 2021].
- Wallace, M. and Wray, A. (2016) *Critical Reading and Writing for Postgraduates* (3rd edn). London: Sage.
- Wikipedia (2021) *Wikipedia: General Disclaimer*. Available at https://en.wikipedia.org/wiki/Wikipedia:General_disclaimer [Accessed 20 March 2021].

Further reading

- Colquitt, J.A. (2013) 'Crafting references in AMJ submissions', *Academy of Management Journal*, Vol. 56, No. 5, pp. 1221–4. This short editorial provides extremely useful insights regarding citations and references in academic writing.
- Neville, C. (2016) *The Complete Guide to Referencing and Plagiarism* (3rd edn). Maidenhead: Open University Press. Chapter 6 is a very helpful guide on what constitutes plagiarism and how it can be avoided. The chapter ends with some useful exercises designed to ensure that the reader does not fall into some common traps.
- Ridley, D. (2018) *The Literature Review: A Step-by-Step Guide for Students* (2nd edn). London: Sage. This has numerous examples and offers a wealth of practical advice with practice tasks. The sections on being critical and on foregrounding writers' voice are particularly helpful.
- Siddaway, A.P., Wood, A.M. and Hedges, L.V. (2019) 'How to do a systematic review: A best practice guide for conducting and reporting narrative reviews, Meta-analyses, and meta-syntheses' *Annual Review of Psychology*, Vol. 70, pp. 747–70. This practical guide outlines how to plan, conduct and present both qualitative and quantitative systematic reviews, outlining core principles and describing common problems.

Generating a Research Idea & Developing your Research Proposal

Saunders et al. (2023; 28-66)

Chapter 2



Generating a research idea and developing your research proposal

Learning outcomes

By the end of this chapter you should be able to:

- identify the characteristics of a good research idea;
- generate your own research ideas;
- refine your research ideas;
- express your research idea as a clear overarching research question, a research aim and research objectives or investigative research questions;
- recognise the relationship between the overarching research question, research aim and research objectives or investigative research questions;
- recognise the role of theory in developing the overarching research question, a research aim and research objectives or investigative research questions;
- develop a research proposal that outlines your proposed research project.

2.1 Introduction

Many students think that researching an idea of their choice is the most exciting part of their course. It is something that you get to decide for yourself rather than having to complete a task decided by your tutors. We will stress in this chapter that it is important to choose something that will sustain your interest throughout the months that you will need to complete it. You may even decide to do some research on something that forms part of your leisure activities!

Before you start your research, you need to have at least some idea of what you want to do. This is probably the most difficult, and yet the most important, part of your research project.



Research ideas, contemplative moods and mindfulness

Legend tells us that the law of gravity was prompted into Isaac Newton's mind when he was sitting in contemplative mood under an apple tree and was hit on the head by a falling apple. In reality it appears that, although sitting in an orchard under an apple tree and witnessing a falling apple, Newton was never hit on the head (Nix 2018). Rather this observation, while relaxing and drinking tea under the shade of some apple trees, caused him to contemplate why apples (and other objects) always fell straight downwards. He had an idea!

For most of us, coming up with a research idea is difficult. Fortunately, numerous articles on creativity and the generation of ideas emphasise that we are not alone whether we are generating ideas for research or some other purpose. An article in the *Harvard Business Review* (Schootstra et al. 2017) recognises this, noting that it is hard to keep on having great ideas and posing the question 'What do you do when you run out of good ideas? How do you "get your mojo back"?' Observing that an increasingly popular solution to this is mindfulness meditation, the authors



Malvern Link Common © 2021 Mark NK Saunders

both reviewed the literature and undertook their own research.

Among other findings, the researchers revealed that, for most participants, practising mindfulness helped them clear their minds, focus better, and come up with a wider range of ideas, including original solutions to problems. Similar to Isaac Newton's relaxed contemplation in the peace of the orchard, they found that mindfulness meditation can help enhance creativity and innovation by improving attention and making it easier to recognise both the novelty and usefulness of ideas that come into our minds.

Up until now, most of your studies will probably have been concerned with answering questions that other people have set. The start of this chapter is intended to help you generate your own research idea and decide the direction of your research journey. If you are not clear about what you are going to research, it will be difficult to plan how you are going to research it. This reminds us of a favourite quote in *Alice's Adventures in Wonderland*. This

is part of Alice's conversation with the Cheshire Cat. In this Alice asks the Cat (Carroll 1989: 63–4):

'Would you tell me, please, which way I ought to walk from here?'
'That depends a good deal on where you want to get to', said the Cat.
'I don't much care where', said Alice.
'Then it doesn't matter which way you walk', said the Cat.

Generating a research idea is unlikely to involve you in a single moment of inspiration. Even if, like Isaac Newton, inspiration arrives while you are relaxing and contemplating, developing this is still likely to be iterative, involving you in a process of formulating, clarifying and re-formulating your research idea(s) until it becomes an acceptable and focused overarching research question. Part of this process will involve refining your initial overarching research question so that it is more focused with a related research aim and set of research objectives. Once you have focused your overarching research question, and can operationalise it through your aim and objectives, you will be better able to choose the most appropriate research strategy and data collection and analysis techniques, and develop your research proposal. This process will be time-consuming, and will probably take you up blind alleys (Saunders and Lewis 1997). However, without spending time on this, you are far less likely to achieve a successful project. This is likely to be the case even when you have been given an embryonic research idea, perhaps by an organisation or a tutor. Whether you start with such a research idea or formulate one for yourself, it is also important to work on something that will sustain your interest throughout the months that you will need to complete it.

We commence this chapter by looking at the characteristics of good research ideas (Section 2.2). We then consider techniques for generating (Section 2.3) and refining (Section 2.4) research ideas. Taking your research idea(s) and chosen topic and developing a research proposal will involve a number of tasks. You will need to develop an overarching research question(s), ensuring it is focused (Section 2.5), and express this as a research aim and a set of research objectives (Section 2.6). You will also need to consider how your research topic fits into the existing theory in the literature. Using theory may help you to clarify your research topic. It will also inform your research question, aim and objectives, and your research proposal more generally. We discuss these aspects in Section 2.7. Before you commence the research to meet your aim and objectives, you will be expected to produce a structured plan in the form of a research proposal. We discuss the rationale and a structure for the research proposal in Sections 2.8 and 2.9, respectively. Once approved, your research proposal will act as the guide for the rest of your research project.

2.2 Characteristics of good research ideas

The attributes of a business and management research idea do not vary a great deal between universities, although there will be differences in the emphasis placed on these attributes. Some of these characteristics reflect the need to fulfil the specification set for the research project and meet the assessment criteria. We outline these under the heading, 'Appropriateness'. Other characteristics reflect the feasibility of the idea being operationalised and the associated developmental opportunities. We consider some of these briefly under the headings of 'Capability' and 'Fulfilment'. There may be other characteristics of a good research idea that become evident to you in relation to your own research project. Identifying these should be helpful in terms of generating a research idea and developing your research proposal. You may also find it useful to discuss them with your project tutor.

Appropriateness

The scope of the research idea you generate and the nature of the research proposal that you develop will need to meet the requirements of your examining body (such as your university, professional body or other accredited organisation). This means that you must generate your research idea and develop your research proposal with care. For example, some universities require students to collect their own data as part of their research project, whereas others allow them to base their project on data that have already been collected. Alternatively, some ask you to undertake an organisation-based piece of applied research, while others simply say that it must be within the subject matter of your course or programme. You therefore need to check the assessment criteria for your research project and ensure that your idea and the plan you propose will enable you to meet these criteria. If you are unsure, you should discuss any uncertainties with your project tutor.

It will be important to use existing theory from the academic literature to inform your research idea and in the development of your research proposal. As part of your assessment criteria you are almost certain to be asked to consider the theoretical context of your research in your research proposal. As we discussed earlier, using existing theory should help you to clarify your research idea and to inform your overarching research question, aim and objectives. Using theory should also help you to develop clear definitions of the concepts that you use in your research (Podsakoff et al. 2016). We consider the role of theory further in Section 2.7 and the critical review of the literature that discusses it in Section 3.3.

Most project tutors will argue that one of the characteristics of a good research proposal is a clearly defined research question(s), aim and set of objectives or investigative questions (Section 2.4). These will, along with a good knowledge of the literature (Chapter 3), enable you to assess the extent to which your research is likely to provide fresh insights into the topic. Many students believe this is going to be difficult. Fortunately, there are numerous ways in which such insight can be defined as new (Sections 2.4 and 2.5). We offer a word of warning, an idea offering new insights is far more likely than a completely new research idea. If you come up with a research idea that no-one has ever thought of or researched and is therefore new, there are at least three possibilities:

- Your idea is new and exciting and no one has ever thought of it, let alone researched it.
- Your idea has been researched before and you have been searching for literature about it in the wrong places.
- Your idea has no worth and that is why it has not been researched up to now.

We therefore recommend you consider any such potential ideas extremely carefully, before claiming the idea is both completely new and worthy of researching.

It is also important that your research will have **symmetry of potential outcomes**: that is, your findings will be of similar value whatever you find out (Gill and Johnson 2010). Without this symmetry you may spend a considerable amount of time researching, only to find an answer of little importance. Whatever the outcome, you need to ensure you have the scope to write an interesting project report.

Capability

Your research must also be something you are capable of undertaking. Capability can be considered in a variety of ways. At the personal level you need to feel comfortable that you have, or can develop, the skills that will be required to undertake the research. We hope that you will develop your research skills as part of undertaking your project, such as those related to data analysis. However, some skills, for example, learning a new foreign language, may be impossible to acquire in the time you have available.

Your ability to find the financial and time resources to undertake the research will also affect your capability. This relates, in part, to the concept of feasibility (which we return to in Section 2.8 and also discuss in Section 6.2). Some research projects are unlikely to be possible to complete in the time allowed by your course of study. This may be because they require you to measure the impact of an intervention over a long time period or because of their complexity. Similarly, ideas that are likely to require you to travel widely or use expensive equipment or specialist software not available at your university should also be disregarded unless financial resources permit.

Capability also means you must be reasonably certain of gaining access to any data you might need to collect. Many people start with ideas where access to data will prove difficult. Certain, more sensitive topics, such as financial performance or decision-making by senior managers, are potentially fascinating. However, these may present considerable access problems. You should, therefore, discuss this with your project tutor after reading Sections 6.2–6.4.

Fulfilment

Your research idea needs to be one that excites your imagination and in which you have or will develop a genuine interest. Most research projects are undertaken over at least a four-month period. An idea in which you are only vaguely interested at the start is likely to become one in which you have no interest and with which you will fail to produce your best work. It may also be important to consider your future aspirations. If you wish to obtain employment or pursue a career in a particular subject area, it is sensible to use this opportunity to start to develop some expertise in it.

It is almost inevitable that the extent to which these characteristics apply to you will depend on your research idea and the reasons why you are undertaking the research. However, most are likely to apply. For this reason, it is important that you check and continue to check any potential research idea against the summary checklist contained in Box 2.1.

Box 2.1 Checklist

Characteristics of a good research idea

Appropriateness

- ✓ Does the idea fit the specifications and meet the standards set by the examining institution?
- ✓ Does the idea contain issues that have a clear link to theory?
- ✓ Are you able to state an overarching research question(s), research aim and research objectives or investigative questions clearly?
- ✓ Will the proposed research be likely to provide fresh insights into this topic?
- ✓ Are the findings likely to be symmetrical: that is, of similar value whatever the outcome?

Capability

- ✓ Do you have, or can you develop within the project time frame, the necessary research skills to undertake the research?
- ✓ Is the research achievable within the available time?
- ✓ Is the research achievable within the financial resources that are likely to be available?
- ✓ Are you reasonably certain of being able to gain access to data you are likely to require for this research?

Fulfilment

- ✓ Does the research idea really interest and motivate you?
- ✓ Will the research help towards the achievement of your future aspirations or career goals?

2.3 Generating research ideas

Many business and management students are expected to generate and refine their own research ideas, whereas some others, particularly those on professional and post-experience courses, are provided with an embryonic research idea by their employing or sponsoring organisation. In the initial stages of their research they are expected to refine this to a clear and feasible idea that meets the requirements of the examining organisation. If you have already been given a research idea, we believe you will still find it useful to read this section as many of the techniques can also be used for refining research ideas.

As outlined in the opening vignette, ideas can come to us at anytime and anywhere; recognising them is made easier if our minds are clear. In addition, if you have not been given a research idea, there are a variety of techniques that can be used to generate one you would like to research. These can be divided into two groups: those that predominantly involve **rational thinking techniques**, being based on a systematic approach and those that involve more **creative thinking techniques** favouring individual preferences or spontaneous ideas (Table 2.1).

The precise techniques you choose to use, and the order in which you use them, are entirely up to you. However, by using one or more creative techniques, you are more likely to ensure that your heart as well as your head is in your research project. In our experience, it is usually better to use both rational and creative techniques. In order to do this, you need to have some understanding of the techniques and the ways in which they work. We therefore list the techniques in Table 2.1 and then discuss possible ways they might be used to generate research ideas. These techniques will generate one of two outcomes:

- one or more possible project ideas that you might undertake;
- few ideas that relate to your interests. In this case you may want to revise the area in which you are interested, either by choosing another area or by refining and perhaps narrowing or widening your original area of interest.

In either instance, we suggest that you make some notes and arrange to talk to your project tutor.

Rational thinking techniques

Examining your own strengths and interests

It is important that you select an idea in which you are likely to do well and, if possible, already have some academic knowledge. One way of doing this is to look at completed assignments for which you have received good grades (Box 2.4). Many, if not most, of these assignments are likely to be in subject areas in which you have an interest. These

Table 2.1 More frequently used techniques for generating (and refining) research ideas

| Rational thinking | Creative thinking |
|---|--|
| Examining your own strengths and interests | Keeping a notebook of your ideas |
| Examining academic staff research interests | Exploring personal preferences using past projects |
| Looking at past project titles | Exploring relevance to business of ideas in the literature |
| Discussion | Relevance trees |
| Searching existing literature | Brainstorming |
| Scanning the media | |

assignments will provide you with subject areas in which to search and find a research idea. In addition, you may, as part of your reading, be able to focus more precisely on the sort of ideas about which you wish to conduct your research.

As noted in Section 2.2, there is the need to think about your future. If you plan to work in financial management, it would be sensible to choose a research project in the financial management field. One part of your course that will inevitably be discussed at any job interview is your research project. A project in the same field will provide you with the opportunity to display clearly your depth of knowledge and your enthusiasm.

Examining academic staff research interests

Your university's website will have profile pages of academic staff, which may be helpful in exploring and generating research ideas that could be of interest for your own project. These pages usually outline the subject area or areas taught by each member of staff (e.g. accounting, international management, marketing, responsible business), and are also likely to list their particular research interests (e.g. regulation of accounting standards, transnational management, pricing and price promotions, organisational learning). In many cases, academic staff provide short commentaries on their research interests offering more insights. Lists of publications and conference papers with hyperlinks to copies may also be included. Working through this information may allow you to generate ideas for your own research and guide you to some initial reading to test this interest.

Looking at past project titles

Many of our students have found looking at past projects a useful way of generating research ideas. For undergraduate and taught master's degrees these are often called **dissertations**. For research degrees they are termed **theses**. A common way of doing this is to scan a list of past project titles for anything that captures your imagination. Titles that look interesting or that grab your attention should be noted, as should any thoughts you have about the title in relation to your own research idea. In this process the fact that the title is poorly worded or the project report received a low mark is immaterial. What matters is the fact that you have found an idea that interests you. Based on this you can think of new ideas in the same general area that will enable you to provide fresh insights.

Scanning actual research projects may also produce research ideas. However, you need to beware. Just because a project is in your library is no guarantee of the quality of the arguments and observations it contains. In some universities all projects are placed in the library whether they are bare passes or distinctions.

Discussion

Colleagues, friends and university tutors are all potentially good sources of possible research ideas. Often project tutors will have ideas for possible student projects, which they will be pleased to discuss with you.

Ideas can also be obtained by talking to people who work in, or have direct experience of, the topic area in which you are interested to develop a research idea. People who have experience of a topic area may include managers and other practitioners such as accountants, business analysts, marketing executives, human resource administrators, purchasing or sales staff as well as others. Entrepreneurs and small business owners may be useful to talk to, and members of professional groups or workplace representatives may also provide you with insights that help to generate research ideas. Your contact with such people at this early stage may be fortuitous, relying on being able to talk to someone you already know such as those in an organisation in which you have undertaken a work placement.

It is important that as well as discussing possible ideas you also make a note of them. What seemed like a good idea in the coffee shop may not be remembered quite so clearly after the following lecture!

Searching existing literature

As part of your discussions, relevant literature may also be suggested. There are various types of literature that are of particular use for generating research ideas. These include:

- articles in peer-reviewed academic journals;
- articles in trade and professional journals;
- reports;
- books and e-books.

Peer-reviewed academic journal articles nearly always contain a section that reviews literature relevant to the article's topic area. Given the nature of published research, such articles are generally highly specialised, focusing on a particular aspect of a management subject. You will need to be prepared to undertake an extensive search lasting some hours (or even days) to find articles that might be helpful in generating research ideas related to your broader topic of interest. The (advanced) search facilities available to you through your university library's or professional association's databases and search interfaces will be very helpful here (Sections 3.4 to 3.6). You may also consider signing up to and using one or more of the social networking platforms used by academics to share their research papers such as Academia.edu and ResearchGate. Although these copies are unlikely to be the final version for copyright reasons, they can provide access to those articles, reports and conference papers that are not available through those databases for which your university's library has subscriptions.

Browsing journals and using available search facilities should help you to identify possible research ideas and potential topics.

Of particular use are academic **review articles**. Some journals such as the *International Journal of Management Reviews* only publish review articles – so look out for these! These articles contain a considered review of the state of knowledge in a particular topic area and are therefore likely to contain a wealth of ideas about that area (Box 2.2). These ideas will act as pointers towards aspects where further research needs to be undertaken.

For many subject areas your project tutor will be able to suggest recent review articles, or articles that contain recommendations for further research. Journal articles reporting research often also indicate where further research is needed. Both may be phrased in the form of questions and, even if not, may suggest possible questions for you. Reports may also be of use. The most recently published are usually up to date and, again, often contain recommendations that may form the basis of your research idea. Books by contrast may be less up to date than other written sources. They often, however, contain a good overview of research that has been undertaken, which may suggest ideas to you.

Alvesson and Sandberg (2011) report that articles published in academic management journals are predominantly based on research that finds new ways to investigate existing theoretical perspectives. They call this approach 'gap spotting', suggesting it results in incremental changes in theory. They identify a more critical and reflexive but rarer approach to research that challenges the assumptions underpinning existing theoretical perspectives and that has the potential to lead to more interesting and high-impact theories. Given the difficulties associated with designing an assumption-challenging study, it is much more likely that you will adopt the 'gap spotting' approach. We discuss this further when we consider the importance of theory in writing research questions and objectives (Section 2.7).



Box 2.2 Focus on management research

Achieving the United Nations Sustainable Development Goals

An article first published in ‘early view’ online in *Decision Sciences* in February 2021 reviews management literature about the emerging nutraceutical (food that provides medical or health benefits) industry within the context of global goals to end hunger, in particular the United Nations Sustainable Development Goals (Chaurasia et al. 2021). The authors of this review undertook a content analysis (see also Section 12.2) of 138 studies published in peer-reviewed journals to establish the growth of the literature over time, the methodologies

adopted, disciplines considered, theoretical lenses used and sustainability issues considered.

Following their analysis, the authors draw their ideas together in a ‘conclusions and future directions’ framework. In this they summarise trends in the literature and identify potential research areas and gaps to address for the future. These include issues such as sustainable distribution and supply chain network design, which they highlight are not explored in the literature. They also emphasise the need to study public policy-driven issues across geographies in relation to supply chain performance and sustainability practices. Any researcher setting out to explore and generate research ideas relating to supply chains, sustainability and management should therefore consider including this review article in their first batch of reading, after conducting a preliminary search of the existing literature on this topic.

Searching for publications is only possible when you have at least some idea of the area in which you wish to undertake your research. One way of obtaining this is to re-examine your lecture notes and course textbooks and to note those subjects that appear most interesting (discussed earlier in this section) and the names of relevant authors. This will give you a basis on which to undertake a **preliminary search** (using techniques outlined in Sections 3.5 and 3.6). When you have located a series of articles, reports and other relevant items, it is often helpful to look for statements on the absence of research and possibly unfounded assertions in some types of publication, as these are likely to contain or suggest ideas that may help you to develop a research idea.

Scanning the media

Keeping up to date with items in the news can be a very rich source of ideas. The stories that occur every day in the ‘broadsheet’ or ‘compact’ newspapers, in both online and traditional print versions, may provide ideas that relate directly to the item (e.g. the extent to which items sold by supermarkets contravene the principles of ‘green consumerism’ by involving excessive ‘food miles’ in order to import them). Please note, however, that some of these online media are only available by subscription. The stories in these media may also suggest other ideas that flow from the central story (e.g. the degree to which a company uses its claimed environmental credentials as part of its marketing campaign).

Creative thinking techniques

Keeping a notebook of your ideas

One of the more creative techniques that we all use is to keep a **notebook of ideas**. This involves simply noting down any interesting research ideas as you think of them and, of equal importance, what sparked off your thought. You can then pursue the idea using more rational thinking techniques later.

Exploring personal preferences using past projects

One way to generate and evaluate possible project ideas is to explore your personal preferences by reading through a number of past project reports from your university. To get started you need to search through these and select a number that you like and a number that you do not like.

For each project that you like, note down your first thoughts in response to each of the following questions:

- 1** What do you like in general about the project?
- 2** Why do you like the project?
- 3** Which ideas in the project appeal to you?

For each project that you do not like, note down your first thoughts in response to each of the following questions:

- 1** What do you dislike in general about the project?
- 2** Why do you dislike the project?
- 3** Which ideas in the project do not appeal to you?

When you have completed this task, you may find it helpful to spend some time reflecting on each set of notes for the projects you like and those you do not. By reflecting on and thinking about each list you will begin to understand those project characteristics that are important to you and with which you feel comfortable. Of equal importance, you will have identified those with which you are uncomfortable and should avoid.

This process has two benefits. First, it may help you to generate possible research ideas. Second, you may use the project characteristics that emerge from exploring your personal preferences as parameters against which to evaluate possible research ideas.

Exploring relevance to business of ideas in the literature

There is an enormous amount of research published in business and management journals. The nature of these journals varies considerably, ranging from those with a more applied focus to those that are more esoteric. As a result, there will be many ways in which you may explore the relevance to business of ideas published in the literature. The real-world benefits to business practice of academic business research remains a key issue (Box 2.3). Yet, even more esoteric journal articles contain a wealth of ideas that may be explored for their relevance to business. Such articles can contain ideas that you may be able to translate, make operational and test in practice in a given setting, such as a particular organisation, albeit using a simpler methodology than that in the published study. The 'Discussion' section in many business and management journals routinely includes an 'Implications for practice' sub-section, which may guide you towards developing a research idea to explore the relevance of the theory in the article to a particular business setting, such as your employing organisation.

Articles based on empirical studies may also provide you with research ideas. A published empirical study may have been undertaken as a case study. It may have been based in a particular sector or industry, and it may have been based in a particular organisation or type of organisation. Reading it may lead you to think that you could undertake a similar study, albeit possibly scaled down, in a different type of organisation, in a different industry or sector.

There may be scope for you to undertake a case study that seeks to apply the findings from a large sample statistical study to a particular organisational context or type of organisation. This will allow you to test the applicability of these previous findings and to convert them into a relevant and accessible form for a particular context.



Box 2.3 Focus on research in the news

Does business school research deliver real-world benefits?

By Andrew Jack

When Richard Locke at MIT's Sloan School of Management was researching Nike's approach to corporate responsibility in the early 2000s, he came across data on labour standards in its factories that sparked reforms far beyond the sportswear manufacturer.

His experience provides a pointer to how business schools can work with business to bring about positive social change, bridging a divide between ideas and practice that critics argue remains far too wide.

After lengthy negotiations to gain access to corporate records and freedom to publish his findings, Prof Locke, now provost at Brown University, was able to demonstrate the limited effectiveness of labour audits alone in improving working conditions. Far greater progress came when they were combined with measures to tackle underlying problems, such as training and enabling suppliers to schedule their work better.

The conclusions, disseminated over a number of years in seminars and in consultations with managers, unions and policymakers as well as in academic journals and more accessible publications, helped spark new policies at multiple companies. 'It's extremely important for scholars in business schools to try to address some of society's great challenges through their research,' he says. 'By bringing a rigorous methodology, you can both show your academic skills and generate new research to not only change the way we think but do so with implications in the real world.'

For many, such examples remain too rare. In a 2018 article in *BizEd*, a journal of the Association to Advance Collegiate Schools of Business, William Glick from Rice University, Anne Tsui from the University of Notre Dame and Gerald Davis from the University of Michigan delivered a damning verdict. 'With a few notable exceptions,' they wrote, 'scholarly research rarely reaches the worlds of business or policy, and academic journals are neither read nor cited widely beyond the academic community.'



Source of extract: Jack, A. (2021) 'Does business school research deliver real-world benefits?', *Financial Times*, 27 January. Copyright 2021 The Financial Times Limited

Creatively approaching the literature to convert existing work into a relevant and specifically applied study, in the ways we have described, may provide you with a rich and valuable research idea.

Relevance trees

Relevance trees may also prove useful in generating research ideas. In this instance, their use is similar to that of mind mapping (Buzan 2018) in which you start with a broad concept from which you generate further (usually more specific) topics. Each of these topics forms a separate branch from which you can generate further, more detailed sub-branches. As you proceed down the sub-branches, more ideas are generated and recorded. These can then be examined and a number selected and combined to provide a research idea. This technique is discussed in more detail in Section 3.5 (and illustrated in Box 3.6).

Brainstorming

The technique of **brainstorming** (Box 2.4), taught as a problem-solving technique on many business and management courses, can also be used to generate and refine research ideas. It is best undertaken with a group of people, although you can brainstorm on your own. Brainstorming involves a number of stages:

- 1 Defining the problem.** This will focus on the sorts of ideas you are interested in – as precisely as possible. In the early stages of formulating a topic this may be as vague as, ‘I am interested in marketing but don’t have any ideas for my research.’
- 2 Asking for suggestions.** These will relate to the problem.
- 3 Recording suggestions.** As you record these you will need to observe the following rules:
 - No suggestion should be criticised or evaluated in any way before all ideas have been considered.
 - All suggestions, however wild, should be recorded and considered.
 - As many suggestions as possible should be recorded.
- 4 Reviewing suggestions.** You will seek to explore what is meant by each as you review these.
- 5 Analysing suggestions.** Work through the list of ideas and decide which appeal to you most as research ideas and why.



Box 2.4 Focus on student research

Brainstorming

George’s main interest was football. In his university city, he worked part-time in the retail store of the local football club and thought he would like to carry out his research project in this setting.

When he finished university, he wanted to work in marketing, preferably for a sports goods manufacturer or retailer. He had examined his own strengths and discovered that his highest marks were in marketing. He wanted to do his research project on some aspect of marketing, preferably linked to the football club, but had no real research idea. He asked three friends, all taking business management degrees, to help him brainstorm the problem.

George began by explaining the problem in some detail. At first the suggestions emerged slowly. He noted them down on some flipchart sheets. Soon a number of sheets of paper were covered with

suggestions and pinned up around the room. George counted these and discovered there were over 100.

Reviewing individual suggestions produced nothing that any of the group felt to be of sufficient merit for a research project. However, George recalled an article they had been asked to read based on a case study of an English Premier League football club (Ogbonna and Harris 2014). He had found this interesting because of its subject. He recalled that it was about organisational culture being perpetuated within organisations that have a long history of success, and stakeholder groups such as football fans who have a strong sense of identity.

George’s recollections of this article encouraged the group to discuss their suggestions further. Combining a number of suggestions from the flipchart sheets with their discussion about organisational cultural perpetuation, George noted a possible research idea as: ‘The impact of factors that perpetuate organisational culture on the development of marketing strategies – help or hindrance?’

George thought this idea could be based on his local football club.

George arranged to see his project tutor to discuss how to refine the idea they had just generated.

2.4 Refining research ideas

The Delphi technique

An approach that our students have found useful to refine their research ideas is the **Delphi technique** (Box 2.5). The standard Delphi method involves a researcher using a purposive sample of participants who are knowledgeable about the topic to be discussed; asking these participants to write down their answers anonymously to some initial questions to gather their opinions and perceptions; analysing these answers thematically; using this to generate a second round of questions to gain participants' feedback to the initial responses; repeating this process until a consensus is reached about the topic in order to inform decision-making, policy or practice. The initial round of questions is likely to be 'open' or 'semi-open', while subsequent rounds of questions are likely to be more focused and structured (Brady 2015). This process works well, not least because people enjoy trying to help one another. In addition, it is very useful in forming cohesive groups.

Preliminary inquiry

Having generated a research idea, you will need to refine it and express it as a clear overarching research question(s), a research aim and research objectives. This will involve searching for and evaluating literature and other related sources. Even if you searched the literature to generate your research idea, it is likely to be necessary to conduct another search of it in order to refine this idea into a workable research question. Once you have your research idea you can re-visit the literature with a much clearer focus to understand how this helps you to refine your research idea, and to develop the research question(s), aim and set of objectives (Sections 2.5 and 2.6).



Box 2.5 Focus on student research

Using a Delphi Group

Tim explained to the group that his research idea was concerned with understanding the decision-making processes associated with mortgage applications and loan advances. His briefing to the three other group members, and the questions that they asked him, considered aspects such as:

- the particular situation of potential first-time house buyers;
- the way in which the nature of contact between potential borrowers and financial institutions might influence decision-making.

The group then moved on to generate a number of more specific research ideas, among which were the following:

- the effect of being a first-time house purchaser on mortgage application decision-making;
- the effect of online only applications on mortgage decisions;
- the attributes that potential applicants look for in financial institutions operating in the mortgage market.

These ideas were considered and commented on by all the group members. At the end of the second cycle Tim had, with the other students' agreement, refined his research idea to:

- an evaluation of the factors that influence potential first-time buyers' choice of lending institution.

Tim now needed to pursue this idea by undertaking a preliminary search of the literature.

This search activity to refine, focus and operationalise your initial research idea into an overarching research question(s), aim and objectives or investigative questions involves a **preliminary inquiry** or initial inquiry. This may lead to the first iteration of your critical literature review or help to inform it (Figure 3.1). Unfortunately, because journal articles are word limited and the ‘methods’ section of an article only describes the research methodology and techniques used in the actual study, research is often presented as an unproblematic (and not a ‘messy’) process.

For some researchers, the preliminary inquiry may include informal discussions with people who have personal experience of and knowledge about your research ideas. It may also involve **shadowing** (following and observing) employees who are likely to be important in your research and who may therefore be able to provide some initial insights. If you are planning on undertaking your research within an organisation, it is also important to gain a good understanding of your host organisation (McDonald 2005).

At this stage, you should test your ideas using the checklist in Box 2.1 and, where necessary, amend them. It may be that after a preliminary inquiry, or discussing your ideas with colleagues, you decide that the research idea is no longer feasible in the form in which you first envisaged it. If this is the case, do not be too downhearted. It is far better to revise your research ideas at this stage than to have to do it later, when you have undertaken far more work.

Integrating ideas

Another, or complementary, way to refine, focus and operationalise your research idea into a research question(s), aim and objectives is to integrate ideas generated using a number of different techniques. Integrating ideas will help your research to have a clear purpose and direction. Jankowicz (2005: 34–6) offers an integrative process that our students have found most useful. This he terms ‘working up and narrowing down’. It involves classifying each research idea first into its area, then its field and finally the precise aspect in which you are interested. These represent an increasingly detailed description of the research idea. For example, your initial area, based on examining your coursework, might be accountancy. After searching through relevant journal articles and holding a discussion with colleagues, this might become more focused on the field of financial accounting methods. After a further literature search and reading, the use of a Delphi technique and discussion with your project tutor you decide to focus on the aspect of activity-based costing.

Refining ideas given by your employing organisation

As a part-time student, your manager may provide you with an embryonic research idea. This may be something that affects your work and in which you have an interest. You may have discussed this with your manager and relish the opportunity to research it further.

It may, however, be something in which you are not particularly interested. In this case you will have to weigh the advantage of doing something useful to the organisation against the disadvantage of a potential lack of personal motivation. You therefore need to achieve a balance. Often the research project your manager wishes you to undertake is larger than is appropriate for your course. In such cases, it may be possible to complete both by isolating an element of the larger organisational project that you find interesting and treating this as the project for your course.

One of our students was asked to do a preliminary investigation of the strengths and weaknesses of her organisation’s pay system and then to recommend consultants to design

and implement a new system. She was not particularly interested in this project. However, she was considering becoming a freelance human resources consultant. Therefore, for her research project she decided to study the decision-making process in relation to the appointment of human resources consultants. Her organisation's decision on which consultant to appoint, and why this decision was taken, proved to be a useful case study against which to compare management decision-making theory. In this event you would write a larger report for your organisation and a part of it for your project report. Section 14.4 offers some guidance on writing two separate reports for different audiences.

Other problems may involve your political relationships in the organisation. For example, there will be those keen to commission a project that justifies their particular policy position and see you as a useful pawn in advancing their political interests. It is important to have a clear stance with regard to what you want to do, and your personal objectives, and to stick to this. A further potential problem may be one of your own making: to promise to deliver research outcomes to your employer and not do so.

Conducting research in your own organisation may also be problematic because of your role as an internal researcher (Tietze 2012). We return to discuss a range of issues related to this role in Section 5.12.

2.5 Developing your overarching research question

You will know when you have a clear research idea as you will be able to say, 'I'd like to do some research on . . .'. Obviously, there is still a big gap between this and being able to start serious work on your research project. You will, however, be in a position to develop your overarching research question(s), before expressing it as a research aim and set of research objectives.

It will be important for you to express your research topic as one, or occasionally two or three, clearly defined **research questions** that your research will address before commencing the research process. Within this chapter we refer to this as your overarching research question to emphasise that this is the overview key question that your research will address. As a student, you are likely to be required to include a research question in your written research proposal (Sections 2.8 and 2.9). The importance of creating a clearly defined question or questions cannot be overemphasised. An **overarching research question** will allow you to say what the issue or problem is that you wish to study and what your research project will seek to find out, explain and answer. One of the key criteria of your research success will be whether you have developed a set of clear conclusions from the data you have collected. The extent to which you can do that will be determined largely by the clarity with which you have posed your question (Box 2.6).

The overarching research question will be at the centre of your research project. It will influence the literature you review, your research design, the access you need to negotiate, your approach to sampling, your choice of data collection and analysis methods, and help to shape the way in which you write your project report. It will also be used to generate a set of more detailed research objectives or investigative (research) questions to guide your research, discussed later.

However, it is also important to recognise that some research approaches and research strategies start off in a more exploratory and emergent direction (Chapter 5). For a researcher undertaking this type of research, their final research question may only emerge during the process of data collection and analysis as they discover the exact focus of the research project and refine its direction. Some of the mainly qualitative strategies



Box 2.6 Focus on student research

Defining the overarching research question

Imran was studying for a BSc in Business Management and undertaking his placement year in an advanced consumer electronics company. When he first joined the company, he was surprised to note that the company's business strategy, which was announced in the company newsletter, seemed to be inconsistent with what Imran knew of the product market.

Imran had become particularly interested in corporate strategy in his degree. He was familiar with some of the academic literature that emphasised the importance of 'fit' between the corporate strategy and the external environment in which the organisation operated. He wanted to do some research on corporate strategy in his organisation for his research project to better understand the concept of 'fit'.

After talking this over with his project tutor, Imran decided on the following overarching research question: 'To what extent does [organisation's name]'s corporate strategy "fit" their external operating environment and why?'

(discussed in Chapter 5), including Ethnography Studies and Grounded Theory, are exploratory and emergent and will often lead you, where you use one of these, to refine your initial overarching research question as you progress. Most tutors will say here that it is part of the process to refine the question as your project progresses to reflect the emerging focus of your research. It is always advisable to discuss such developments with your project tutor! The key point is that if you use such a research approach it is still important to define an initial overarching research question at the outset of your project to focus your research, even if you then refine your question accordingly.

Developing the overarching research question or questions, rather like generating and refining research ideas (Sections 2.3 and 2.4), is not a straightforward matter. It is important that a question is sufficiently involved to generate the sort of project that is consistent with the standards expected of you (Box 2.1). A question that prompts a descriptive answer – for example, 'What is the proportion of graduates entering the UK civil service who attended elite universities?' – is far easier to answer than: 'Why are graduates from elite universities more likely to enter the UK civil service than graduates from other universities?' However, answering the first question is unlikely to satisfy your examining body's requirements as it only needs description.

Questions may be divided into ones that are exploratory, descriptive, explanatory or evaluative (Section 5.7). Overarching research questions you ask are likely to begin with or include either 'what', 'when', 'where', 'who', 'why', 'how' or 'to what extent'. Some of these can result in an answer that is partly or entirely descriptive, such as: 'What was the cost of the marketing campaign for the new range of products?' Exploratory questions are likely to begin with 'How' or 'What'. For example, 'How has the corporate rebranding strategy affected consumer attitudes?' Questions that seek explanations will either commence with 'Why' or contain this word within the question. For example, a question may ask customers what they think about a new product and why they like or dislike it. Questions that are evaluative are also likely to begin with 'How' or 'What' but unlike the 'How much...?' or 'How has...?' questions, an evaluative question might ask, 'How effective was the marketing campaign for the new range of products?' Another way of wording this type of question might be, 'To what extent was the marketing campaign effective and why?'

The Goldilocks test

While some questions may be too simple, it is perhaps more likely that you might fall into the trap of asking research questions that are too difficult. The question cited earlier, ‘Why are graduates from elite universities more likely to enter the UK civil service than graduates from other universities?’, is a case in point. It would probably be very difficult to gain sufficient access to the civil service to get a good grasp of the subtle ‘unofficial’ processes that go on at staff selection that may favour one type of candidate over another. Over-reaching yourself in the definition of research questions is a danger.

Clough and Nutbrown (2012) use what they call the ‘**Goldilocks test**’ to decide if research questions are either ‘too big’, ‘too small’, ‘too hot’ or ‘just right’. Those that are too big probably need significant time and demand too many resources. Questions that are too small are likely to be of insufficient substance, while those that are too ‘hot’ may be so because of sensitivities that might be aroused as a result of doing the research. This may be because of the timing of the research or the many other reasons that could upset key people who have a role to play, either directly or indirectly, in the research context. Overarching research questions that have been written to take into account the researcher’s capabilities and the availability of resources, including time and the research setting, are more likely to be about right.

The pitfall you must avoid at all costs is asking an overarching research question that will not generate new insights. This raises the question of the extent to which you have consulted the relevant literature. It is perfectly legitimate to replicate research because you have a genuine concern about its applicability to your research setting (for example, your organisation or your country’s context). However, it certainly is not legitimate to display your ignorance of the literature.

The Russian doll principle and the AbC rule

To clarify a research question, Clough and Nutbrown (2012) talk of the Russian doll principle. This means refining your initial overarching research question until it reflects the essence of your research idea without including any unnecessary words or intentions. By stripping away any unnecessary layers (the larger outer dolls), the clearly defined research question (the smallest doll) that you reveal should provide you with an appropriately focused starting point for your research project. Dudau’s (2016) AbC rule offers further insightful advice on the components of your research question, emphasising it should include:

- one or two clearly stated **Abstract** concepts;
- the **Context** in which the research is undertaken.

For example, ‘to what extent has organisational citizenship impacted on turnover intentions in the public sector and why?’ Here ‘organisational citizenship’ and ‘turnover intentions’ are the abstract concepts and the ‘public sector’ the context. More than two **Abstract** concepts (or the absence of any abstract concept) make for an undefined theoretical focus, whereas lack of context misses the opportunity for an empirical or theoretical anchor to your study. **Abstract concepts** and **Context** are highlighted for each of the focused overarching research questions in Table 2.2.

Writing your overarching research question(s) will be, in most cases, your task, but it is useful to get other people to help. An obvious source of guidance is your project tutor. Consulting your project tutor will avoid the pitfalls of the questions that are too easy or too difficult or have been answered before. Discussing your question with your project tutor will lead to it becoming much clearer.

Table 2.2 From research idea to overarching research question

| Research idea | (Initial) overarching research question | (More focused) overarching research question |
|---|---|--|
| Media campaign following product recalls | In what ways can media campaigns be designed to increase consumer trust, value and loyalty in exchange relationships following product recalls? | How does brand equity in media campaigns increase consumer trust, value and loyalty following product recalls? |
| Graduate recruitment post pandemic | How have firms responded post pandemic to ensure effective recruitment and selection of graduates? | Can cryptocurrency predict person environment fit and competence in post-pandemic graduate recruitment ? |
| Supermarket coupons as a promotional device | To what extent do supermarket coupon promotions influence buyer behaviour? | Does couponing affect buyer motivation in supermarkets ? |
| Small business start-up funding | To what extent is small business start-up borrowing influenced by the characteristics of the owner? | To what extent are small business start-ups' borrowing decisions influenced by the owners' need for independence ? |

Key: **A**bstract concept, **C**ontext

2.6 Writing a research aim and set of research objectives

Research aim

As well as your overarching research question, you may also be required to formulate a research aim. A **research aim** is a brief statement of the purpose of the research project. It is often written as a sentence stating what you intend to achieve through your research. To illustrate this, the examples of research questions in Table 2.2 have been matched to their research aims in Table 2.3. You will see the close relationship between these: one stated as a question, the other as an aim.

Table 2.3 From overarching research question to research aim

| Overarching research question | Research aim |
|---|---|
| How does brand equity in media campaigns increase consumer trust, value and loyalty following product recalls? | ... to assess the impact of brand equity on consumer trust, value and loyalty relationships in media campaigns following a series of product recalls. |
| Can cryptovetting predict person environment fit and competence in post pandemic graduate recruitment? | ... to explore and explain the relationship between cryptovetting and graduates' person-environment fit and competence in post-pandemic recruitment. |
| Does couponing affect motivation to purchase products in supermarkets? | ... to establish the impact of supermarket couponing on buyer motivation. |
| To what extent are small business start-ups' borrowing decisions influenced by the owners' need for independence? | ... to examine the extent to which small business start-ups' borrowing decisions are affected by the owners' need for independence. |

Research objectives and investigative questions

Your overarching research question and research aim are complementary ways of saying what your research is about, providing a clear indication of the abstract concepts you will consider (and associated theory) and the context. However, neither gives sufficient detail about purpose and direction: the steps you will need to take to undertake your research. To do this you will need to devise a set of research objectives or more detailed investigative (research) questions. Many project tutors have a preference for either research objectives or investigative questions, but it may be that either is satisfactory. Do check whether your examining body has a preference.

Research objectives allow you to **operationalise** your question, that is to state the steps you will take to answer it. Each **research objective** therefore provides a clear specific statement of an aspect of the research that you need to undertake to meet your research aim. In contrast, **investigative questions** state with precision and depth the questions that by answering also mean you have answered your overarching research question and met your research aim.

Writing useful research objectives or investigative questions both require you to meet a number of fit-for-purpose criteria. Table 2.4 sets out criteria to help you devise research objectives and investigative questions to answer your overarching research question and meet your research aim. Each of these criteria is also rephrased as a short question, which you can use as a checklist to evaluate your own draft research objectives or investigative questions. Box 2.7 provides an example set of objectives at the stage when a student's overarching research question and aim were developed into a sequence of research objectives.

Table 2.4 Research objective and investigative question criteria

| Criterion | Purpose |
|--|--|
| Transparency (What does it mean?) | The meaning is clear and unambiguous. |
| Specificity (What I am going to do?) | The purpose is clear and easily understood, as are the actions required to fulfil it. |
| Relevance (Why I am going to do this?) | The link to the aim and overarching research question and wider research project is clear. |
| Interconnectivity (How will it help to complete the research project?) | Taken together as a set, the research objectives and/or investigative questions outline the steps in the research process from its start to its conclusion, without leaving any gaps. In this way, they form a coherent whole. |
| Answerability (Will this be possible?) (Where shall I obtain data?) | The intended outcome is achievable. Where this relates to data, the nature of the data required will be clear or at least implied. |
| Measurability (When will it be done?) | The intended outcome will be evident when it has been achieved. |



Box 2.7

Focus on student research

Writing research objectives

Diane worked for a medium-sized technology company that had been taken over by a much larger, multi-divisional firm. This company was gradually being integrated into its larger parent. Originally, the company had been incorporated as one division of the parent firm, although recently one of its most successful product areas had been reallocated into a different division. This had adverse consequences for many employees because it narrowed their scope to work across the company and to seek development opportunities. Many of the original employees had already left the company's employment. However, a significant number of the original employees remained and others who had joined since the take-over had been assimilated into the organisational culture that still prevailed from before the merger.

Diane was undertaking a management course as a part-time student. This course included a substantial research project and Diane thought that the changes at this company and how these affected employees' perceptions of working there would be a fascinating topic and provide the context for her research. She now needed to be clear about the related abstract concepts. She began to review literature related to organisational change, and in particular the impact of organisational structures on those who work within them, and decided that organisational culture and climate were potential abstract concepts. She brainstormed some ideas related to this topic and spent time evaluating these. She then decided to discuss her research idea with two people whose advice she valued.

First, she spoke to one of her tutors. Her tutor felt that this idea had merit but took time to discuss two possible concerns that focused around access and ethical issues. One focused on the likely sensitive nature of this research topic for both the company and those employees from whom Diane would need to collect data. The other focused on the fact that, as Diane worked for the company and alongside those from

whom she would need to collect data, there were concerns about confidentiality. However, as they discussed this research idea they agreed that, if addressed in a sensitive way that absolutely ensured confidentiality, it could be possible to use her ideas to develop a suitable research project.

Second, she spoke to one of the senior managers in the company. This manager knew that as a part-time student on a management course Diane needed to undertake an organisationally based research project. She explained her research idea to this manager carefully and her justification for wishing to undertake it. This manager had been employed in the company for several months, having been recruited from outside both the company and the parent firm. The manager was aware that some employees had spoken of their concerns about the ways in which the nature of work and scope for development had changed as a result of the take-over and structural changes.

This manager told Diane that there were significant concerns associated with her research idea. In particular, the manager thought there would be a risk of generating greater negativity with real consequences for the company. However, the manager also said that if conducted with sensitivity, her project might prove to be helpful. This, the manager felt, might be achieved by promoting a positive outcome from the research project by asking research participants to indicate how they could become re-engaged with and more committed to the company in spite of the changes that had occurred.

This manager also recognised that if employees felt the company was seen as being behind this research idea, they might be suspicious of Diane and refuse to share meaningful data with her. They discussed this and agreed that if she was going to proceed with this research idea, she should let potential participants know that the data produced would be used only for her dissertation; she would separately produce a short summary document for senior managers that would only focus on recommendations for improvement based on an aggregated level of analysis to ensure confidentiality and anonymity.

These two discussions gave Diane a great deal to think about. She decided to undertake a preliminary





Box 2.7 **Focus on student research (continued)**

Writing research objectives

inquiry to help refine her ideas about her research topic. Without referring to the discussion she had held with the manager to ensure confidentiality, she also held a Delphi group with a small trusted group of colleagues to refine her ideas and to build in scope for a positive focus. She also sought to integrate her ideas by working them up and narrowing them down. After this, she made an appointment to see her tutor. She took a draft overarching research question, aim and set of objectives to this meeting.

The research question read, 'How have employee engagement, commitment and development been affected in a medium-sized technology company, and how may these be promoted following recent organisational change?' It included both abstract concepts (employee engagement, employee commitment and employee development) and a context (medium-sized technology company) that had undergone

organisational change). The research aim was, 'The aim of this research is to evaluate employee engagement, commitment and development in a medium-sized technology company and explore how these may be promoted following recent organisational change.' The set of objectives were:

- 1** to describe the nature and cause of recent organisational change;
- 2** to define clearly the concepts used (employee engagement, employee commitment and employee development) to evaluate the impact of organisational change;
- 3** to evaluate the impact of recent organisational change on employee engagement and commitment;
- 4** to evaluate the impact of recent organisational change on employees' perceptions about their scope for development and future progression;
- 5** to explore ways to promote employee engagement, commitment, development and progression following recent organisational change in the company;
- 6** to make recommendations to promote employee engagement, commitment, development and progression in the company.

2.7 The importance of theory

We have already alluded briefly to the importance of theory in our earlier discussion of overarching research questions and abstract concepts. We now consider in more detail, what it is, why it is important, and how it informs the research objectives and investigative questions before discussing types of theoretical contribution. This is developed further in the discussion of approaches to theory development in Section 4.5.

What theory is

The term **theory** is used to refer to 'a systematic body of knowledge grounded in empirical evidence which can be used for explanatory or predictive purposes' (Saunders et al. 2015: 37). Theories are therefore based upon the development and examination of (abstract) concepts, the clear definition of these concepts being essential for testing and developing theory (Box 2.7). A theory uses related facts and concepts to provide an explanation or predict an outcome. The explanatory power of a theory is based on its ability to explain relationships between concepts. These explanations need to be capable of being confirmed, refined or contradicted as understandings develop and change based on further research.

To explore the question ‘what is theory?’ in more detail we use the influential work of Whetten (1989). Whetten identified that theory is composed of four elements, related to ‘what’, ‘how’, ‘why’ and a fourth group of ‘who’, ‘where’ and ‘when’. The first of these may be summarised as: what are the concepts or variables that the theory examines? For example, in Box 2.7, the concepts in Diane’s research question are organisational change, employee engagement, employee commitment and employee development.

The second element may be summarised as: how are these concepts or variables related? Diane’s research question was designed to examine the relationships between organisational change, on the one hand, and employee engagement, commitment and development, on the other hand. A key aspect here is **causality**. Theory is concerned with cause and effect. In her research, Diane was interested to explore how organisational change affected employee engagement, employee commitment and employee development. In other words, how did change have an effect on each one of these?

The third element may be summarised as: why are these concepts or variables related? This is the critical element in a theory because it explains the reasons for relationships between the concepts or variables. According to Whetten, ‘what’ and ‘how’ are descriptive; it is ‘why’ that explains the relationship. This point is worth developing, as you may be asking, ‘what is the difference between “how” and “why” in this context?’ In the case of Diane’s research, she found that organisational change had affected employee engagement, commitment and development, respectively. Diane’s data allowed her to recognise a number of relationships that she could describe. However, this description did not explain why these outcomes had occurred and, in fact, the reasons for them were complex. For example, different categories of employee provided different explanations for the impact of organisational change upon themselves. Diane needed to analyse her data further (and where necessary to extend its collection) to answer the question, ‘why do these relationships exist in my data?’

Once a good theory has been developed it may be used not only to explain why any relationship exists, but also to predict outcomes in a similar situation or where one or more of these theoretical variables are manipulated (changed). In the case of Diane’s research, her theory may be used to predict a similar impact on employee engagement and commitment where change in another albeit similar organisation is implemented in the same way. Her theory may also be used to predict different outcomes for employee engagement and commitment where organisational change is managed differently.

While good theory has the power to explain and predict, it may also be subject to limitations. The scope of many theories will be limited by one or more constraints. The fourth group of elements that Whetten identified may therefore be summarised as: who does this theory apply to; where does this theory apply; when does this theory apply? In the case of Diane’s research, she recognised that some of her theoretical conclusions applied more to professional-grade staff but less so to administrative staff. She also recognised that with the introduction of new policies to re-engage employees and offer development opportunities, the applicability of some of her conclusions would need to be re-evaluated in the future.

In this way, the explanations of the cause-and-effect relationships between variables in a theory may be contextual and time limited, indicating constraints to their generalisability. Another important contribution that addresses the question ‘what is theory?’ starts from the opposite perspective by discussing ‘What theory is not’ (Sutton and Staw 1995). This is a helpful contribution to our understanding and provides a complementary approach to that of Whetten (1989) (Box 2.8).



Box 2.8 Focus on management research

Clarifying what theory is not

Sutton and Staw (1995) make a helpful contribution to the question 'What is theory?' by defining what it is not. In their view theory is not:

- 1 **References.** Listing references to existing theories and mentioning the names of such theories may look impressive. However, alluding to the theory developed by other researchers may only provide a smokescreen. Instead researchers need to identify the concepts, causal relationships and logical explanations that they are using from previous theoretical work in relation to their own work.
- 2 **Data.** Data are important to be able to confirm, revise or overturn existing theory and to be able to develop new theory. However, data are used to describe the relationships or patterns that are revealed from their collection and analysis. Description by itself does not equal theory. Theory also requires logical explanations to discuss why such relationships or patterns were revealed, or

why they might be expected to be revealed when testing existing theory (Section 4.3).

- 3 **Lists of variables.** Variables are important in the process of theory development but simply presenting or listing these by themselves does not represent a theory.
- 4 **Diagrams.** Diagrams are often helpful to show observed or expected causal relationships and how different relationships are related or how they are expected to be related. However, by themselves diagrams or figures are not theory. Sutton and Staw (1995: 376) state: 'Good theory is often representational and verbal.' They say that clear explanations can be represented graphically but that, to be able to develop a rich theoretical understanding, these will also require written discussion to explain why these relationships exist.
- 5 **Hypotheses or predictions.** In a similar manner to point 3, hypotheses are an important part of the process of developing and testing theory, in particular theoretical approaches (Experiment in Section 5.8), but they do not constitute a theory by themselves.

You are likely to use objectives or investigative questions rather than hypotheses in your research design and we would add to point 5 that the propositions or concepts that inform your research questions are also not theory by themselves.

Why theory is important

There is probably no word that is more misused and misunderstood in education than the word 'theory'. It is thought that material included in textbooks is 'theory', whereas what is happening in the 'real world' is practice. Students who used previous editions of this book remarked that they were pleased that the book was not too 'theoretical'. What they meant was that the book concentrated on giving lots of practical advice. Yet this book is full of theory. Advising you to carry out research in a particular way (variable A) is based on the theory that this will yield effective results (variable B). This is the cause-and-effect relationship referred to in the definition of theory developed earlier, and is very much the view of Kelly (1955). Kelly argues that as individuals attempt to solve the daily problems we face, we go about this activity in much the same way as the scientist. Both continuously make and test hypotheses and revise their concepts accordingly. Both organise their results into what are called schemata and then into a system of broader schemata, which are called theories. Kelly asserts that we need such schemata and theories in order to make sense of the complexity of the world in which we live. Without these organising frameworks we would be overwhelmed by the unconnected detail we would have to recall.

Implicitly, each of us uses theory in our lives and in the jobs that we undertake, for example the marketing manager who believes that issuing coupons (couponing) in the supermarket chain for which they work motivates customers to return, be more loyal, and be less likely to shop regularly at a competitor supermarket (Table 2.2).

| | | |
|---|--|---|
| Supermarket coupons as a promotional device | To what extent do supermarket coupon promotions influence buyer behaviour? | Does couponing affect buyer motivation in supermarkets ? |
|---|--|---|

This is a theory even though the marketing manager would probably not recognise it as such. They are less likely to refer to it as a theory, particularly in the company of fellow managers. Many managers are very dismissive of any talk that smacks of 'theory'. It is thought of as something that is all very well to learn about at business school, but that bears little relation to what goes on in everyday organisational life. Yet the coupons example shows that it has everything to do with what goes on in everyday organisational life. By issuing coupons (variable A), the supermarket is attempting to influence the buyer motivation (variable B). As every supermarket chain issues their own coupons, the marketing manager's personal theory that this encourages loyalty may begin to seem inadequate when confronted by a range of other complementary and innovative strategies to encourage customers to switch where they shop.

The use of coupons may become just one variable among many as supermarkets compete by offering extra loyalty card bonus points on particular goods, double or treble points if customers spend over a certain amount, the opportunity to redeem the value from accumulated bonus points against a range of discounted offers, and so on. In this case, research will provide the marketing manager with a much greater understanding of the effectiveness of couponing strategies used within her or his supermarket chain. The data collected will allow theoretical explanations to be developed, based on causal relationships that may then be used to predict which of these strategies is more effective. It may also indicate that different strategies will be effective in different locations and perhaps that specific strategies are more effective at particular times of the year, or that specific strategies should be targeted at particular socioeconomic groups. The ability to make these predictions potentially allows the supermarket chain to compete more effectively against its rivals. Valid theoretical explanations may lead to predictions that offer the supermarket chain increased opportunities for influence and control, and the possibility of increasing market share.

If theory is so rooted in our everyday lives, it is something that we need not be apprehensive about. If it is implicit in all of our decisions and actions, then recognising its importance means making it explicit. In research, the importance of theory must be recognised: therefore, it must be made explicit.

How theory informs research

So far we have defined the elements of theory and discussed the need to recognise it in your research, even as you start to plan this. At this point, you may be asking, 'why is it important for me to recognise theory at this early stage, when still developing my overarching research question and research objectives?' This relates to the capacity of theory to inform your research ideas (discussed earlier), your overarching research question, research aim and objectives, or investigative questions.

Theory published in the literature may inform your overarching research question in several ways. It will help you to formulate a question that should lead to a theoretical explanation, rather than just a descriptive answer. It will allow you to find out whether others have asked similar questions to the question you propose. Where you find that a

similar research question to yours has been addressed in the literature, you will be able to learn about the context within which it was explored and how the research was conducted. This may help to focus your question to provide you with a set of variables to test, or concepts to explore, to determine whether, how and why they are related in the context of your own research project (Box 2.9).

Using relevant theory to inform your research question will also sensitise you to the nature and level of importance of the research undertaken surrounding your question. You may find that a considerable body of relevant work exists, either in business and management or in another subject domain, for example in psychology, economics or sociology. Discovering this may help you to focus so that later on you can firmly connect your findings and conclusions to this existing theory. It is unlikely that you will fail to find any literature that relates to your proposed question, although where you find that you are working in a more specialised topic area, this discovery may also help to focus your research question to relate to the theory that you locate. It will be important to discuss how the results of your research relate to theory, to be able to assess that theory in the context of your work and to demonstrate the theoretical contribution, no matter how limited, of your research.

Where you simply find it difficult to formulate an overarching research question from your research idea, using existing theory may also help you to achieve this.

How theory is developed

How theory is developed provides a crucial reason for recognising relevant theory when writing your research question and objectives. Your research project will be designed (even if this appears implicit) to test a theory or to develop a theory. Where you wish to



Box 2.9 Focus on student research

Writing an overarching research question based on theory

Justine was a final-year marketing undergraduate who was interested in the theory of cognitive dissonance (Festinger 1957). She wanted to apply this to consumer-purchasing decision-making in the snack foods industry (e.g. buying potato crisps) in the light of the adverse publicity that the consumption of such foods has as a result of 'healthy eating' campaigns.

Justine applied Festinger's theory by arguing in her research project proposal that a consumer who learns that eating too many snacks is bad for her health will experience dissonance because the knowledge that eating too much snack food is bad for her health will be dissonant with the cognition

that she continues to eat too many snacks. She can reduce the dissonance by changing her behaviour, i.e. she could stop eating so many snacks. (This would be consonant with the cognition that eating too many snacks is bad for her health.) Alternatively, she could reduce dissonance by changing her cognition about the effect of snack overeating on health and persuade herself that it does not have a harmful effect on health. She would look for positive effects of eating snacks, for example by believing that it is an important source of enjoyment that outweighs any harmful effects. Alternatively, she might persuade herself that the risk to health from overeating snacks is negligible compared with the danger of car accidents (reducing the importance of the dissonant cognition).

Justine's research question was, 'To what extent does adverse "healthy eating" campaign publicity affect the consumer's decision to purchase snack foods and why?'

adopt a clear theoretical position that you will test through the collection of data, your research project will be theory driven and you will be using a **deductive approach**. Where you wish to explore a topic and develop a theoretical explanation as the data are collected and analysed, your research project will be data driven and you will be adopting an **inductive approach**. We discuss approaches to theory development, also introducing the abductive approach, in much greater detail later (Section 4.5). However, it is useful to introduce this fundamental difference in the way theory is developed now to show why you need to think about this when developing your overarching research question, research aim, research objectives and investigative questions. A deductive approach will require you to identify a clear theoretical position when you draft the research question that you will then test. This is the approach we outlined earlier (Box 2.9).

An inductive approach does not rely on identifying an existing theoretical position, but it is likely that if you adopt this approach you will still need to familiarise yourself with theory in your chosen subject area before you draft your research question. Using an inductive approach does not mean disregarding theory as you formulate your research question and objectives. An inductive approach is intended to allow meanings to emerge from data as you collect them in order to identify patterns and relationships to build a theory, but it does not prevent you from using existing theory to formulate your research question and even to identify concepts that you wish to explore in the research process (Section 4.5). In this way, all researchers are likely to commence their research with knowledge of relevant literature and the theory it contains.

There is a further relationship between theory and your research that is important to recognise when developing your research proposal. In our discussion of theory, we recognised that it is crucial to be able to explain how variables or concepts are related and why they are related. Overarching research questions and investigative questions can play a crucial role in encouraging research that is designed to produce theoretical explanations, no matter how limited these explanations might be. A question that only encourages a descriptive outcome will not lead to a theoretical explanation. For example, compare the following questions. ‘How satisfied are employees with recent changes in the department’s business strategy?’ ‘What are the implications of recent changes in the department’s business strategy for employee satisfaction and why?’ The first question will result in a descriptive outcome. The second question has the potential to explore and test relationships and to arrive at theoretical explanations to explain why these might exist.

Types of theoretical contribution

Our discussion of theory has probably left you asking, ‘What does this mean for me?’ While you will be expected to produce a theoretical explanation, you will not be expected to develop a momentous theory that leads to a new way of thinking about management! Not all theoretical contributions are the same and it is reassuring to look at the threefold typology of theories shown in Figure 2.1.

‘Grand theories’, such as Newton’s theory of gravity discussed in the opening vignette, are usually thought to be the province of the natural scientists. These may be contrasted with ‘middle-range theories’, which lack the capacity to change the way in which we think about the world but are nonetheless of significance. Some theories such as Maslow’s (1943) hierarchy of needs and Herzberg et al.’s (1959) two-factor theory of motivation are well known to managers and would be in this category. However, most of us are concerned with ‘substantive theories’ that are restricted to a particular time, research setting, group or population, or problem.

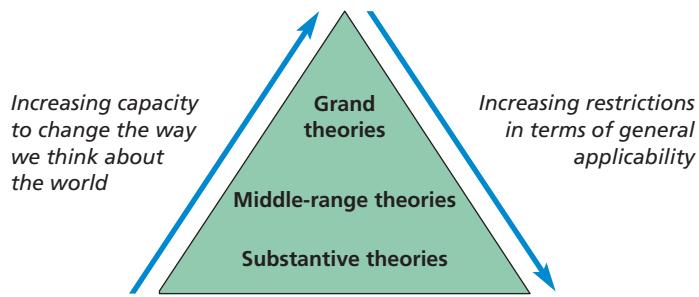


Figure 2.1 Grand, middle-range and substantive theories

For example, studying the implications of a cost-saving strategy in a particular organisation would be an example of a substantive theory. Although they may be restricted, a host of ‘substantive theories’ that present similar propositions may lead to ‘middle-range theories’. By developing ‘substantive theories’, however modest, we are doing our bit as researchers to enhance our understanding of the world about us. A grand claim, but a valid one!

Another way to examine the theoretical contributions of research into business and management is to assess its practical usefulness for organisations and those who work in them (Box 2.10). This is beautifully stated by Lewin (1945: 130) as ‘nothing is as practical as a good theory’.



Box 2.10 Focus on management research

What makes a theoretical contribution?

Corley and Gioia’s (2011) study found that the theoretical contribution of management research can be measured along two dimensions. One of these relates to what they call the ‘originality’ of the contribution. This they divide into a contribution that is either ‘incremental’ or ‘revelatory’. An incremental theoretical contribution is one that adds to or builds on a theory, perhaps by applying it in a new context. A revelatory theoretical contribution is more profound, offering a new theory to make sense of a problem or explain a phenomenon. The other dimension relates to what they call the ‘utility’ or ‘usefulness’ of a contribution. This they divide into a contribution that has ‘scientific usefulness’ or ‘practical usefulness’. A scientifically useful contribution is one that emphasises methodological rigour and usefulness to an academic audience. A practically useful contribution emphasises

organisational application and usefulness to organisational practitioners.

However, while theoretical contributions on the first dimension are likely to be exclusively either ‘incremental’ or ‘revelatory’ (it’s unlikely that a contribution could be both!), this does not have to be the case on the second dimension. Corley and Gioia focus much of their article on discussing how to achieve research that is capable of being both academically and practically useful. They refer to designing research that has scope to be theoretically relevant to both academics and organisational practitioners. They consider ways in which this type of theoretical contribution may be achieved. This includes a continuing emphasis on examining the links between theoretical abstractions and practice implications. In interpreting their results, researchers also need to go beyond narrow generalisations and look for insights that can inform organisational practice. In a similar way, when developing theory, researchers need to look not only at the validity of their theories, but also their usefulness and applicability. Researchers also need to exercise some foresight when choosing their research topics so they pursue research that has, and will have, relevance to the problems and issues faced by organisations and organisational practitioners.

2.8 The need for a research proposal

The **research proposal**, occasionally referred to as a protocol or outline, is a structured plan of your proposed research project. In this section we discuss why it is necessary; but it is important to recognise that a competent research proposal will draw on material discussed in subsequent chapters. Before you can write your research proposal, you will need to be aware of available literature and appropriate theory (Section 2.7 and Chapter 3), the research philosophy and approach that you wish to use (Chapter 4), your research design including methodological choice, research strategy and time frame (Chapter 5), access and ethical issues (Chapter 6), sample selection (Chapter 7), data collection methods and data analysis techniques (Chapters 8–13).

Providing a clear specification to guide your research project

Your research project is likely to be a large element in your course. It is also yours! You will be responsible for conceiving, conducting and concluding this project and creating the report, dissertation or thesis. From this perspective, developing a research proposal offers you the opportunity to think carefully about your research project (Box 2.11). We do not suggest that you use the questions in Box 2.11 as headings under which you write responses, but we feel that they should be helpful as a guide and as a checklist against which to evaluate your research proposal before submitting it to your tutor. A well-thought-out and well-written research proposal has the potential to provide you with a clear specification of the what, why, how, when and where of your research project.

Developing a research proposal is demanding: thinking through what you wish to do and why, identifying and synthesising literature and then envisaging all of the stages of your research will be time consuming, as will the necessary revisions to create a coherent and clearly written proposal. However, the effort is likely to prove to be very worthwhile. As you juggle several activities during the period of your research project, there may be occasions when you pick up your research proposal and feel glad that you spent so much time producing a clear specification to guide your project through its various stages.

Ensuring assessment criteria can be met

It is likely that your research proposal will be assessed before you are allowed to carry on with your research project. A proportion of the overall marks available for your project report may be given for the research proposal. Alternatively, a research proposal may be subject to approval before you are permitted to proceed with your research project. In either case, it will be necessary to reach a certain standard before being allowed to progress. There are potentially a number of different criteria that may be used to assess a research proposal. These may include criteria that are specific to each of the components of the proposal, which we describe in Section 2.9. Part of the assessment and approval process may also centre on criteria that focus on more general concerns. We first consider three such criteria that are likely to be used to assess your research proposal: coherence, ethical considerations and feasibility.



Box 2.11 Checklist to guide and evaluate your research proposal

- ✓ Have I explained what am I going to do?
- ✓ Have I explained why I am doing this?
- ✓ Have I said why it is worth doing?
- ✓ Have I explained how it relates to the research that has been done before in the subject area?
- ✓ Have I stated which theory or theories will inform what I am doing and how I will use it or them?
- ✓ Have I stated my overarching research question(s), research aim and my research objectives or investigative questions?
- ✓ Have I outlined how I will conduct my research?

- ✓ Have I outlined my research design?
- ✓ Have I outlined what data I need?
- ✓ Have I stated who and where my intended respondents or participants are?
- ✓ Have I explained how I will select my respondents or participants?
- ✓ Have I explained how I will gain access?
- ✓ Have I outlined how I will collect my data?
- ✓ Have I outlined how I will analyse my data and use this to develop theoretical explanations?
- ✓ Have I outlined what data quality issues I might encounter?
- ✓ Have I outlined how I will seek to overcome these data quality issues?
- ✓ Have I considered the ethical issues I might encounter at each stage of my research?
- ✓ Have I outlined how I will address these?

Coherence

Undertaking research is a complex and time-consuming activity. As we indicated earlier, you are likely to benefit from creating a clear specification to guide your research project. Your project tutor and any other assessor will be looking for evidence of coherence and lucidity in the way you have written your research proposal, to demonstrate that it will be fit for purpose and able to direct your research activity.

Ethical considerations

Part of the approval process for your research proposal may involve it being considered and given a favourable opinion by a research ethics committee. Your university's code of ethical practice is likely to require all research involving human participants to be considered and 'approved', especially where research involves young or vulnerable participants. You will also probably have to state how data will be stored, whether they will be kept after the research is completed and under what conditions, in order to ensure the continuing anonymity of the participants and confidentiality of their data. Section 6.5 discusses ethical issues related to the design stage of a research project. You will need to be aware of and abide by the ethical requirements of your university. These requirements will add to the time that you will need to allow for the planning stage of your research project. As a professional student you may also need to be aware of and abide by the ethical requirements of your professional institute.

Feasibility

You may have devised a coherent and well-structured research proposal that would create much interest but it may not be possible to achieve, or sensible to contemplate. Feasibility is a multifaceted criterion that your assessors will be concerned about. Your proposal may not be possible to achieve in the time available to undertake the research project and produce your dissertation or management report. It may be that data collection would not be possible because you would not be able to gain access to participants, or it might not be practical and your tutor will tell you so! The proposal may require resources that

are not available, finance commitments that are unaffordable, or skills that you have not developed and would not be able to acquire in the timescale of the project.

It is always helpful to discuss your research proposal with a tutor. Where there are concerns about any of the issues just considered, it will be possible to discuss these to work out how the proposed research may be amended. For example, in relation to feasibility something more modest in scope may be discussed. Your task will then be to amend initial ideas and convince your tutor that the proposed research is achievable within the time and other resources available.

Ensuring that your research project isn't based on preconceived ideas

Your research project offers a valuable way to learn the skills involved in this activity. These skills are transferable to many other situations, including the world of work. It is about process as well as outcome. Concerns about feasibility (related to over enthusiasm) lie at one end of a continuum, at the other end of which lies a very occasional concern about sincerity. Do not be like the student who came to Phil to talk over a research proposal and said, ‘Of course, I know what the answer will be’. When asked to explain the purpose of doing the research if he already knew the answer, he became rather defensive and eventually looked for another supervisor and, probably, another topic.

Approval of your research proposal implies that if it is followed, the research is likely to be satisfactory. While this is no guarantee of subsequent success, it will reassure you to know that you have started your research journey with an appropriate destination and journey plan. It will be for you to ensure that you do not get lost!

2.9 Structuring your research proposal

There are potentially different ways to structure your research proposal. Different philosophical traditions and approaches to theory development (Chapter 4) will involve a variety of research designs (Chapter 5), and may lead to different ways of structuring your proposal and, as you write up your research, your project report (Chapter 14). We describe what many think of as the standard approach to structuring your research proposal. You will therefore need to check if your assessment criteria require a different structure. Whichever structure you are required to adopt, this will need to be driven by and focused on your research question, aim and research objectives, and you will need to ensure that you produce a coherent proposal.

Title

The title should simply and concisely summarise the research question. It should avoid unnecessary phrases such as, ‘A study to explore . . .’ Instead it should reflect the concepts or variables in your research question (Box 2.12). If your research question changes, this will naturally lead to a change to your title.

Background

This section has three related functions which are to:

- introduce the reader to the research issue or problem;
- provide a rationale for your overarching research question and aim;
- ground your research in the academic literature.



Box 2.12 **Focus on student research**

Devising research proposal titles

Imran (Box 2.6) reworded his research question into the following title for his research proposal:

Reasons for a lack of fit between corporate strategy and the external environment.

Diane (Box 2.7) devised this title for her research proposal by rewording her research question:

An evaluation of employee engagement, commitment and development, and scope for their promotion, following organisational change.

Justine (Box 2.9) used her research question to develop this title for her proposal:

The effect of 'healthy eating' publicity on snack foods purchasing decisions.

Introducing the reader to the research issue addresses the question, 'what am I going to do?' You also need to provide a rationale for your overarching research question and research aim and justify this. This may be composed of two elements, one relating to you and the other relating to the value of the work and referencing the literature. Your reader will be looking for some evidence that this is a topic in which you have sufficient interest to sustain the effort that will be required from you over the period of the research project. This may relate to the need to tackle a problem, to your intellectual curiosity or to your intended career direction. It relates to the question, 'why am I going to do this?' The rationale will also need to address the question, 'why is it worth doing?' This will relate to one of the following types of justification using the academic literature: the application of a theory to a particular context (such as within an organisation); the development of a theory within a research setting; testing a theory within a given context. Your research may propose other such justifications depending on its nature.

Grounding your research in the academic literature means demonstrating how your research relates to relevant previous research. In achieving this you will show your knowledge of relevant literature and clarify where your proposal fits into the debate in this literature (Section 3.3). You will also be able to begin to show which theory or theories will inform what you are doing and show they will be used. The intention will be not to write a detailed review of the literature but rather to provide an overview of key literature sources from which you will draw and the theory or theories within them. This will not be the same as the critical literature review (Sections 3.2 and 3.3) that you will present in your final project report but the start of the process that leads to it.

Overarching research question(s), research aim and objectives or investigative questions

The Background section should lead logically into a statement of your research question(s), research aim and, usually, either research objectives or investigative questions. These should leave the reader in no doubt about what your research seeks to achieve. Be careful here to ensure that your objectives are precisely written and will lead to observable outcomes (Box 2.7).

Method

The Background and Method will be the longest sections of your proposal. The Method section will explain how you will undertake your research. It may be divided into subsections that deal with research philosophy, research design, participants, techniques and

procedures, and ethical considerations. This final element may need to be dealt with in a discrete section of your research proposal.

Research philosophy is discussed in Chapter 4. It involves you in being clear about your own research philosophy, how this may impact on your subsequent research design and ensuring these are consistent with subsequent decisions regarding your method. Research design is discussed in Chapter 5. You will need to make a methodological choice between a quantitative, qualitative or mixed methods design. You will also need to select one or more research strategies (e.g. an experiment, a case study, a survey, a Grounded Theory strategy), and determine an appropriate time frame for your project depending on the nature of your research. You will need to describe each of these and justify your choice by the way these elements fit together to form a coherent whole and are consistent with your research philosophy.

How you design your research will affect the type of data you require, where you intend to locate them and from whom you will collect them. Your data may be collected from human participants, or they may be secondary data (Chapter 8) such as from archival research (Section 5.8) or a combination of these. You will therefore need to be clear regarding type of data you need. If you are using secondary data you will need to explain what these are, where they are located, any issues related to access and justify this choice. If you intend to collect data from people, you will need to be clear who these are likely to be. You may be intending to conduct research in a single organisation or across a number of organisations. You will need to explain and justify the nature of the organisation or organisations and possibly the sector or sectors within which it, or they, operate. The people from whom you intend to collect data may be located within a specific part of an organisation or be drawn from across it. You will need to explain and justify this.

You will also need to explain the nature of your research population and why you chose it. For example, they may be entrepreneurs, managerial employees, non-managerial employees, a particular occupational group, trade union officials or some combination of these. Where you need to select a sample from within a research population you will need to explain how they will be selected. Chapter 7 discusses types of probability and non-probability sampling and you will need to describe and justify your sampling technique(s) and sample size.

You will also need to describe the data collection and analysis techniques you intend to use and how these will enable you to develop theoretical explanations. Data collection techniques include examination of secondary data, observation, questionnaires, interviews and diaries (Chapters 8 to 11). You will not need to explain the precise details of each technique you intend to use, such as including a copy of your questionnaire, an interview checklist or the content of an observation schedule, but you will need to describe how you will use it. For example, if you are using interviews, the structure, mode and medium you will use, how many you will conduct, their intended duration and how you will record the data. You will also need to describe, albeit briefly, how you intend to analyse each type of data that you collect.

It will also be important to discuss ethical considerations so that you anticipate these and demonstrate to your tutor and ethics committee that your research design and proposal have been formulated to minimise ethical concerns and avoid unethical practice. This will be essential where you are dealing with human participants, and sometimes even if using secondary data already collected from human participants. There may be a reduced need for some of you undertaking certain types of research (e.g. where this is based on macro-level, completely anonymised data) but in nearly all cases this requirement is very likely to mean that you need to be sensitive to ethical concerns.

Timescale

It is very helpful to divide your research project into its constituent stages or tasks. You may estimate the amount of time that each stage or task should take to complete. Allocating each stage or task so much time should help you and your tutor decide on the feasibility of the research project, by giving you a clear idea as to what needs to be achieved during the time allowed. Experience shows that however well the researcher's time is organised, the whole process seems to take longer than anticipated. Devising a timescale allows you to monitor your progress and indicates where you need to allocate more working hours to keep up with your intended schedule (Box 2.13).

Many researchers use a **Gantt chart** to produce a schedule for their research project. Developed by Henry Gantt in 1917, this provides a simple visual representation of the stages or tasks that make up your research project, the timings to be allocated to each of these and the relationship between them. It is a simple but effective tool used in various types of project management. In a grid of columns and rows, tasks are listed under each other in the first column. Each row therefore starts with a short description of a task and the remainder of the row indicates a timescale (Figure 2.2).

The time estimated that each task will take is represented by the length of its associated horizontal bar, while each task's start time and finish time is indicated by the beginning and end of the bar. As we can see from the first bar of the chart in Figure 2.2, the student has decided to schedule in two weeks of holiday. The first of these occurs over the Christmas and New Year period, and the second occurs while their tutor is reading a draft copy of the completed project in April. We can also see from the second and fourth bar that, like many of our students, they intend to begin to draft their critical literature review while still reading new articles and books. However, they have also recognised that some activities must be undertaken sequentially. For example, bars 9 and 10 highlight that before they can administer their questionnaire (bar 10) they must complete all the revisions highlighted as necessary by the pilot testing (bar 9). Finally, this student has noted that their project assessment criteria include a reflective essay and has decided to keep a reflective diary throughout the research project (bar 20).

Resources

These are another facet of feasibility (Box 2.1 and also our earlier discussion in this section). Including this discussion in your research proposal will allow you and your tutor to assess whether what you are proposing can be resourced. Resource considerations may be categorised as finance, data access and equipment.

Conducting research costs money. This may include, for example: travel, subsistence, help with transcription or, perhaps, postage for mailed questionnaires. Think through the expenses involved and ensure that you can meet them.

Assessors of your proposal will need to be convinced that you have access to the data you need to conduct your research (Sections 6.2 and 6.3). This may be unproblematic if you are carrying out research in your own organisation. Many project tutors wish to see written approval from host organisations in which researchers are planning to conduct research. You will also need to convince your reader of the likely response rate to any questionnaire that you send.

It is surprising how many research proposals have ambitious plans for large-scale data collection with no thought given to how the data will be analysed. It is important that you convince the reader of your proposal that you have access to the necessary computer software to analyse your data. Moreover, it will be necessary for you to demonstrate that

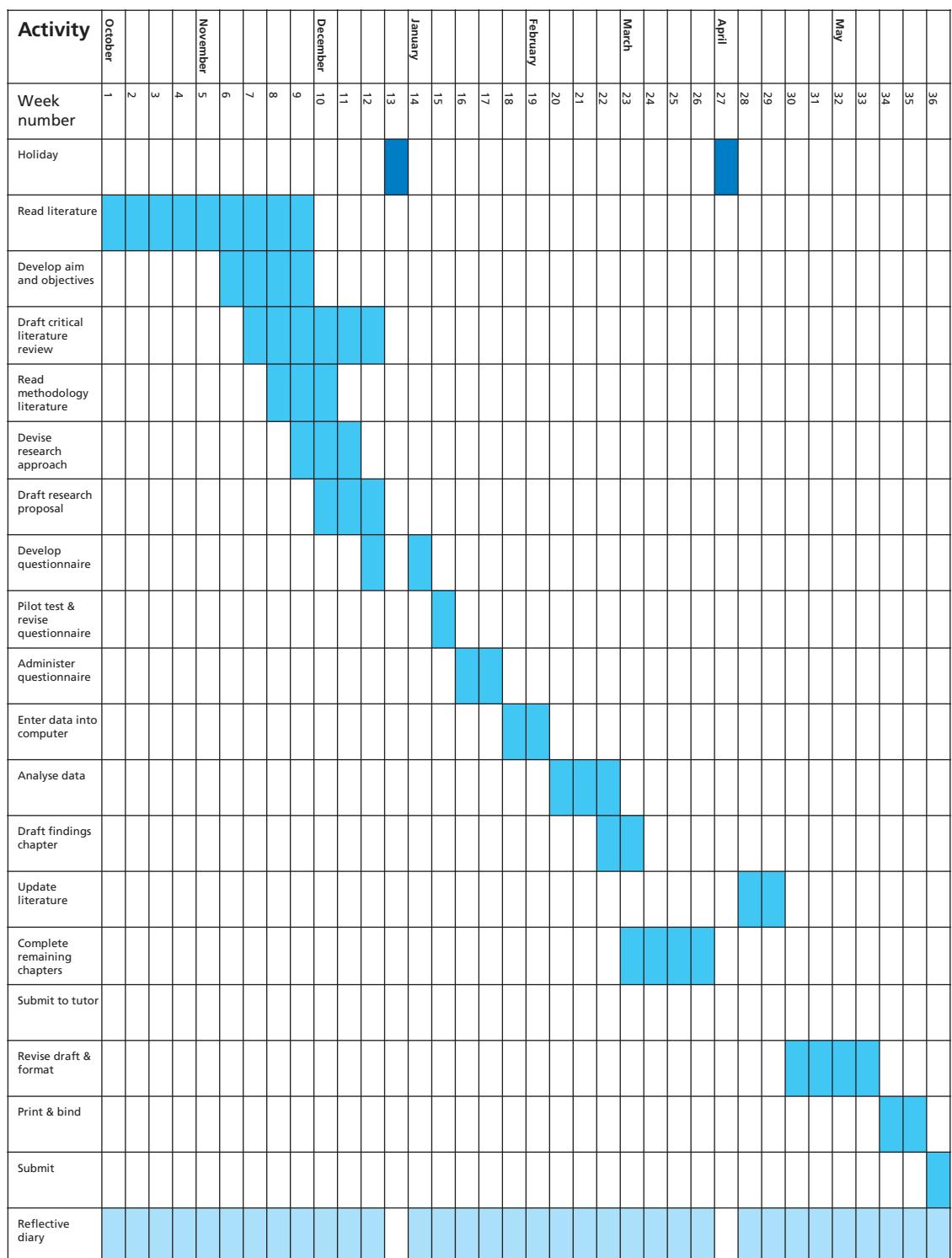


Figure 2.2 Gantt chart for a student's research project



Box 2.13 Focus on student research

Louisa's research timescale

As part of the final year of her undergraduate business studies degree, Louisa had to undertake an 8000–10,000-word research project. In order to assist her with her time management, she discussed the following task list, developed using Microsoft task and reminder setting app 'To Do', with her tutor, noting she had still to fully develop her overarching research question.

The screenshot shows a Microsoft To Do task list titled 'Planned'. The tasks are listed by due date:

- Fri 7 Oct: Develop clear overarching research question
- Fri 25 Nov: Read and noted literature
- Fri 16 Dec: Draft literature review leading to clear research objectives emailed to project tutor
- Fri 23 Dec: Read methods literature for both primary and secondary data collection
- Fri 10 Feb: Literature review revised in response to project tutor's comments
- Fri 3 Mar: Secondary and primary data collected and analysed to meet research objectives
- Fri 28 Apr: Analysis completed and findings chapters written
- Fri 26 May: Draft of project report completed (everything!) and sent to tutor
- Fri 9 Jun: Draft revised in response to project tutor's comments
- Mon 12 Jun: Project submitted (11:59 a.m. deadline)

you have either the necessary skills to perform the analysis or can learn the skills in an appropriate time, or you have access to help.

References

It is not necessary to try to impress your proposal reader with an enormous list of references. A few key literature sources to which you have referred in the background section and which relate to the previous work and theory that directly informs your own proposal, as well as references to the methods literature, should be all that is necessary. We provide more detail on how to reference in Appendix 2.

2.10 Summary

- Generating a research idea and developing your research proposal are key parts of your research project.
- Characteristics of a good research topic include appropriateness, capability and fulfilment. However, the most important is that it will meet the requirements of the examining body.

- Generating and refining research ideas makes use of a variety of techniques. It is important that you use a variety of techniques, including those involving rational thinking and those involving creative thinking.
- Further refinement of research ideas may be achieved through using a Delphi technique, conducting a preliminary inquiry and integrating ideas by working these up and narrowing them down.
- A clearly defined overarching research question expresses what your research is about and will become the focal point of your research project.
- A research aim is a brief statement of the purpose of the research project. It is often written as a sentence stating what you intend to achieve through your research.
- Well-formulated research objectives or investigative questions operationalise how you intend to conduct your research by providing a set of coherent and connected steps to answer your research question.
- It will be important to use academic theory to inform your research irrespective of the approach you will use to conduct your research project.
- A research proposal is a structured plan for your proposed research project.
- A well-thought-out and well-written research proposal with a clear time frame has the potential to provide you with a clear specification of the what, why, how, when and where of your research project.

Self-check questions

Help with these questions is available at the end of the chapter.

- 2.1** You have decided to search the literature to 'try to come up with some research ideas in the area of operations management'. How will you go about this?
- 2.2** A colleague of yours wishes to generate a research idea in the area of accounting. They have examined their own strengths and interests on the basis of their assignments and have read some review articles but have failed to find an idea about which they are excited. They ask you for advice. Suggest two techniques that your colleague could use and justify your choices.
- 2.3** You are interested in undertaking some research on the interface between business organisations and schools. Write three research questions that may be appropriate.
- 2.4** For the workplace project for her professional course, Karen had decided to undertake a study of the effectiveness of the joint negotiating and consultative committee in her NHS Trust. Her title was 'An evaluation of the effectiveness of the Joint Negotiating and Consultative Committee in Anyshire's Hospitals NHS Foundation Trust'. Draft some objectives that Karen may adopt to complement her title.
- 2.5** How may the formulation of an abstract concept with clear links to theory help in the development of a research proposal?
- 2.6** How would you demonstrate the influence of relevant theory in your research proposal?

Review and discussion questions

- 2.7** Together with a few colleagues discuss the extent to which a number of research ideas would each constitute a 'good research idea' using the checklist in Box 2.1. The set of ideas you use may be past project titles obtained from your tutor that relate to your course. Alternatively, they may be those that have been written by you and your colleagues as preparation for your project(s).

- 2.8 Look through several of the academic journals that relate to your subject area. Choose an article that is based upon primary research. Assuming that the research aim and objectives are not made explicit, infer from the content of the article what the aim and objectives may have been.
- 2.9 Watch the news on television or access a news website. Look for a news item based on research that has been carried out to report a current issue related to business. Spend some time investigating other news websites (e.g. <http://www.news.google.com>) to learn more about the research that relates to this business news story. Study the story carefully and decide what further questions the report raises. Use this as the basis to draft an outline proposal to seek answers to one (or more) of these questions.



Progressing your research project

Choosing a research topic and developing your research proposal

- If you have not been given a research idea, consider the techniques available for generating and refining research ideas. Choose a selection of those with which you feel most comfortable, making sure to include both rational and creative thinking techniques. Use these to try to generate a research idea or ideas. Once you have got a research idea(s), or if you have been unable to find an idea, talk to your project tutor.
- Evaluate your research idea(s) against the characteristics of a good research project (Box 2.1).
- Refine your research idea(s) using a selection of the techniques available for generating and refining research idea(s). Re-evaluate your research ideas against the characteristics of a good research project (Box 2.1). Remember that it is better to revise (and in some situations to discard) ideas that do not appear to be feasible at this stage. Integrate your ideas using the process

of working up and narrowing down to form one research idea.

- Use the AbC rule to help write an overarching research question based on your research idea. Where possible this should be a 'how?' or a 'why?' rather than a 'what?' question.
- Refine this overarching research question and write a research aim and a set of connected research objectives or investigative questions.
- Write your research proposal making sure it includes a clear title and sections on:
 - the background to your research;
 - your overarching research question(s), related aim and research objectives or investigative questions;
 - the method you intend to use including research design, sample, data collection techniques and analysis procedures, and ethical considerations;
 - the timescale for your research;
 - the resources you require;
 - references to any literature to which you have referred.
- Use the questions in Box 1.4 to guide your reflective diary entry.

References

- Alvesson, M. and Sandberg, J. (2011) 'Generating research questions through problematization', *Academy of Management Review*, Vol. 36, No. 2, pp. 247–71.
- Brady, S.R. (2015) 'Utilizing and adapting the Delphi method for use in qualitative research', *International Journal of Qualitative Methods*, Vol. 14, pp. 1–6.
- Buzan, T. (2018) *Mind Map Mastery: The complete guide to learning and using the most powerful tool in the universe*. (new edition). London: Watkins Publishing.

- Carroll, L. (1989) *Alice's Adventures in Wonderland*. London: Hutchinson. (Originally published by Macmillan and Co. in 1866).
- Chaurasia, S., Pati, R.K., Padhi, S.S., Jensen, J.M.K. and Gavirneni, N. (2021) 'Achieving the United Nations Sustainable Development Goals-2030 through the nutraceutical industry: A review of managerial research and the role of operations management', *Decision Sciences* (early view) pp. 1–16.
- Clough, P. and Nutbrown, C. (2012) *A Student's Guide to Methodology* (3rd edn). London: Sage.
- Corley, K.G. and Gioia, D.A. (2011) 'Building theory about theory building: What constitutes a theoretical contribution?', *Academy of Management Review*, Vol. 36, No. 1, pp. 12–32.
- Dudau, A.I. (2016) 'The research topic identification.' YouTube. Available at <https://www.youtube.com/watch?v=Yn-OdyISUPU> [Accessed 1 Sept. 2020].
- Festinger, L. (1957) *A Theory of Cognitive Dissonance*. Stanford, CA: Stanford University Press.
- Gill, J. and Johnson, P. (2010) *Research Methods for Managers* (4th edn). London: Sage.
- Herzberg, F., Mausner, B. and Snyderman, B.B. (1959) *The Motivation to Work*. New York: John Wiley.
- Jankowicz, A.D. (2005) *Business Research Projects* (4th edn). London: Thomson Learning.
- Kelly, G.A. (1955) *The Psychology of Personal Constructs*. New York: Norton.
- Lewin, K. (1945) 'The research centre for group dynamics at the Massachusetts Institute of Technology', *Sociometry*, Vol. 8, No. 2, pp. 126–36.
- Maslow, A.H. (1943) 'A theory of human motivation', *Psychological Review*, Vol. 50, No. 4, pp. 370–96.
- McDonald, S. (2005) 'Studying actions in context: A qualitative shadowing method for organisational research', *Qualitative Research*, Vol. 5, No. 4, pp. 455–73.
- Nix, E. (2018) 'Did an apple really fall on Isaac Newton's head?' *History Stories*. Available at <https://www.history.com/news/did-an-apple-really-fall-on-isaac-newton-s-head#:~:text=Legend%20has%20it%20that%20a,with%20his%20law%20of%20gravity.&text=occasion'd%20by%20the%20fall,sat%20in%20a%20contemplative%20mood.%E2%80%9D> [Accessed 9 May 2021].
- Ogbonna, E. and Harris, L.C. (2014) 'Organizational cultural perpetuation: A case study of an English Premier League football club', *British Journal of Management*, Vol. 25, No. 4, pp. 667–86.
- Podsakoff, P.M., MacKenzie, S.B. and Podsakoff, N.P. (2016) 'Recommendations for creating better concept definitions in the organizational, behavioral and social sciences', *Organizational Research Methods*, Vol. 19, No. 2., pp. 159–203.
- Saunders, M.N.K., Gray, D.E, Tosey, P. and Sadler-Smith, E. (2015) 'Concepts and theory building', in L. Anderson, J. Gold, J. Stewart and R. Thorpe, (eds) *Professional Doctorates in Business and Management*. London: Sage, pp. 35–56.
- Saunders, M.N.K. and Lewis, P. (1997) 'Great ideas and blind alleys? A review of the literature on starting research', *Management Learning*, Vol. 28, No. 3, pp. 283–99.
- Schootstra, E., Deichmann, D. and Dolgová, E. (2017) 'Can 10 minutes of meditation make you more creative?' *Harvard Business Review*, 29 August, pp. 2–5.
- Sutton, R. and Staw, B. (1995) 'What theory is not', *Administrative Science Quarterly*, Vol. 40, No. 3, pp. 371–84.
- Tietze, S. (2012) 'Researching your own organisation', in G. Symon and C. Cassell (eds) *Qualitative Organizational Research*. London: Sage.
- Whetten, D. (1989) 'What constitutes a theoretical contribution?', *Academy of Management Review*, Vol. 14, No. 4, pp. 490–5.

Further reading

Alvesson, M. and Sandberg, J. (2011) 'Generating research questions through problematization', *Academy of Management Review*, Vol. 36, No. 2, pp. 247–71. This article discusses the established way in which research questions are generated by researchers and how this approach can be challenged. While the way in which you generate your research question is likely to be related to the established way they discuss, reading this will provide you with a deeper understanding of research questions and their relationship to theory.

Podsakoff, P.M., MacKenzie, S.B. and Podsakoff, N.P. (2016) 'Recommendations for creating better concept definitions in the organizational, behavioral and social sciences', *Organizational Research Methods*, Vol. 19, No. 2., pp. 159–203. This is a significant paper that discusses the importance of concepts and conceptual clarity in undertaking research and their role in theory. It is well worth reading this paper and considering its guidance as you develop your research proposal.

Saunders, M.N.K., Gray, D.E., Tosey, P. and Sadler-Smith, E. (2015) 'Concepts and theory building' in L. Anderson, J. Gold, R. Thorpe and J. Stewart (eds) *Professional Doctorates in Business and Management*. Los Angeles, Sage, pp. 35–56. This chapter is written from the viewpoint that theory is an essential component of all research, being both practical and useful. It will be particularly useful for those undertaking a research project for an organisation.

Sutton, R. and Staw, B. (1995) 'What theory is not', *Administrative Science Quarterly*, Vol. 40, No. 3, pp. 371–84. This is a helpful article to read to gain some insights into the role of theory if you find this aspect daunting. In telling us what theory is not, they provide a very helpful discussion about what it is by referring to their own experiences. They also go further than this and evaluate the role of theory.

Formulating the Research Design

Saunders et al. (2023; 176-228)

Chapter 5



Formulating the research design

Learning outcomes

By the end of this chapter you should be able to:

- recognise the importance of your decisions when designing research and the need to achieve methodological coherence throughout your research design;
- identify the differences between exploratory, descriptive, explanatory and evaluative research, and recognise the purpose(s) of your research design;
- distinguish and choose between quantitative, qualitative and mixed methods research designs;
- develop an appropriate research strategy or strategies and achieve coherence throughout your research design;
- consider the implications of the time frames required for different research designs;
- identify the main ethical issues implied by your research design;
- utilise appropriate criteria to evaluate the quality of your research design;
- recognize your role as researcher in your research design;
- progress your research project by formulating your research design.

5.1 Introduction

In Chapter 4 we introduced the research onion as a way of depicting the issues underlying your choice of data collection method or methods and peeled away the outer two layers – research philosophy and approach to theory development. In this chapter we uncover the next three layers: methodological choice, research strategy or strategies, and the time horizon for your research. As we saw in Chapter 4, the way you answer your research question will be influenced by your research philosophy and approach to theory development. Your research philosophy



and approach to theory development will, whether this is deliberate or by default, invariably influence your selections shown in the next three layers of the research onion (Figure 5.1). These three layers can be thought of as aspects of research design, which is the way you turn your overarching research question or aim and your objectives into a research project. The key to these selections will be achieving a coherent design that is consistent with your research philosophy and approach to theory development.

Your research design is the overall plan for your research project. The tactics, through which you will bring your plan to fruition, are precisely how you will collect and analyse the data (the centre of the research onion). They comprise the detailed design and operationalisation of your data collection procedures and analysis techniques, and are discussed in later chapters.

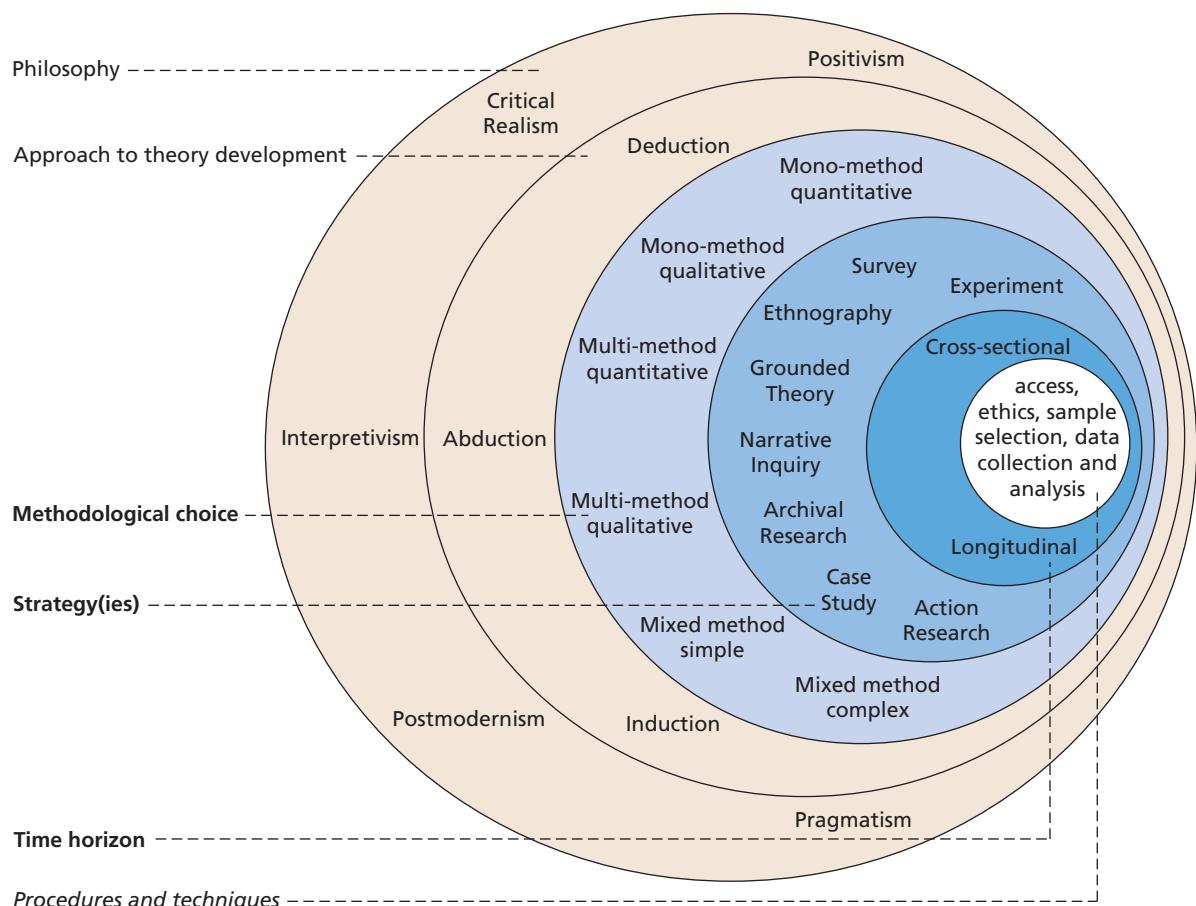


Figure 5.1 The research onion

Source: © 2022 Mark NK Saunders; developed from Saunders et al., 2019.

The research process is like a journey

The cover photographs of recent editions of this book have shown that the research process as a journey. Like many such journeys, there is generally a choice of paths or roads to take. When you are thinking about setting out on a car journey of some distance, you will probably enter the destination into your satnav and consider possible route options to get to your destination. A number of criteria will influence your decision about which route to take, including time, fuel economy and your preferences for avoiding motorways, ferries and toll roads. The route you choose will be calculated by the satnav to meet your given preferences. As you undertake your journey, you will find yourself interacting with the reality of your planned route. Some parts of the journey will go according to plan; other parts may not, and you may need to amend your route, perhaps because of traffic congestion or a road being closed due to roadworks. In many ways, designing research is like planning a journey. Formulating the most appropriate



Source: © Mark Saunders 2021

design to answer your research question is similar to planning the route to your destination. The research aim is your destination and the research objectives are your route criteria. Your research design is your route and it needs to enable you to reach your destination and crucially be consistent with your objectives. Like your route, your research design may need to be amended due to unforeseen circumstances. Both will be interactive experiences.

5.2 Achieving a coherent research design

Your **research design** is the plan of how you will go about answering your research question, achieving your research aim and meeting your objectives. It will specify the source or sources from which you intend to collect data, how you propose to collect and analyse these, and discuss ethical issues and the constraints you will inevitably encounter (e.g. access to data, time, location and money). Crucially, it will demonstrate that you have thought through the elements of your particular research design.

The purpose of your research design will be driven by your overarching research question and will be either exploratory, descriptive, explanatory, evaluative or a combination of these (Section 5.3). In planning your research, you will need to decide whether you follow a quantitative, qualitative or mixed methods research design. Each of these methodological choices will require a careful consideration of strategies and time horizons along with data collection procedures and analysis techniques, to achieve a coherent research design. We introduce this methodological choice in Section 5.4 considering quantitative, qualitative and mixed methods research designs. This will be operationalised through one or more research strategies (Section 5.5), which need to be consistent with

your research philosophy and your methodological choice. Your research questions will also determine the selection of an appropriate time horizon (Section 5.6). Your research design will need to be ethical, and it will be important to identify potential concerns and minimise or overcome them. We consider the importance of anticipating potential ethical concerns briefly in Section 5.7, discussing them in greater detail in Sections 6.5 and 6.6. It is also important to ensure the quality of your research design, and we discuss criteria to evaluate this in Section 5.8. Finally, we recognise that your own role as researcher will affect your research design in Section 5.9.

Your research design is likely to be assessed by your university or examining institution as part of your research proposal and you will need to achieve a pass before you are allowed to proceed. You therefore need to produce a clear and coherent design with valid reasons for each of your research design decisions, even if your design changes subsequently. Your justification for each element in your research design should be based on your research question(s) and objectives, and be consistent with your research philosophy.

5.3 The research purpose

In Chapter 2 we encouraged you to think about your research project in terms of an overarching research question, a research aim and objectives. Your research question indicates whether the purpose is exploratory, descriptive, explanatory or evaluative or a combination of these. In this section we discuss each purpose in more detail to help you evaluate the purpose of your own research study.

Exploratory studies

An **exploratory study** explores or clarifies understanding of an issue, problem or phenomenon (Box 5.1). The overarching research question is likely to start with 'What' or 'How' (Section 2.4). Questions that you ask during data collection to explore an issue, problem or phenomenon will also be likely to start with 'What' or 'How' (Chapters 10 and 11).

Exploratory research has the advantage that it is flexible and adaptable to change. If you are conducting exploratory research, you must be willing to change your direction as a result of new data that appear and new insights that occur to you. A quotation from the travel writer V.S. Naipaul (1989: 222) illustrates this point beautifully:

I had been concerned, at the start of my own journey, to establish some lines of enquiry, to define a theme. The approach had its difficulties. At the back of my mind was always a worry that I would come to a place and all contacts would break down . . . If you travel on a theme the theme has to develop with the travel. At the beginning your interests can be broad and scattered. But then they must be more focused; the different stages of a journey cannot simply be versions of one another. And . . . this kind of travel depended on luck. It depended on the people you met, the little illuminations you had. As with the next day's issue of fast-moving daily newspapers, the shape of the character in hand was continually being changed by accidents along the way.

Exploratory research may commence with a broad focus, but this will become narrower as the research progresses. It may be that time spent on exploratory research might show that the research is not worth pursuing!



Box 5.1 Focus on management research

An exploratory study combining research methods

Research by Sun et al. (2021: 28) published in the *Journal of Marketing*, focussing on the clothing and accessories industries, explores three aspects of sustainable luxury consumption. These comprise three questions:

- 1 'Whether high-end products are more sustainable by virtue of their longer product lifecycles?
- 2 How consumers process information regarding the durability of these high-end products?
- 3 How marketeers can help consumers overcome a failure to consider product durability and promote the purchase of fewer, higher-end products that will last longer? (Sun et al., 2021: 29)'

To answer these questions data were collected using multiple quantitative methods. These included:

- An automatic web crawler scraping price, brand and detailed product category data from 20 online retailers selling new and second-hand shoes and handbags for 4,600 products. This provided evidence of whether high-end goods are more durable, and hence more sustainable.
- An online survey of 1,800 United States (US) Amazon Mechanical Turk (MTurk) crowdsourcing

platform respondents to classify these brands as high, middle or low-end products.

- An online survey using the Qualtrics platform of 340 wealthy US women who answered questions about their own belongings to find further support regarding high-end goods being more sustainable than low-end goods.
- A paid online survey of 201 US respondents using MTurk, collecting data about two products at different price points and different time horizons. This investigated why consumers prefer multiple mid-range products over a high-end product, neglecting product durability.
- A paid online survey of 421 US respondents using Prolific Academic crowdsourcing platform, collecting data about responses to product information for a fictitious high-end and mid-end item. This further established neglect of product durability.
- A survey completed by 162 US graduate students for course credit, collecting data on their choices in relation to a range of attributes for an actual high-end product. This evaluated consumers' preferences for durability in relation to specific trade-offs relative to product attributes such as price and style.
- A survey of 106 real consumers using a US clothing company's email list. This collected data on particular attributes of a product to determine whether durability could be framed as a dimension of sustainability.

Descriptive studies

A **descriptive study** is designed to gain an accurate profile of events, persons or situations. One of the earliest well-known examples of a descriptive survey is the *Domesday Book*, which described the population of England in 1085. As we noted in Section 2.4, research questions that are descriptive are likely to begin with, or include, 'Who', 'What', 'Where', 'When' or 'How'. Questions that you ask during data collection are also likely to start with, or include, 'Who', 'What', 'Where', 'When' or 'How' (Chapters 10 and 11). A descriptive study can extend an exploratory study or contextualise an explanatory study. However, it is necessary to have a clear picture of the phenomenon on which you wish to collect data prior to the collection of the data.

Project tutors are often wary of work that is too descriptive. There is a danger of their saying 'That's very interesting . . . but so what?' They will want you to go further and draw conclusions from the data you are describing. They will encourage you to develop

the skills of evaluating data and synthesising ideas. These are higher-order skills than those of accurate description. Description in business and management research has a very clear place. However, it should be thought of as a means to an end rather than an end in itself. This means that if your research project utilises description it is likely to be a precursor to explanation, a **descriptive-explanatory** study.

Explanatory studies

An **explanatory study** establishes causal relationships between variables, the overarching research question being likely to begin with, or include, 'Why' or 'How' (Section 2.4). Questions that you ask during data collection to gain an explanatory response will also be likely to start with, or include, 'Why' or 'How' (Chapters 10 and 11).

The emphasis in explanatory research is to study a situation or a problem in order to understand it or explain relationships between variables. You may find, for example, that a cursory analysis of quantitative data on manufacturing scrap rates shows a relationship between scrap rates and the age of the machine being operated. You could analyse these data quantitatively in order to get a clearer view of the statistical significance of the relationship. Alternatively, you might collect further qualitative data by asking machine operators why some scrap rates are higher than others?

Evaluative studies

An **evaluative study** finds out how well something works. Investigative research questions that seek to evaluate answers are likely to begin with 'How', or include 'What', in the form of 'To what extent' (Section 2.4). Evaluative studies in business and management are likely to be concerned with assessing the effectiveness of an organisational or business strategy, policy, programme, initiative or process; for example, evaluating a marketing campaign, a personnel policy, a costing strategy, or the delivery of a support service. An evaluative study may also make comparisons between events, situations, groups, places or periods. Questions that you ask during data collection will be likely to start with, or include, 'What', 'How' or 'Why' (Chapters 10 and 11). It can produce a theoretical contribution where emphasis is placed on understanding not only 'how effective' something is, but also 'why', comparing this explanation to existing theory.

Combined studies

A research study may combine more than one purpose in its design. This may be achieved by the use of multi or mixed methods in the research design (Section 5.4). Alternatively, a single method research design may be used in a way that provides scope to facilitate more than one purpose.

5.4 Methodological choice: choosing a quantitative, qualitative or mixed methods research design

Research designs are frequently referred to as 'quantitative', 'qualitative' or 'mixed' methods. These labels highlight whether the data collected are numeric (numbers), non-numeric



Box 5.2 Focus on management research

Methodological choices in international business research

A review of 50 years of methodological trends in international business research by Nielson et al. (2020) emphasises the importance of variety in methodological choices.

Analysing articles in the *Journal of International Business* from 1970 to 2019, they note three prevailing patterns:

- increased use of large-scale, longitudinal cross-national quantitative archival data;

- increased use of complex statistical analysis techniques with the use of multiple analytical techniques in one study;
- decline in the diversity of methods although, within this, a greater variety of qualitative methods being used.

Taking a pluralist approach, they consider that these changes jeopardise the quality of international business research. In particular, they argue the study of any phenomenon needs a range of research designs; if the systematic biases, errors and limitations of any single option are to be avoided. They propose that international business as a field needs to encourage the use of a greater variety of alternative research designs to investigating a phenomenon.

(words, images, audio recordings, video clips and other similar material) or a combination of both. Quantitative refers to any data collection procedures (such as a questionnaire) or analysis techniques (such as graphs or statistics) that generate or use numerical data. In contrast, ‘qualitative’ refers to any data collection procedures (such as an interview or unstructured observation) or analysis techniques (such as narrative analysis or grounded theory) that generate or use non-numerical data. Where more than one quantitative data collection procedure and corresponding analysis technique is used, it is termed a **multi-method quantitative** design. Where more than one qualitative data collection procedure and corresponding analysis technique is used, it is a **multi-method qualitative** design. Where quantitative and qualitative data collection techniques and analysis procedures are used, it is a **mixed methods** design (Figure 5.2). Within a number of fields in business and management there has been a decline in the diversity of methodological choices (Box 5.2).

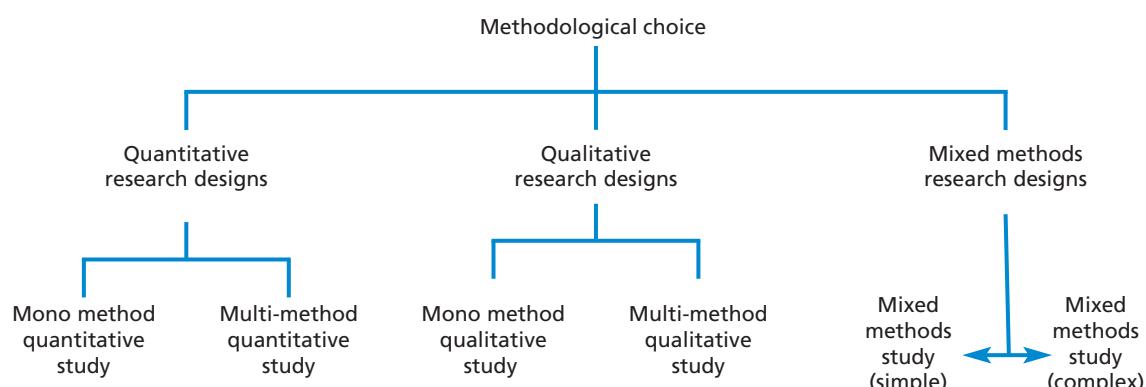


Figure 5.2 Methodological choice



Box 5.3 Focus on research in the news

Mandating vaccinations could backfire, study says

By Leke Oso Alabi

Health and social care workers who felt under pressure from their employers to receive Covid-19 vaccines were more likely to decline them, according to new research led by the London School of Hygiene & Tropical Medicine.

Nearly 2,000 people were asked whether they agreed with the statement ‘I feel/felt under pressure from my employer to get a Covid-19 vaccine’.

Those who agreed most with the statement were more likely to have declined a Covid-19 vaccine.

‘Our work shows a move towards mandating Covid-19 vaccination is likely to harden stances and negatively affect trust in the vaccination, provider, and policymakers,’ said Sandra Mounier-Jack, associate professor in health policy at LSHTM.

The LSHTM said its study also identified structural barriers to vaccination uptake. Black African and mixed black African workers were not offered vaccination at the same rates as white British and white Irish participants.

The authors of the study – which has not yet been peer reviewed – acknowledged limitations, including survey recruitment from social media and professional bodies, which may have led to some ethnic or professional groups being over or under-represented.



Source: Alabi, L.O. (2021) ‘Mandating vaccinations could backfire, study say’, *Financial Times*, 12 May. Copyright © 2021 The Financial Times Limited

As methodological pluralists we consider, dependent upon the research question, all three research designs can offer valuable insights.

We now examine these three research designs considering how each is associated with research philosophy (Sections 4.2 to 4.4) and approaches to theory development (Section 4.5), as well as looking at their characteristics and associated research strategies.

Quantitative research designs

Philosophical assumptions

Quantitative research designs are associated generally with positivism, especially when used with highly structured data collection techniques where a large number of people are asked the same questions (Box 5.3). However, it is wrong to suggest the link between positivism, deduction and a quantitative research design is exclusive (Walsh et al. 2015a). Quantitative research designs are also undertaken within the realist and pragmatist philosophies (Section 4.4).

Approach to theory development

Quantitative research is usually associated with a deductive approach, where data are collected and analysed to test theory. However, it may also incorporate an inductive approach, using data to develop theory (Section 4.5). For example, a researcher may analyse quantitative data to generate hypotheses to test through subsequent research. Alternatively, data analysis may reveal the original hypothesis was poorly framed or be used to suggest possible hypotheses to test. Walsh et al. (2015b: 621) refer to using analysis to suggest hypotheses as “**Harking**” – hypothesising after the results are known.’ While you may find it necessary to refine your original hypothesis, we recommend you do not use harking without discussing it with your project tutor.

Characteristics

Quantitative research examines relationships between variables, which are collected in a standard manner, measured numerically and analysed using a range of statistical and graphical techniques. It often incorporates controls to ensure the validity of data, as in an experimental design. Quantitative designs often use probability sampling techniques to ensure statistical generalisability (Section 7.2), the researcher being considered independent from those being researched. The characteristics of quantitative research designs are summarised in Table 5.1.

A quantitative research design may use a single data collection technique, such as a questionnaire, and corresponding quantitative analytical procedure. This is known as a **mono method quantitative study** (Figure 5.2). Where more than one quantitative data collection technique and corresponding analytical procedure are used, this is termed

Table 5.1 Characteristics principally associated with quantitative and qualitative research designs

| Characteristic | Quantitative research | Qualitative research |
|--|--|--|
| Independence of researcher from those being researched | independent | not independent |
| Terms for those taking part | respondents | participants, informants |
| Focus of study | variables and relationships between them | attributed meanings and associated variables |
| Sampling techniques | probability | non-probability |
| Generalisation | statistical | to theory |
| Data collection method(s) | rigorously defined, highly structured | unstructured or semi-structured |
| Data | numerical and standardised | non-standardised, generally requiring classification |
| Analysis | through statistics and diagrams | through conceptualisation |
| Derivation of meaning | numbers | words (spoken/text) and images |

a **multi-method quantitative study** (Figure 5.2). You might, for example, decide to collect quantitative data using both questionnaires and structured observation, analysing these data statistically. Using more than one method is likely to overcome weaknesses associated with a mono method, allowing for richer data collection, analysis and interpretation (Bryman 2006).

Research strategies

Quantitative research is principally associated with survey and experiment research strategies (Section 5.5). The survey strategy is normally conducted through the use of questionnaires, structured interviews or, possibly, structured observation. However, it is important to note that quantitative data and analysis techniques can and are used in research strategies that are often associated with qualitative, designs such as action research, case study research and grounded theory (Section 5.5).

Procedures and techniques

Procedures and techniques associated with quantitative designs are considered in Chapters 9, 11 and 12. Structured observation is discussed in Section 9.4; Chapter 11 focusses on the use of questionnaires including structured interviews; and Chapter 12 is devoted to analysing data quantitatively.

Qualitative research designs

Philosophical assumptions

Qualitative research designs are often associated with interpretivism (Denzin and Lincoln 2018), researchers making sense of subjective and socially constructed meanings. Such research is sometimes referred to as naturalistic since researchers need to operate within a natural setting, or research context, in order to establish trust, participation, access to meanings and in-depth understanding. Like quantitative research, qualitative research may also be undertaken within realist and pragmatist philosophies.

Approach to theory development

Qualitative research often uses an inductive approach to theory development, research being used to build theory or develop a richer theoretical perspective than already exists in the literature. However, some qualitative research strategies start with a deductive approach, testing an existing theory (Yin 2018). In practice, much qualitative research also uses an abductive approach to theory development, inferences being developed inductively and tested deductively in an iterative process (Section 4.5).

Characteristics

Qualitative research studies participants' meanings and the relationships between them, using a variety of data collection techniques and analytical procedures, to develop a conceptual framework and theoretical contribution. The success of the qualitative researcher's role is dependent not only on gaining physical access to those who take part, but also building rapport and demonstrating sensitivity to gain cognitive access to their data (Section 6.2). Those who consent to take part in qualitative research are seen as participants in the collection of data, the researcher not being considered independent.

In qualitative research, meanings are derived from words and images, not numbers. Since words and images may have multiple meanings as well as unclear meanings, it is often necessary to explore and clarify these with participants. Methods used are unstructured or semi-structured (Sections 9.3 and 10.3), so that research questions, procedures can alter or emerge in a naturalistic and interactive process. Qualitative designs are likely to use non-probability sampling techniques (Section 7.3). The qualitative data that are collected will be non-standardised and generally require being classified into categories for analysis. The characteristics of qualitative research are summarised in Table 5.1.

A qualitative research design may use a single data collection technique, such as semi-structured interviews, and corresponding qualitative analytical technique. This is known as a **mono method qualitative** study (Figure 5.2). Where more than one qualitative data collection procedure and corresponding analytical technique are used, this is termed a **multi-method qualitative** study (Figure 5.2). You might, for example, decide to collect qualitative data using in-depth interviews and diary accounts, analysing these data using qualitative techniques (Box 5.4).

Research strategies

Qualitative research designs are associated with a variety of strategies including Action Research, case study research, ethnography, Grounded Theory and Narrative Inquiry (Section 5.5). Some of these strategies, such as case study, are also used in quantitative research designs.

Procedures and techniques

Procedures and techniques associated with qualitative designs are considered in Chapters 9, 10 and 13. Observation is considered in Chapter 9; collecting qualitative data using semi-structured and in-depth interviews is considered in Chapter 10; while Chapter 13 focuses on techniques to analyse data qualitatively.



Box 5.4 Focus on student research

Multi-method qualitative study

Tom wanted to establish how supervisors managed teams that had adopted hybrid working. To do this he thought it essential that he should have the clearest possible grasp of supervisors' interactions with their teams. This involved him in:

- observing online meetings between five supervisors and their teams for a week to establish practice (qualitative data);

- conducting online interviews with each of the supervisors to establish their views (qualitative data);
- interviewing a sample of team members reporting to the five supervisors to establish team members' views (qualitative data).

This gave Tom a much better grasp of how supervisors managed teams' hybrid working. It also did much to enhance his credibility with the supervisors and team members, a number of whom emailed him with photographs of their home and office work environments; further qualitative data.

Mixed methods research designs

Philosophical assumptions

Mixed methods research designs integrate the use of quantitative and qualitative data collection procedures and analysis techniques in the same research project (Figure 5.2), being often associated with pragmatism and critical realism. Pragmatists choose methods because they will enable credible reliable and relevant data to be collected to address the research problem (Section 4.4). For pragmatists, the nature of the research question, the research context and likely research consequences are driving forces determining the most appropriate methodological choice (Nastasi et al. 2010). This means pragmatists do not always use mixed methods designs. Critical realism, like pragmatism, can support the use of mixed methods research and, again, critical realists do not always use mixed methods designs. They may, for example, use initially qualitative research methods to explore perceptions. This could be followed by quantitative analysis of officially published data (Section 8.2) to establish the relationship between socially constructed knowledge and possible underlying causal structures, processes and forces.

Researchers using mixed methods have a **pluralist** view of research methodology. They believe that flexibility in selection and use of methods (both quantitative and qualitative) is legitimate and that researchers should be tolerant of each other's preferred methods even when they differ from their own. These views can be contrasted with the **unitarist** view; there is, or should be, one legitimate method that should be followed.

Approach to theory development

Mixed methods research designs may use deductive, inductive or abductive approaches to theory development. For example, quantitative research may be used to test theory statistically, followed by qualitative research to develop a richer theoretical understanding. Theory may also be used to provide direction for the research. In this way a particular theory may be used to provide a focus for the research and to provide boundaries to its scope (Tashakkori and Teddlie 2010).

Characteristics

Mixed methods research draws from the characteristics of both quantitative and qualitative research (Table 5.1), combining quantitative and qualitative procedures and techniques in a variety of ways that range from simple, concurrent forms to more complex and sequential forms (Figure 5.2).

Research designs

Different combinations of quantitative and qualitative research lead to various mixed methods research designs. The principal mixed methods research designs are: concurrent triangulation design, concurrent embedded design, sequential exploratory design, sequential explanatory design (Creswell and Plano Clark 2017) and sequential, multi-phase design.

Concurrent mixed methods research uses quantitative and qualitative methods within a single phase of data collection and analysis (**single-phase research design**) (Figure 5.3). This allows both sets of results to be interpreted together to provide a richer, more comprehensive response to the research question compared to a mono method design. Where you collect qualitative and quantitative data in the same phase of research in order to

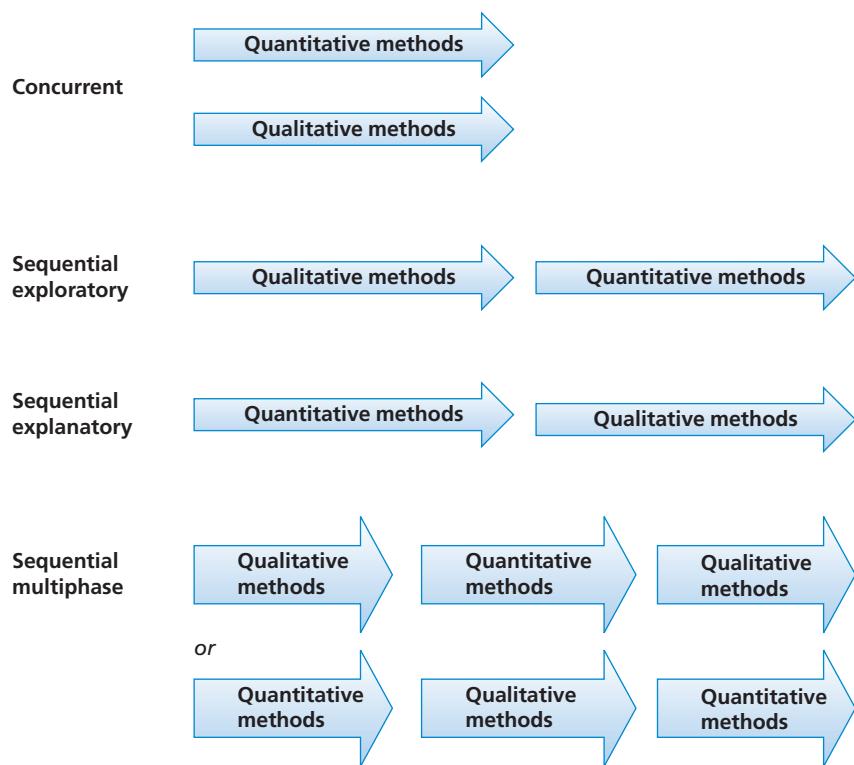


Figure 5.3 Mixed methods research designs

compare how these data sets support one another, you will be using a **concurrent triangulation design**.

Sequential mixed methods research involves more than one phase of data collection and analysis (Figure 5.3). In this, the researcher uses one method followed by another in order to expand or elaborate on the initial set of findings. This often takes more time than a concurrent design. In a **double-phase research design** two alternative mixed methods research strategies are used, either a **sequential exploratory research design** (qualitative followed by quantitative) or a **sequential explanatory research design** (quantitative followed by qualitative). A **sequential, multi-phase design** is more complex using multiple phases of data collection and analysis (e.g. qualitative followed by quantitative, then by a further phase of qualitative) (Box 5.5).

Using sequential multi-phase mixed methods research design suggests a dynamic approach to the research, where one phase subsequently informs and influences the next phase of data collection and analysis. The exact nature of the interaction and iteration influences the way in which qualitative and quantitative methods are chosen and integrated at each phase of the research (Greene 2007; Nastasi et al. 2010; Teddlie and Tashakkori 2009).

Where you mix quantitative and qualitative methods at every stage of your research (design, data collection and analysis, interpretation and presentation of the research), you will be using a **fully integrated mixed methods research** design. Where you use quantitative and qualitative methods at only one stage or particular stages of your research, you will be using a **partially integrated mixed methods research** approach (Nastasi et al. 2010; Teddlie and Tashakkori 2009, 2011).



Box 5.5 **Focus on student research**

Mixed methods research

Andreas conducted research into organisational change in an IT company, using a sequential mixed methods research design consisting of four stages:

- 1** *Initial exploratory telephone discussions* with key senior managers to negotiate access, agree the scope of the project and gain essential contextual data. These data were analysed qualitatively in order to get a picture of important internal and external organisational issues.
- 2** *Individual in-depth online interviews* with a sample of 28 directly employed staff (excluding contractor staff), representing the organisation across its departments and throughout its grade

structure. These data were analysed qualitatively to establish the issues that were important to staff, to help to inform the content of the questionnaire.

- 3** *An online questionnaire* designed, pilot-tested, amended and delivered to a representative sample of directly employed staff, producing a 42 per cent response rate. The quantitative data collected were analysed statistically to allow the views of employee groups to be compared for differences by age, gender, length of service, occupation and grade.
- 4** A fourth stage consisted of real-time *Internet presentations of initial findings to employees*. These allowed employees' questions to be answered and discussion to occur to clarify the content of some of the questionnaire results. Notes from these presentations were analysed qualitatively.

Mixed methods research may use quantitative research and qualitative research equally or unequally (Creswell and Plano Clark 2017), one methodology having a dominant role and the other a supporting role. This prioritisation reflects the research purpose, researcher preferences and the expectations of those who commission the research (such as your project tutor or the managers in an organisation).

The purpose of the research will emphasise the initial use and prioritisation of qualitative research (as in an exploratory study, where qualitative precedes quantitative) or quantitative research (as in a descriptive study, before using qualitative research to explain particular findings). The purpose will also emphasise the dominance of either quantitative or qualitative research such as in a sequential project which commences with a qualitative, exploratory phase, followed by a quantitative, descriptive phase, and is completed by a further qualitative, explanatory phase. For other research projects the purpose will suggest more equal use of quantitative and qualitative research methods. Similarly, the approach to theory development will lead to prioritising either quantitative or qualitative methods; an inductive approach designed to generate theoretical concepts and to build theory may emphasise the use of qualitative methods.

Merging quantitative and qualitative methods may involve either using '**quantitised**' qualitative data (e.g. specific events in the data are counted as frequencies and numerically coded for statistical analysis) or '**qualitised**' quantitative data (e.g. frequencies are turned into text, although this is extremely rare in practice) or both. Qualitative data may be presented diagrammatically (Box 12.9) and quantitative data presented as text. This approach to mixing methods may be risky as the value of each form of data may be diluted; for example, excessively 'quantitising' qualitative data may lead to loss of its exploratory or explanatory richness.

Embedded mixed methods research refers to one methodology supporting the other (Creswell and Plano Clark 2017). One methodology may be embedded within the other during a single means to collect data (e.g. some quantitative questions are included in

an interview schedule, or some questions within a questionnaire require a qualitative response) – a **concurrent embedded design**. Alternatively, a single-phase research design may use both quantitative and qualitative methods concurrently, collecting these separately, the analysis of one informing the other. Within a multi-phase, sequential research design, both quantitative and qualitative methods will be collected and analysed, one after the other, with one being used in a supporting role.

The characteristics that help to define mixed methods research designs highlight how quantitative and qualitative methods may be combined in a number of ways to provide you with better opportunities to answer your research question (Table 5.2).

Procedures and techniques

Quantitative data collection procedures and analytical techniques that may be used as part of mixed methods research are considered in Chapters 9, 11 and 12. Structured observation

Table 5.2 Reasons for combining methods in a mixed methods design

| Reason | Explanation |
|------------------|--|
| Initiation | The initial method defines the nature and scope of subsequent research. Can also provide contextual background and to better understand the research problem (e.g. Box 5.3). It may inform redrafting of research questions, sample selection and data collection procedures |
| Facilitation | One method generates new insights that inform and are followed up using another method |
| Complementarity | Meanings and findings are elaborated, enhanced, clarified, confirmed, illustrated or linked |
| Interpretation | One method (e.g. qualitative) is used to help to explain relationships emerging from the other (e.g. quantitative) |
| Generalisability | Helps establish the generalisability of findings (e.g. qualitative followed by quantitative) or the credibility of a study or produce more complete knowledge (e.g. quantitative followed by qualitative) |
| Diversity | Allows for a greater diversity of views to inform and be reflected in the study |
| Problem-solving | An alternative method helps when the initial method generates insufficient understanding |
| Focus | One method focuses on one attribute (e.g. quantitative on macro aspects), while the other method focuses on another attribute (e.g. qualitative on micro aspects) |
| Triangulation | Ascertains if the findings from one method mutually corroborate the findings from the other method |
| Confidence | Findings may be affected by the method used. Mixed methods ascertains and cancels out this 'method effect' leading to greater confidence in your conclusions |

Source: Developed from Bryman (2006), Greene (2007), Molina-Azorin (2011) and authors' experience

is discussed in Section 9.4; Chapter 11 discusses the use of questionnaires, including structured interviewing; and Chapter 12 considers the analysis of quantitative data.

5.5 Developing a coherent research strategy

In this section we turn our attention to your choice of research strategy (Figure 5.1). Your **research strategy** is the methodological link between your philosophy and subsequent choice of methods to collect and analyse data (Denzin and Lincoln 2018).

Within business and management there are a variety of research strategies with a range of methodological choices resulting in alternative combinations of quantitative, qualitative or mixed methods research designs (Table 5.3). Particular research strategies are associated with particular research philosophies and approaches to theory development; however, boundaries between research philosophies, approaches to theory development and research strategies are, at least to some extent, open.

As pluralists, we believe a particular research strategy should not be seen as inherently superior or inferior to any other. Rather, the key to the choice of a research strategy or strategies is achieving a reasonable level of coherence throughout your research design and ensuring the research question(s) are answered and the research objectives met. The coherence between research question(s) and objectives, and your philosophy and approach to theory development is crucial; alongside more pragmatic concerns including the extent of existing knowledge, the amount of time and other resources you have available and being able to obtain data. The strategy or strategies you adopt should not be thought of as being mutually exclusive or exclusive to one philosophy (Section 5.4). For example, it is possible to use the survey strategy within a case study or combine a number of different strategies within a mixed methods design.

In our experience the choice between strategies associated principally with qualitative research decisions that is likely to cause the greatest confusion. Such confusion is often unsurprising given the diversity of qualitative strategies (many more than those we consider), with their conflicting tensions and ‘blurred genres’ (Denzin and Lincoln 2018: 10).

Table 5.3 Research strategies and methodological choice

| Strategy | Principal associated research designs |
|-------------------|---|
| Experiment | Quantitative mono- and multiple methods |
| Survey | Quantitative mono- and multiple methods |
| Ethnography | Qualitative mono- and multiple methods |
| Grounded Theory | Qualitative mono- and multiple methods |
| Narrative Inquiry | Qualitative mono- and multiple methods |
| Archival | Quantitative mono- and multiple methods, qualitative mono- and multiple methods, mixed methods |
| Case study | Qualitative multiple methods, quantitative multiple methods, mixed methods |
| Action Research | Quantitative mono- and multiple methods, quantitative mono- and multiple methods, mixed methods |

We now draw out the distinctions between a range of strategies and their associations with quantitative, qualitative, quantitative and mixed methods research designs.

Experiment

An **experiment** studies the probability of a change in an independent variable causing a change in another, dependent variable. This strategy that owes much to the natural sciences, although it features strongly in psychological and social science research, and, with its roots in natural science, laboratory-based research, is often seen as the ‘gold standard’ against which the rigour of other strategies is assessed.

Hypotheses

In an experiment you hypothesise whether or not a relationship will exist between the variables, formulating two opposing hypotheses that could explain the relationship and testing these statistically (Section 12.6). In a standard experiment two types of (opposing) hypotheses are formulated and tested: the null hypothesis and the hypothesis. The **null hypothesis** is the explanation that there is no relationship or difference between the variables, for example:

User satisfaction of online customer support is not related to the amount of training support staff have received.

The **hypothesis** (also referred to as the **alternative hypothesis**) is the explanation that there is a relationship or difference between the variables, for example:

User satisfaction of online customer support is related to the amount of training support staff have received.

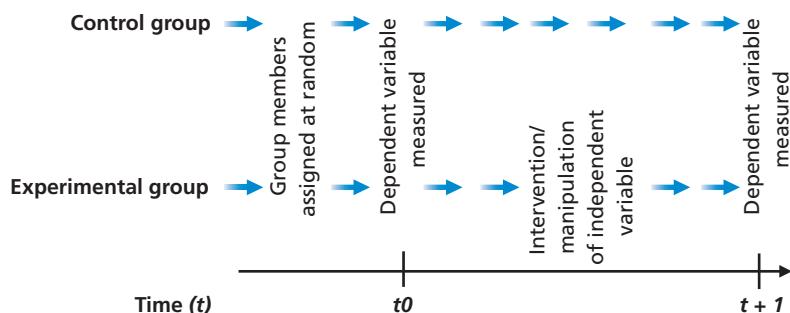
In an experiment, testing statistically the compatibility of the data with the null hypothesis is based on the probability of these data or data more extreme occurring by chance (Wassenstein and Lazar, 2016). In effect, this measures the probability that the data are compatible with the null hypothesis. The smaller the probability (termed the *p*-value), the greater the statistical incompatibility of the data with the null hypothesis. This ‘incompatibility’ casts doubt or provides evidence against the null hypothesis and its associated underlying assumptions. Where this probability is greater than a prescribed value (usually $p = 0.05$), the null hypothesis is usually accepted, and the hypothesis is rejected. Where the probability is less than or equal to the prescribed value (usually $p = 0.05$), this indicates that the hypothesis can be accepted. The simplest experiments are concerned with whether there is a relationship or difference between two variables, a dependent variable such as user satisfaction of online customer support and an independent variable such as the amount of training support staff have received (Table 5.4). More complex experiments also consider the change in the dependent variable and the relative importance of two or more independent variables as well as in some designs the impact of other types of variable such as mediating and moderator variables (Table 5.4).

Experimental designs

Experimental designs include classical experiments, quasi-experiments and within-subject designs. In a **classical experiment**, a sample of participants is selected and assigned randomly to an experimental or control group (Figure 5.4). The **experimental group**

Table 5.4 Types of variable

| Variable | Meaning |
|------------------|---|
| Independent (IV) | Variable manipulated or changed to measure its impact on a dependent variable |
| Dependent (DV) | Variable that may change in response to changes or manipulation in other independent variables |
| Mediating (MV) | Variable that transmits the effect between the independent and dependent variables |
| Moderator | Variable that affects the nature of the relationship between the independent variable and dependent variable |
| Control | Additional observable and measurable variables that need to be kept constant to avoid them influencing the effect of the independent variable on the dependent variable |
| Confounding | Extraneous but difficult to observe or measure variables that can potentially undermine the inferences drawn between the independent and dependent variables |

**Figure 5.4** Classical experimental design

receives the planned intervention or manipulation. The **control group** does not receive the intervention. Random assignment means each group should be similar in all aspects relevant to the research other than their exposure to the planned intervention. In assigning the members to the control and experimental groups at random and using a control group, you try to control (that is, remove) the possible effects of an alternative explanation to the planned intervention and eliminate threats to internal validity. Because the control group is subject to exactly the same external influences as the experimental group other than the planned intervention, this intervention is the only explanation for any changes to the dependent variable.

A **quasi-experiment** still uses an experimental group(s) and a control group, but the researcher does not assign participants randomly to each group, perhaps because participants are only available in pre-formed groups (e.g. existing work groups). Differences in participants between groups may be minimised by the use of matched pairs. **Matched pair analysis** leads to a participant in an experimental group being paired with a participant in the control group based on matching factors relevant to the experiment such as age,

gender, occupation, length of service and grade. This minimises the effect of extraneous variables on the experiment's outcomes.

The basic experimental procedure in classical and quasi-experiments is the same (Figure 5.4), with the exception of random assignment. Let us look at an example related to the introduction of a sales promotion. The dependent variable is purchasing behaviour and is measured for members of both the experimental and control groups before any intervention. This provides a **pre-test** measure of purchasing behaviour. A planned intervention of a 'buy two, get one free' promotion is then made to members of the experimental group. In the control group, no such intervention is made. The dependent variable, purchasing behaviour, is measured after the manipulation of the independent variable (the use of the 'buy two, get one free' promotion) for both the experimental and control groups, so that a pre-test and post-test comparison can be made. Any difference between the experimental and control groups for the dependent variable (purchasing behaviour) can then be attributed to the intervention of the 'buy two, get one free' promotion. This is termed a **between-subjects design**, where participants belong to either the experimental group or control group but not both. In a between-subjects design, if more than one intervention or manipulation is to be tested, a separate experimental group will be required for each test (known as **independent measures**). If the experiment was designed to compare two separate interventions, such as 'buy one, get one free' and 'buy two, get one free' promotions, two experimental groups would be used alongside the control group.

In a **within-subjects design**, or within-group design, there is only a single group, and every participant is exposed to the planned intervention or series of interventions. For this reason, this approach is known as **repeated measures**. The procedure involves a pre-intervention observation or measurement, to establish a baseline (or control for the dependent variable). This is followed by a planned intervention (manipulation of the independent variable) and subsequent observation and measurement (related to the dependent variable). Within-subject designs may be more practical than a between-subjects designs, requiring fewer participants. However, the design may lead to carryover effects where familiarity or fatigue with the process distorts the validity of the findings. This may lead to a counterbalanced design, where some of the participants undertake tasks in a different order to see if familiarity or fatigue affects the outcomes.

Experiments in business and management, or related disciplines such as organisational psychology, are sometimes conducted in laboratories rather than in the field (e.g. in an organisation). This offers greater control over aspects of the research process such as sample selection and the context within which the experiment occurs. This improves the **internal validity** of the experiment, that is, the extent to which the findings can be attributed to the interventions rather than any flaws in your research design. However, **external validity**, the extent to which the findings from the study can be generalised to all relevant contexts is likely to be more difficult to establish (Section 5.8). Laboratory settings are less unlikely to be related to the real world of organisations.

The suitability of an experimental strategy will depend on your research question. Most business and management research questions ask about the relationships between variables, rather than test a predicted relationship. This emphasises a difference between experiments and other research strategies and, within quantitative research designs, highlights a key difference between an experimental strategy and a survey strategy.

Survey

The **survey** strategy is usually associated with a deductive research approach and is most frequently used to answer 'what', 'who', 'where', 'how much' and 'how many' questions.

It therefore tends to be used for descriptive, exploratory and explanatory research, suggesting possible reasons for relationships between variables and producing models of these relationships. Survey strategies using questionnaires are popular as they enable the collection of standardised data from a large number of respondents economically, allowing easy comparison. Using a survey strategy should give you more control over the research process and, when probability sampling is used, it is possible to generate findings that are statistically representative of the target population at a lower cost than collecting the data from them all (Section 7.2). In addition, the survey strategy is perceived as authoritative by people and is comparatively easy both to explain and to understand. Every day a news bulletin, news website or newspaper reports the results of a new survey that is designed to find out how a group of people think or want in relation to a particular issue (Box 5.6).



Box 5.6 Focus on research in the news

Executive education 2021: FT survey shows what employers want

By Andrew Jack

Coronavirus disrupted face-to-face contact and forced companies to trim their costs, but it has also reinforced many employers' commitment to training for a broader range of their middle and senior managers.

From a self-selective poll organised by the FT, more than a quarter of chief learning officers (CLOs) around the world said they intended to increase their budgets for executive education in 2021, while over half said they would maintain spending at 2020 levels. Just 17 per cent planned a reduction.

The findings come from a pioneering survey conducted by the FT in partnership with Unicon, the international consortium for university-based executive education, along with the Association to Advance Collegiate Schools of Business and the European Foundation for Management Development – the two leading accreditation agencies – as well as the Society for Human Resource Management.

Of the 363 respondents surveyed in February and early March 2021, the majority worked for companies based in the US and Canada, but Europe, Latin America, the Middle East and Africa were also well represented. Respondents worked in organisations of various sizes: more than two-fifths oversaw workforces of fewer than 1,000 people and more than a fifth were responsible for training in groups with more than 20,000 staff. Respondents from finance, banking, healthcare and industrial businesses dominated.

Leadership was the top learning priority identified for executive education, cited as important by 82 per cent of respondents. Change management followed, at 57 per cent. Other longstanding priorities, including digital transformation, strategy and innovation, were also ranked highly by more than two-fifths of respondents. In a sign of the issues brought to the fore during the pandemic, the need for training around resilience, well-being and remote or online collaboration were also cited by many.

Another emerging theme – diversity and inclusion – was a high priority, cited by 55 per cent, placing it third overall. Among US respondents, it was still higher – in second



place after leadership. That reflects the growing focus on a subject that has mobilised senior managements in recent months, triggered partly by the killing of George Floyd last May and the Black Lives Matter movement.

Business schools can take some comfort from the fact that just over half of the CLOs surveyed said they would turn to universities during 2021 for their executive learning programmes. However, a larger proportion pointed to alternatives: more than two-thirds said they would use internal resources and nearly three-quarters planned to use non-university training partners.



Source of extract: Jack, A. (2021) 'Executive education 2021: FT survey shows what employers want', Financial Times, 9 May. Copyright © 2021 The Financial Times Limited

The survey strategy allows you to collect data that you can analyse quantitatively (Chapter 12) using a variety of data collection methods including questionnaires and structured interviews (Chapter 11) and structured observation (Section 9.4). You will need to ensure your sample is representative, design and pilot your data collection instrument, and try to ensure a good response rate. Invariably, there is a limit to the number of questions that any questionnaire or structured interview can ask; and preparing and analysing the data will also be time-consuming. Despite this, perhaps the biggest drawback with using a questionnaire or structured interview as part of a survey strategy is the capacity to do it badly!

Ethnography

Ethnography is a strategy that focuses on describing and interpreting the social or cultural world of a group through first-hand study. It means, literally, a written account of a people or ethnic group. Ethnographers study people in groups, who interact with one another and share the same space, whether this is at street level, within a work group, in an organisation or within a society. It is the earliest qualitative research strategy, with its origins in colonial anthropology.

We present our subsequent discussion of ethnography as a developmental account. Ethnography is a demanding strategy to use, not least because of the time scale and intensity involved. However, if you are working in an organisation, there may be scope to undertake participant observation of your workgroup or another group in the organisation (Chapter 9). Alternatively, where you have recently undertaken a work placement, you will be familiar with the context and complexity of this workplace and you may be able to negotiate access based on your credibility to undertake an ethnographic study related to a work group. Ethnography is relevant for modern organisations. Alternatively, for example, you may wish to gain an in-depth understanding of their markets and the experiences of their consumers.

Being successful with this strategy is likely to include making sure that the scale or scope of your proposed ethnographic research project is achievable. This will relate to your research question and objectives. To collect sufficient data, you will need to make detailed notes of everything you observe and spend considerable time reflecting on what you have observed. You will also need to make additional notes to elaborate on these and supplement the process of observation, by conducting informal discussions and interviews

to explore what you have observed and collect any documentation that supports your data collection (Delamont 2007).

From the 1700s to the early 1900s, ethnography was developed to study cultures in societies that had been brought under the rule of a colonial power, to facilitate imperialist control and administration. Early anthropologists treated those among whom they lived and conducted their fieldwork as subjects and approached their ethnography in a detached way, believing that they were using a scientific approach, reminiscent of a positivism, producing what were meant to be accurate and timeless accounts of different cultures (Tedlock 2005). However, as we highlight in the opening vignette of Chapter 4, such studies (and their associated interpretations) privileged the colonialist ethnographers' taken for granted beliefs and assumptions.

From the 1920s the use of ethnography changed through the work of the Chicago School (University of Chicago), which used ethnographic methods to study social and urban problems within cultural groups in the USA. A seminal example of this work is Whyte's (1993) 'Street Corner Society' originally published in 1943, which examined the lives of street gangs in Boston. This approach to ethnography involved researchers living among those whom they studied, observing and talking to them to produce detailed cultural accounts of their shared beliefs, behaviours, interactions, language, rituals and the events that shaped their lives (Cunliffe 2010). This use of ethnography adopted a more interpretive and naturalistic focus using the language of those being studied in writing up cultural accounts. However, the researcher remained the arbiter of how to tell the story and what to include, leading many to question the impact of the researcher's socialisation and values (Geertz 1988).

This problem of 'representation' (Denzin and Lincoln 2018) means ethnography, as well as qualitative research more generally, remains in a fluid developmental state. In the second half of the twentieth century, researchers developed a 'bewildering array' (Cunliffe 2010: 230) of qualitative research strategies, associated with a great deal of 'blurring' across these strategies (Denzin and Lincoln 2018). Conflict about how best to achieve focus led to a range of ethnographic strategies of which we consider four: realist ethnography, impressionist or interpretive ethnography, critical ethnography and autoethnography.

Realist ethnography

Realist ethnography is the closest to the ethnographic strategy described earlier. The realist ethnographer believes in objectivity, factual reporting and identifying 'true' meanings. She or he reports the situation observed using 'facts' or data about structures and processes, practices and customs, routines and norms, artefacts and symbols. Such reporting is likely to use standardised categories that produce quantitative data from observations. The realist ethnographer's account is written in the third person, portraying their role as the impersonal reporter of 'facts'. It presents a detailed contextual background, the nature of the cultural interactions observed, and the patterns of behaviour and social processes identified. Quotations are used dispassionately without personal bias or seeking to act as an agent for change. The realist ethnographer's written account is his or her representation of what he or she has observed and heard.

Interpretive ethnography

In contrast, **interpretive ethnography** places greater emphasis on subjective impressions than on perceived objectivity. The interpretive ethnographer believes in the likelihood of multiple meanings rather than being able to identify a single, true meaning. Multiple meanings are located in different participants' socially constructed interpretations. This suggests a more pluralistic approach focused on understanding meanings, with those being

observed treated as participants rather than subjects. It requires the researcher to engage in continuous reflexivity to try to ensure quality in this research process (Delamont 2007) (Section 5.8). The research report will reflect the participation of both the ethnographer (writing in the first person), editing themselves into the text (rather than out of it) and those being observed, through devices such as personalisation, use of dialogue and quotations, dramatisation and presentation of different perspectives. It will also involve contextualisation, orderly and progressive description, factual reporting, analysis and evaluation.

Critical ethnography

Critical ethnography has a radical purpose, designed to explore and explain the impact of power, privilege and authority on those who are subject to these influences or marginalised by them (Section 5.4). Critical ethnographers often adopt an advocacy role in their work to try to bring about change. You may be able to adopt a constrained or bounded version of critical ethnography to explore the impact of a problem or issue within an organisation or work group, with a view to advocating change. Such an issue could be concerned with strategy development, decision-making procedures, regulation, governance, organisational treatment, reward and promotion, communication and involvement and so forth.

Autoethnography

Autoethnography describes and systematically analyses personal experience in order to understand cultural experience. It therefore combines the characteristics of autobiography and ethnography in which you write analytically about past experiences usually using hindsight. In writing an autoethnography, you may interview others as well as consult other sources such as texts and photographs and diaries. Crucially, you need to write your autoethnography using research methods and the academic literature to analyse and contextualise the insights provided by your cultural experience (Box 5.7).

Grounded Theory

'Grounded theory' can refer to a strategy, a method, the theory developed through the strategy, and a research process (Bryant and Charmaz 2007; Charmaz 2011; Strauss



Box 5.7 **Focus on management research**

Mothers and researchers in the making

Huopalainen and Satama (2019) undertook an autoethnographic study to answer the question 'How do early-career academic mothers balance the demands of contemporary motherhood in academia?' Using their autoethnographic diary notes gathered during, before and after the birth of their babies as data,

they provide a detailed understanding of how they negotiated becoming new mothers in their respective universities. Inspired by feminist thinking and matricentric feminism, they review the academic literature on maternal embodiment and the 'new' academia. This is used to help frame the analysis of their own diary notes. In their analysis they use the method of memory work, collectively analysing the individual memories of their day-to-day experiences and making explicit connections between their own experiences and existing research. Their article reveals their ongoing negotiation of striving to become both good academics and good mothers is a process loaded with gendered norms, expectations and beliefs.

Table 5.5 Common key elements of Grounded Theory strategy

- use of an abductive approach that seeks to gain insights to create new conceptual possibilities that are then examined;
- early commencement of data collection;
- concurrent collection and analysis of data;
- developing codes and categories from the data as these are collected and analysed;
- use of constant comparison and writing of self-memos to develop conceptualisation and build a theory;
- use of theoretical sampling and theoretical saturation to develop theory;
- initial use of literature as a complementary source to the categories and concepts emerging in the data, rather than as the source to categorise these data;
- later use of literature to review the place of the grounded theory in relation to existing theories;
- development of theory that is grounded in the data.

and Corbin 2008; Walsh et al 2015a). A **Grounded Theory** strategy, often referred to as ‘Grounded Theory (Methodology)’ offers distinctive, sequential guidelines for using qualitative methods inductively to develop theory from data. In contrast, ‘**grounded theory (method)**’ is the data collection procedures and analysis techniques used to derive meaning from the subjects and settings being studied (Section 13.9). The outcome of this method is a **grounded theory**, that is a theory that is grounded in or developed inductively from a set of data. In this sub section we consider the development of Grounded Theory, its key elements, and the implications of adopting it as a strategy. In this section we use capital letters (**Grounded Theory**) to distinguish the research strategy from the outcome, a **grounded theory** (no capital letters). The common key elements of the strategy are summarised in Table 5.5.

Development

Grounded Theory was developed by Glaser and Strauss (1967) as a response to the ‘extreme positivism’ of much social research at that time (Suddaby 2006: 633). Believing positivism is suited to research in the natural sciences, they considered social research should use a different philosophy. By adopting an interpretive approach in social research to explore human experience, ‘reality’ is seen as being socially constructed through the meanings that social actors ascribe to their experiences and actions. Grounded Theory was therefore developed as a strategy to analyse, interpret and explain the meanings that social actors construct to make sense of their everyday experiences in specific situations (Glaser and Strauss 1967; Suddaby 2006; Charmaz 2014).

Grounded Theory is used to build theoretical explanations of social interactions and processes in a wide range of contexts. As many aspects of business and management are about people’s behaviours, for example consumers’ or employees’, a Grounded Theory strategy can be used to explore a wide range of business and management issues. As the title of Glaser and Strauss’s (1967) book *The Discovery of Grounded Theory* indicates, the aim is to ‘discover’ or generate theory grounded in the data produced from the accounts of social actors.

Not only did Glaser and Strauss (1967) challenge traditional philosophical assumptions about conducting social research at that time, they also developed a set of principles and guidelines to conduct Grounded Theory. These provide a systematic and emergent approach to collect and analyse qualitative data.

Grounded Theory is usually considered as using an inductive approach, although, as we discuss later, it may be more appropriate to think of it as abductive, moving between

induction and deduction (Charmaz 2011; Strauss and Corbin 1998; Suddaby 2006). Data collection starts by collecting data such as from an initial interview or observation and then analysing these as soon as possible and before collecting more data. This is known as collecting and analysing data simultaneously. Analysis commences by identifying analytical codes that emerge from the initial interview or observation data. Each code is used to label pieces of data (such as a line, sentence or paragraph in an interview transcript) with the same or similar meaning. Coding also allows related fragments of data from different interviews or observations to be linked together to facilitate the on-going process of analysis (Section 13.9).

Grounded Theory is a useful and widely recognised research strategy and yet it has been the subject of much evaluation, criticism and even misunderstanding (Box 5.8). This is partly due to the development of different approaches to grounded theory method. Glaser and Strauss, who developed Grounded Theory, each went on to develop different approaches to its use. Strauss has become associated with the development of a



Box 5.8 Focus on management research

What is Grounded Theory?

A symposium held to debate the question 'What is Grounded Theory?' is reported in a dedicated section of an issue of *Organizational Research Methods*. Following an Introduction, this section is composed of five related articles that seek to address and debate this question. The first article contains the edited comments of the six panel members who contributed to this symposium (Walsh et al. 2015a). These contributors include Barney Glaser, one of the originators of Grounded Theory. Three further articles form commentaries on the symposium (Corley 2015; Dougherty 2015; Locke 2015). The final article is a rejoinder by the panel members to the three commentaries (Walsh et al. 2015b). Walsh's introductory comments in Walsh et al. (2015a) provide the rationale for this symposium, 'In 2006, Suddaby wrote a very interesting piece detailing what Grounded Theory "is not" . . . It has now become even more essential and urgent to understand the full reach and scope of Grounded Theory and to clarify what GT "is" as different applications of GT have led to a rather blurred picture of it.'

Walsh says that approaches to Grounded Theory vary from the way in which it was originally conceived, some using all of its methodological elements and others only using particular elements such as a coding

procedure. Variations to Grounded Theory are referred to as remodelling. The original, orthodox version is referred to as 'Classic' Grounded Theory, defined in Holton's comments in Walsh et al. (2015a) as the grounded theory methodology outlined in Glaser and Strauss (1967) and then developed in the subsequent work of Glaser (e.g. 1978, 1992). The scope of this approach is seen to be:

- philosophically flexible: it can be used by either positivist or interpretive researchers;
- a general methodology: it can be used with qualitative or quantitative data, or both, providing that theoretical sampling occurs in its collection;
- one that emphasises the study of a phenomenon in its context over the use of prior existing theory;
- a theory-building method that implies use of an exploratory and inductive data-driven process that may incorporate deduction to build theory.

The debate between the six panel members of this symposium and the authors of the three articles who offered their comments provides further insight into the question 'What is Grounded Theory?'. In seeking to address this question, Locke (2015: 615) points readers interested in developing 'a fuller picture of the grounded theory arena [to] consult *Developing Grounded Theory: The Second Generation* (Morse et al 2009) compiled by six grounded theory practitioners . . . who apprenticed with Glaser and Strauss . . . and embody the distinctions and tussles that have evolved in the domain'.

particularly prescriptive approach to grounded theory method (e.g. Corbin and Strauss 2008; Strauss and Corbin 1998).

Key elements

Grounded Theory means that the process of data collection and analysis becomes increasingly focused, leading to the generation of a contextually based theoretical explanation (Bryant and Charmaz 2007). It comprises a number of key elements, namely coding, constant comparison, memo writing, theoretical sampling (including theoretical saturation), and theoretical sensitivity.

Coding: In the Grounded Theory strategy of Strauss and Corbin (1998) there are three coding stages: the reorganisation of data into categories is called **open coding**, the process of recognising relationships between categories is referred to as **axial coding** and the integration of categories to produce a theory is labelled **selective coding**. More recently, Corbin has altered the approach in Corbin and Strauss (2008), with axial coding being combined within open coding and selective coding simply becoming ‘integration’ (Section 13.9). Charmaz’s (2014) approach is more flexible, involving two principal coding stages known as initial coding and focused coding.

Constant comparison: Underpinning coding is the process of **constant comparison**. Each item of data collected is compared with others, as well as against the codes being used to categorise data. This is to check for similarities and differences, to promote consistency when coding data and to aid the process of analysis. Where appropriate, new codes are created, and existing codes reanalysed as new data are collected. Constant comparison promotes the higher levels of analytical coding we referred to earlier because it involves moving between inductive and deductive thinking. As you code data into categories, a relationship may begin to suggest itself between specific codes (here, inductive thinking links specific codes to form a general proposition). This emerging interpretation is ‘tested’ through collecting data from new cases (here, deductive thinking tests this abstract generalisation, to see if it stands up as an explanatory relationship to form a higher-level code) (Strauss and Corbin 1998). This process of gaining insights to create new conceptual possibilities that are then examined is termed abduction (Charmaz 2011; Suddaby 2006) (Section 4.5).

Memo writing: **Memo writing** aids the development of grounded theory throughout a research project as you define or make notes about:

- the codes being used;
- how codes change through the research process;
- how codes might be related, helping identify theoretical relationships and the emergence of higher-level codes and categories;
- any other ideas that occur to the researcher that help him or her to develop the research process and analyse the data.

Where you use a Grounded Theory strategy, your collection of self-memos will provide you with a chronological record of the development of your ideas and your project, and show how you arrived at your grounded theory.

Theoretical sampling: When using Grounded Theory, you will also need to decide how to select cases for your research. As you analyse data, the categories being developed will indicate the type of new cases (e.g. new participants) to select. The purpose of **theoretical sampling** is therefore based upon developing and testing the emerging theory and the evolving story line; participants being chosen purposively to inform this rather than to achieve statistical representativeness (Section 7.3). Your core theme, relationship or process around which you focus the research, and the need to test your emerging theory

also provides the focus to select new cases. Sampling continues until theoretical saturation is reached; when further data does not reveal any new properties that are relevant to a category, and where categories have become well developed and understood and relationships between categories have been verified (Strauss and Corbin 1998). This is also termed achieving conceptual density (Glaser 1992) or conceptual saturation (Corbin and Strauss 2008).

Objectivist and subjectivist grounded theory: A further difference has been revealed by Charmaz (2014), who makes a distinction between ‘objectivist grounded theory’ and ‘constructivist grounded theory’. Charmaz views the approach of Glaser, Strauss and Corbin to grounded theory as being ‘objectivist’, which assumes that data indicate an external reality, just waiting to be ‘discovered’. She considers that ‘objectivist grounded theory’ has positivist leanings. According to this view, it is only ‘constructivist’ grounded theory that is truly based on an interpretive approach, because it recognises that the researcher’s role in interpreting the data will affect the development of a grounded theory. In this approach, grounded theories are ‘constructed’, not discovered. This might seem a rather abstract difference, but because Charmaz advocates a ‘constructivist’ approach she also promotes a more flexible approach to grounded theory method (Section 13.9).

Adopting a Grounded Theory strategy

Adopting a Grounded Theory strategy invariably has implications. These concern the data collection; use of existing theory; identifying a core category or categories around which to focus the research; and the time required to undertake this strategy. We briefly consider each of these.

Data collection: In Grounded Theory, data collection can start as soon as the research idea has been developed and the initial research participants have agreed to take part (or the first set of documents have been identified). This means you need to be interested in and committed to your research idea from the start.

Use of existing theory: There is sometimes confusion about the role of published theory in a Grounded Theory research project (Suddaby 2006; Locke 2015). Grounded theorists may use published theory before and during their research project. The idea for such a research project may come from existing theory and your understanding of the theoretical background to your research topic may help to inform the project in general terms. However, existing theory should not be allowed to influence how you code your data, decide on new cases and conduct your analysis. Grounded Theory is an emergent strategy and you need to be guided by concepts emerging from the data you collect rather than being sensitised by concepts in existing theory. This is known as **theoretical sensitivity**, where you focus on interpreting meanings by using *in vivo* and researcher-generated rather than *a priori* codes (Section 13.6) to analyse your data and construct a grounded theory (Glaser 1978). Theoretical sensitivity means that you must be sensitive to meanings in your data, generating a theory grounded in them. You will, however, need to allow yourself sufficient time later on to link your grounded theory to published theories as you write your research report!

Identifying a core category: The emergent nature of the strategy means your identification of a core category or categories around which to focus your research and develop a grounded theory is crucial. This requires rigorous use of coding, constant comparison, theoretical sampling until theoretical saturation is reached, alongside theoretical sensitivity to develop a theoretical explanation.

Time requirements: Using Grounded Theory is time-consuming, intensive and reflective. Before committing yourself to this strategy, you need to consider the time that you have to conduct your research, the level of competence you will need, your access to

data, and the logistical implications of immersing yourself in an intensive approach to research. Kenealy (2012) advises novice Grounded Theory researchers to identify one approach to grounded theory method and follow it without too much adaptation. He also advises researchers to focus on identifying ‘ideas that fit and work’ from their data to develop a grounded theory (Kenealy 2012: 423). Kenealy recognises that using Grounded Theory requires experience but says that the only way to build this is to practise the use of grounded theory method!

Narrative Inquiry

A narrative is a story; a personal account which interprets an event or sequence of events (Box 5.9). Using the term ‘narrative’ requires a distinction to be drawn between its general meaning and the specific meaning here. A qualitative research interview inevitably involves a participant in storytelling, and so the term ‘narrative’ can be applied generally to describe the nature or outcome of a qualitative interview. As a research strategy, however, **Narrative Inquiry** means collecting the experiences of participants as whole accounts, or reconstructing their experiences into narratives.

The purpose of Narrative Inquiry is to derive theoretical explanations from narrative accounts while maintaining their integrity. Where your research question and objectives are consistent with an interpretivist philosophy and a qualitative methodological choice, Narrative Inquiry may be suitable. It will allow you to analyse the linkages, relationships and socially constructed explanations that occur naturally within narrative accounts in



Box 5.9 Focus on student research

Using Narrative Inquiry to explore marketing strategies

Kasia was undertaking a marketing degree and, because of her longstanding interest in fashion and textiles, she hoped to find work in that sector. Kasia’s interests led her to focus her research project on factors that affected the success of marketing strategies in a small sample of fashion companies. After considering her choice of research strategy and discussing this with her project tutor, she decided to adopt a Narrative Inquiry strategy, using online in-depth interviews with the marketing directors from a sample of three medium-sized fashion companies who had agreed to grant her access. Kasia realised that the outcome of her research would very much depend on the quality of these three in-depth interviews. She decided to send each an email briefly outlining this approach and a list of the structural elements of narrative inquiry she had read about.

She was nervous in her first interview and realised that her participant, Hetal, sensed this. Hetal had read Kasia’s email and knew a little about Narrative Inquiry from her own degree studies. Hetal provided Kasia with a full and useful narrative of the factors affecting the outcomes of her employer’s marketing strategy over the past year. However, after the interview Kasia realised that she had asked Hetal unnecessary questions on several occasions, interrupting the flow of Hetal’s narrative account. Kasia wrote and thanked Hetal for her very useful narrative account and resolved to allow her two remaining interviewees to act as narrators, using their own voices to tell their stories.

Kasia started the next online interview with a list of elements and themes in which she was interested but resolved that her second participant, Jorg, should be allowed to use his own voice. Jorg provided Kasia with another full and useful narrative, with Kasia acting as listener rather than traditional interviewer, only seeking clarification occasionally, after explaining the nature and purpose of the process as they started. Kasia left this second interview feeling very pleased and looked forward to the next one.

order ‘to understand the complex processes which people use in making sense of their organisational realities’ (Musson 2004: 42). Chase (2011) distinguishes between asking participants to generalise when answering questions in more structured types of qualitative research and being invited to provide a complete narrative of their experience.

Narrative Inquiry preserves chronological connections and the sequencing of events as told by the narrator (participant) to enrich understanding and aid analysis. Chase (2011: 421) refers to this strategy as providing the opportunity to connect events, actions and their consequences over time into a ‘meaningful whole’. Through storytelling the narrator provides their interpretation of these events, allowing you to analyse the meanings which the narrator places on them. Where there is more than one participant providing a personal account of a given context, the narrative researcher will also be able to compare and to triangulate or contrast these narratives.

A **narrative** is a personal account of an experience that is told in a sequenced way, indicating a flow of related events that, taken together, are significant for the narrator and which convey meaning to the researcher (Coffey and Atkinson 1996). Such narratives are likely to contain ‘thick descriptions’ of contextual detail and social relations. Gabriel (2018) argues that using narratives can offer powerful insights into the meanings accorded to events and experiences.

In Narrative Inquiry, the participant is the narrator, with the researcher adopting the role of a listener facilitating the process of narration (Box 5.9). The narrative provided may be a short story about a specific event, a more extended story (for example, about a work project, managing or setting up a business, or an organisational change programme) or a complete life history (e.g. Chase 2011; Maitlis 2012). While in-depth interviews are the most widely used method to collect stories, other methods such as participant observation (Coffey and Atkinson 1996), autobiographies, authored biographies, diaries, documents and informal discussions (Chase 2011; Maitlis 2012) may also be used. This raises the issue of the researcher adopting the role of narrator in particular circumstances, which we will consider later. Narrative Inquiry may be used as the sole research strategy, or in conjunction with another strategy as a complementary approach (Musson 2004).

Narrative Inquiry may be used in different ways. It may be used with a very small sample of one to three participants, selected as being typical of a much larger population (Chase 2011) (Box 5.9). Alternatively, a small sample may be selected as critical cases or extreme cases, such as company founders or entrepreneurs, from whom much may be learnt. Narrative Inquiry may also be used with slightly larger samples of participants from across an organisation, to analyse how narratives are constructed around an event or series of events. This would allow comparisons between accounts to establish whether they differ, such as between departments, occupational groups or genders.

The strategy is generally associated with small, purposive samples (Section 7.3) because of its intensive and time-consuming nature. It is likely to generate large amounts of data in the form of the narrative account, or of interview transcripts or observational notes.



Box 5.10 Checklist

Structural element questions to facilitate narrative analysis

- ✓ What is the story about?
- ✓ What happened, to whom, whereabouts and why?
- ✓ What consequences arose from this?
- ✓ What is the significance of these events?
- ✓ What was the final outcome?

Source: Developed from Coffey and Atkinson (1996)

The narratives that emerge may not be easy-to-use or in a structural and coherent form (Gabriel 2018). Box 5.10 outlines structural elements to facilitate analysis of narratives.

To achieve analytical coherence in a narrative account you may need to (re)construct the story from one or more in-depth interviews with one participant, or a number of interviews with different participants. This places you as the narrative researcher in a central role in telling the story. Decisions will need to be taken about what to include and what to leave out, and how to connect parts of the account (Section 13.10).

While analysis in Narrative Inquiry does not use the analytical fragmentation of Grounded Theory, neither does it offer a well-developed set of analytical procedures comparable to those used by grounded theorists. Despite this, analytical rigour is still important in order to derive constructs and concepts to develop theoretical explanations. While narrative researchers may believe that predefined analytical procedures are neither advisable nor desirable, this may make the task of analysis more demanding for you (Section 13.10).

Archival and documentary research

Data digitalisation, the rapid growth of online archives, and open data initiatives by governments and businesses, mean there is considerable scope for you to use an archival or a documentary research strategy. It is now possible to access such sources online from around the world through online data archives and gateways to governmental websites (Chapter 8). Organisations' websites may provide access to certain types of documentary sources such as annual reports, company results, financial highlights, press releases and regulatory news. Media websites also provide facilities to search for articles about organisations and business and management topics. Some documents created by individuals may be accessible through data archives (e.g. a collection of papers of a notable businessperson) but use of recently created materials will probably require you to contact a potential participant to negotiate access, where these are not considered to be private or commercially sensitive. An **archival research** strategy uses manuscripts, documents, administrative records, objects, sound and audio-visual materials held in archives, special collections and other repositories as the main sources of data. A **documentary research** strategy uses personal and official documents as the sources of data.

From our initial discussion it clear there is a wide range of potential archival and documentary materials available. Lee (2012: 391) suggests that 'a document is a durable repository for textual, visual and audio representations that may be retained and used in different times and spaces, creating the possibility that meanings may be interpreted differently'. This illustrates the wide range of materials and that their use and interpretation can vary. Categories of textual documents include:

- communications between individuals or within groups such as emails, tweets, letters, social media and blog postings;
- individual records such as diaries, calendars and notes;
- organisational documents such as administrative records, agendas and minutes of meetings, agreements, contracts, memos, personnel records, plans, policy statements, press releases, reports and strategy statements;
- government documents such as publications, reports and national statistics data sets;
- media documents including online and printed articles and other data;
- visual and audio documents include advertising posters, artefacts, audio recordings, audio-visual corporate communications (e.g. YouTube videos), digital recordings, DVDs, films, photographs, products, promotional advertisements and recordings, and television and radio programmes.

Documents used for research are considered secondary sources because they were created original for a different purpose. However, there is a significant difference between re-analysing data collected originally for a research purpose and using secondary sources in an archival or documentary research strategy. Where previously collected research data are re-analysed for a different purpose in a secondary data analysis, the quality of the original research data needs to be assessed. For example, how was the survey sample selected? Was the original research designed to overcome threats to reliability and validity (Section 5.8)? In contrast, where documents are used as secondary data in an archival or documentary research strategy, their original purpose had nothing to do with research and so you will need to be sensitive to their nature and original purpose, the way in which you analyse them and the generalisations that you can draw (Hakim 2000).

Data from archives documents may be analysed quantitatively, qualitatively or both. Analysing textual documents qualitatively can enable you to generate a rich or ‘thick’ description of key events, the context within which these events occurred, the roles of the actors involved, the influence of external influences such as economic or commercial pressures, as well as outcomes. Your scope to achieve such an outcome will depend on whether you find suitable documents. Documents may, for example, allow you to analyse critical incidents or decision-making processes, or evaluate different policy positions or strategies. Using quantitative data from documents such as annual or financial reports may, for example, facilitate comparisons between organisations or across reporting periods. Prior (2007) points out that documents can also be analysed to reveal:

- not only what they contain but what is omitted;
- which facts are used and why these might be emphasised while others are not used;
- how they are used in an organisation and how they are circulated and to whom.

The utility of archival or documentary research strategies will depend on their appropriateness to your research question and objectives and gaining access to sufficient suitable documents. You may be refused access to documents or find some data are restricted for confidentiality reasons. You may also find that the documents you locate vary in quality, especially where they come from different sources. Data may be missing or presented inconsistently, making comparison difficult or potentially leaving gaps in your analysis (Box 5.11). Using an archival or documentary research strategy will therefore necessitate establishing what documents are available and designing your research to make the most of these. This may mean combining this research strategy with another; for example, conducting documentary research alongside a Grounded Theory strategy based on qualitative interviews and using similar procedure to analyse both sets of data. Alternatively, you might use documentary research within a case study strategy.

Case study

A **case study** is an in-depth inquiry into a topic or phenomenon within its real-life setting (Yin 2018). The ‘case’ in case study research may refer to, for example, a person, group, organisation, association, process or event. Choosing the case to be studied and determining the boundaries of the study is a key factor in defining a case study (Flyvbjerg 2011). Once defined, case study research sets out to understand the case within its setting or context (Eisenhardt and Graebner 2007).

The study of a case within its real-life setting or context helps to distinguish this research strategy from others. In an experimental strategy, contextual variables are highly controlled as they are seen as a potential threat to the validity of the results. In a survey strategy, research is undertaken in a real-life setting, but the ability to understand the



Box 5.11 Research in the news

Historians having to tape together records that Trump tore up

An article in *The Guardian* in January 2021 highlighted growing concern that Donald Trump's White House Records would be incomplete due to his habit of ripping up papers before discarding them. This had led to officials spending hours taping records back together. Officials also had to be reminded not to conduct official business using text messaging apps or private emails and to preserve it if they did. While it is estimated that computer systems have captured the vast majority of records, they have not been able to capture those records that were not created or logged into the system.

In 2021 Trump's electronic and paper records were transferred to the United States National Archives. While the Biden administration could see these documents immediately, members of the public are likely to have to wait. Trump, like other presidents, has restricted public access for up to 12 years.



Source: Staff and Agencies (2021) 'Historians having to tape together records that Trump tore up', The Guardian 17 January [online].

impact of context is limited by the number of variables for which data can be collected. In contrast, case study research is often used when the boundaries between the phenomenon being studied and the context within which it is being studied are not always apparent (Yin 2018).

A case study strategy has the capacity to generate insights from intensive and in-depth research into the study of a phenomenon in its real-life context, leading to rich, empirical descriptions and the development of theory (Yin 2018). They can be designed to identify what is happening and why, and to understand the effects of the situation and implications for action; often using both qualitative and quantitative data from a range of sources. Although case studies have been widely used over a long period, including in business and management, they have been criticised by some because of 'misunderstandings' about their ability to produce generalisable, reliable and theoretical contributions to knowledge (Flyvbjerg, 2011). This is largely based on positivist criticisms of using small samples and more generally about using interpretive, qualitative research. This type of criticism has been countered and is generally losing favour as the value of qualitative and mixed methods research is recognised more widely (Denzin and Lincoln 2018).

Case studies are designed in different ways dependent upon their purpose. They have been used for descriptive, exploratory or explanatory purposes by 'positivist' as well as 'interpretivist' researchers both deductively and inductively. Some positivist researchers have also advocated using case studies inductively to build theory and to develop theoretical hypotheses, which can be tested subsequently. In this way, the use of the case study is advocated in the early, exploratory stage of research as a complement to deductive research (Eisenhardt and Graebner 2007). This approach has been called 'indicative case study research', designed to reveal 'specific attributes' rather than rich description (Ridder et al. 2014: 374).

Yin (2018) recognises that case studies may be used not only for exploratory but also descriptive and explanatory purposes. An explanatory case study is likely to use a deductive approach, testing the applicability of theoretical propositions, to build and verify an explanation (Chapter 13). Interpretivist researchers are more interested, at least initially, to develop richly detailed and nuanced descriptions of their case study research (Ridder et al. 2014). For some interpretivists, making comparisons with existing theory is unnecessary. Stake (2005) notes many interpretivist researchers prefer to describe their case study in ample detail, allowing readers to make their own links to existing theory. Other interpretivist researchers inductively analyse their data, identifying themes and patterns and, at some point, locating this in existing literature in order to refine, extend or generate theory (Ridder et al. 2014; Chapter 13). If you are an interpretivist, it is highly likely that you will need to follow this second route and provide a clear link to theory!

Orthodox and emergent designs

Lee and Saunders (2017) differentiate between research designs for ‘orthodox cases’ and ‘emergent cases’. An **orthodox case study** strategy involves an approach that is rigorously defined and highly structured before the research commences, with the intention that it will proceed in a linear way. This reflects the rational approach to conducting research where literature is reviewed first, the research question is defined, the research project is designed, preparation for the conduct of the research undertaken, and data are collected, analysed, interpreted and then reported. This approach to case study strategy is likely to be underpinned by realist philosophical assumptions. An **emergent case study** strategy involves you strategically choosing a case study environment within which research will be conducted and allowing the focus of the research to emerge through the different stages of data collection and analysis incorporating relevant literature. This approach is likely to be underpinned by interpretivist or constructivist philosophical assumptions.

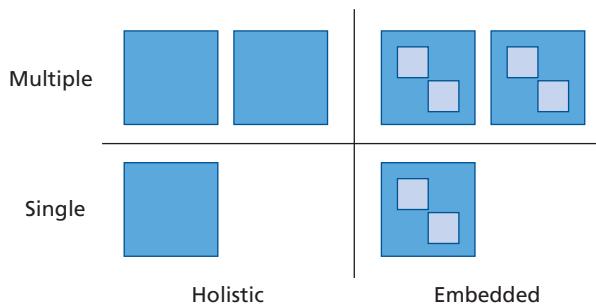
The existence of various case study designs offers both opportunities and challenges. Where you are considering using a case study strategy, you may be able to find earlier work in the social sciences, if not specifically in business and management, which provides guidance in an approach that fits logically with your research idea and question (deductive or inductive, exploratory or explanatory etc.). To achieve an in-depth inquiry and a rich, detailed flow of analytical data, a case study strategy can use a mixed methods research design (although case studies may rely on a multi-method design). Case study research often uses a combination of archival records and documents (Chapter 8), different forms of observation (Chapter 9), interviews and focus groups (Chapter 10), questionnaires (Chapter 11), reflection and the use of research diaries and other research aids (Chapters 1 and 13). Case study research is likely to prove to be challenging because of its intensive and in-depth nature, and your need to be able to identify, define and gain access to a case study setting.

Case study structures

Yin (2018) distinguishes between four case study strategies based upon two discrete dimensions (Figure 5.5):

- single case versus multiple cases;
- holistic case versus embedded case.

A single case is often used where it represents a critical case or, alternatively, an extreme or unique case. Conversely, a single case may be selected purposively because it is typical or because it provides you with an opportunity to observe and analyse a phenomenon that few have considered before (Section 7.3). Many part-time students use the

**Figure 5.5** Case study structures

Source: Developed from Yin (2018)

organisation for which they work as a single case study (Box 5.12). The key here will be to ensure the case study strategy is appropriate for the nature of your research question and objectives.

A case study strategy can also incorporate multiple cases to establish whether findings can be replicated across cases. Here you will select cases on the basis that similar results are predicted to be produced from each one. Where this is realised, this is termed **literal replication** (Yin 2018). Alternatively, cases may be chosen where a contextual factor is deliberately different. The impact of this difference on the anticipated findings is predicted by the researcher. Where this predicted variation is realised, Yin terms this **theoretical replication**.

A multiple case study strategy may combine a small number of cases chosen to predict literal replication and a second small number chosen to predict theoretical replication (Yin 2018). Where all of the findings from these cases are as predicted, this would produce very strong support for the theoretical propositions on which these predictions were based. This approach commences deductively, using data to test theoretical propositions before, possibly, incorporating an inductive or abductive approach (Section 4.5). Where the findings are in some way contrary to the predictions in the theoretical propositions being tested, it would be necessary to reframe these propositions and select further cases to test them. Choosing between a single or multiple case study is not simply related to producing more evidence. While a multiple case study is likely to produce more evidence, the purpose of



Box 5.12 Focus on student research

Using a single organisation as a case study

Simon was interested in discovering how colleagues within his organisation were using a recently introduced financial costing model in their day-to-day work. In discussion with his project tutor, he highlighted that he wished to find out how it was actually

being used in his organisation as a whole, as well as seeing if the use of the financial costing model differed between senior managers, departmental managers and front-line operatives. Simon's project tutor suggested that he adopt a case study strategy, using his organisation as a single case within which the senior managers', departmental managers' and front-line operatives' groups were embedded cases. He also highlighted that, given the different numbers of people in each of the embedded cases, Simon would be likely to need to use different data collection techniques with each.

each approach is different. A single case study approach is chosen because of the nature of the case. A multiple case study approach is chosen to allow replication.

Yin's second dimension, holistic versus embedded, refers to the unit of analysis. You may have chosen to use an organisation in which you worked or are currently employed as your case. If your research is concerned only with the organisation as a whole, then you are treating the organisation as a holistic case study. Conversely, even if you are only researching within a single organisation, you may wish to examine one or more sub-units within the organisation, such as departments or work groups. Your case will inevitably involve more than one unit of analysis and, whichever way you select these units, is called an embedded case study (Box 5.12).

Action Research

Lewin first used the term Action Research in 1946. It has been interpreted subsequently by management researchers in a variety of ways, but a number of common and related themes have been identified within the literature. An **Action Research** strategy is an emergent and iterative process of inquiry that is designed to develop solutions to real organisational problems through a participative and collaborative approach, uses different forms of knowledge, and will have implications for participants and the organisation beyond the research project (Coghlan 2011; Coghlan 2019). Our definition identifies five themes: purpose, process, participation, knowledge and implications, which we now consider.

Purpose

The purpose of an Action Research strategy is to promote organisational learning to produce practical outcomes through identifying issues, planning action, taking action and evaluating action. Coghlan (2019) emphasises Action Research is about research in action rather than research about action. This is because Action Research focuses on 'addressing worthwhile practical purposes' (Reason 2006: 188) and resolving real organisational issues (Shani and Pasmore 1985).

Process

The process of Action Research is both emergent and iterative. An Action Research strategy commences within a specific context and with a research question but because it works through several stages or iterations the focus of the question may change as the research develops. Each stage of the research involves a process of diagnosing or constructing issues, planning action, taking action and evaluating action (Figure 5.6). Diagnosing or constructing issues, sometimes referred to as fact finding and analysis, enables action planning and a decision about the actions to be taken. These are then taken, and the actions evaluated to tease out the issues (cycle 1). This evaluation provides a direction and focus for the next stage of diagnosing or constructing issues, planning action, taking action and evaluating action to understand the customer and project (cycle 2), demonstrating the iterative nature of the process. Subsequent cycles (cycle 3 and possibly beyond) involve further diagnosing or constructing of issues, taking into account previous evaluations, planning further actions, taking these actions and evaluating them; thereby acting on knowledge. Action Research differs from other research strategies because of its explicit focus on action related to multiple stages, to explore and evaluate solutions to organisational issues and to promote change within the organisation.

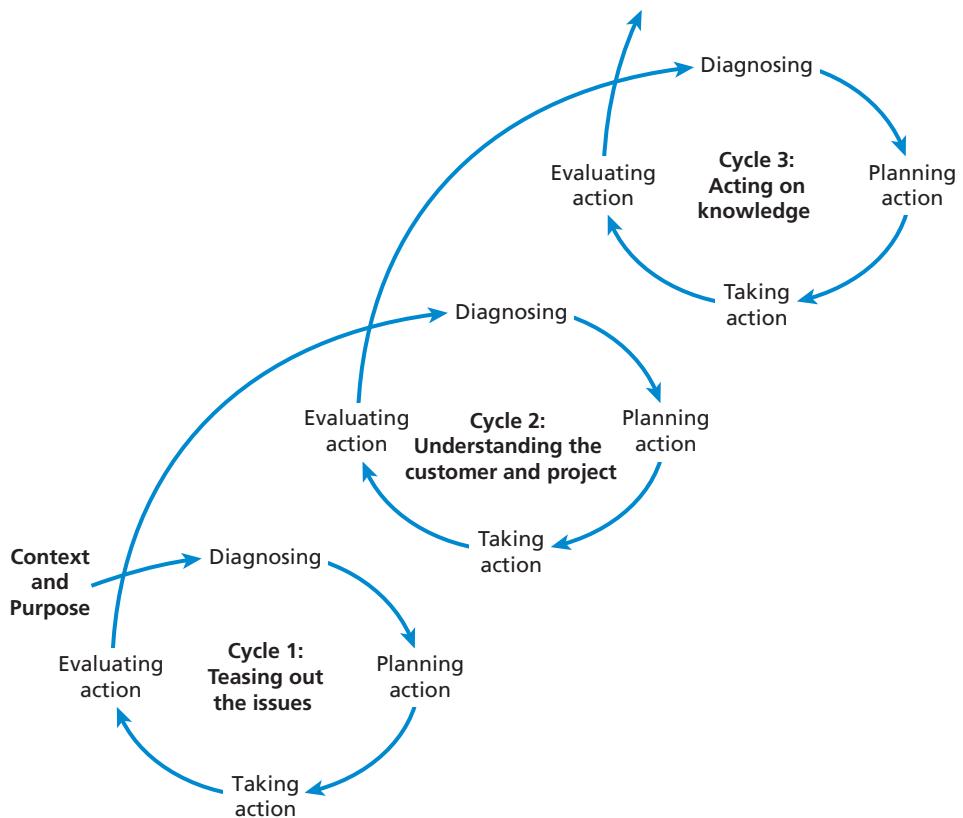


Figure 5.6 The three cycles of the Action Research spiral

Participation

Action Research is a social process in which an action researcher works with members in an organisation, as a facilitator and teacher, to improve the situation for these participants and their organisation. For Greenwood and Levin (2007) it can only be called Action Research if research, action and participation are all present. Organisational members need to cooperate with the researcher to allow their existing work practices to be studied. The process of Action Research then requires collaboration through its iterative cycles (Figure 5.6) to facilitate the improvement of organisational practices. This means building a democratic approach to communication and decision-making throughout each Action Research stage or cycle. The researcher passes on her or his skills and capabilities to organisational members so that they effectively become co-researchers in the Action Research process. Without such participation, this approach simply would not be viable, although creating such participation is likely to be difficult in practice and to meet with resistance at various levels (Reason 2006).

Participation of organisational members results usually from their involvement in ‘a matter which is of genuine concern to them’ (Eden and Huxham 1996: 75). Members of an organisation are more likely to implement change they have helped to create (Schein 1999). Once the members of an organisation have identified a need for change and have widely shared this need, it becomes difficult to ignore, and the pressure for change comes

from within the organisation. In this way, an Action Research strategy combines both data gathering and the facilitation of change.

Knowledge

Action is informed by both abstract theoretical knowledge, known as propositional knowledge; and participants' everyday lived experiences (their experiential knowledge) and knowing-in-action (knowledge that comes from practical application) (Reason 2006). These forms of knowledge are important in each stage or cycle of the Action Research process, encouraged by the collaborative approach that underpins this strategy. They lead to 'actionable knowledge' that has the potential to be useful to organisational practitioners as well as being academically robust (Coghlan 2011: 79). Coghlan believes that Action Research not only affects 'what we know' but emphasises understanding of 'how we know'.

Implications

The implications of Action Research go beyond the research project. Participants in an organisation where action research takes place are likely to have their expectations about future treatment and involvement in decision making raised (Greenwood and Levin 2007). There are also likely to be consequences for organisational development and culture change. Implications from the process may be used to inform other contexts. Academics will use the results from undertaking Action Research to develop theory that can be applied more widely. Consultants will transfer knowledge gained to inform their work in other contexts. Such use of knowledge to inform, we believe, also applies to others undertaking Action Research, such as students undertaking research in their own organisations.

There are a number of practical concerns to consider if you are thinking of using an Action Research strategy. Identifying an accommodating context, the emergent nature of this strategy, the need to engender participation and collaboration, the researcher's role as facilitator, and the stages or iterations involved are some of the reasons that make Action Research a demanding strategy in terms of the intensity involved and the resources and time required. Action Research can be suited to (part-time) students who undertake research in their own organisation, although its longitudinal nature means that it is more appropriate for medium- or long-term research projects. There is the related issue of deciding how many Action Research cycles are sufficient. Where these practical as well as political concerns have been properly anticipated and evaluated in terms of a feasible design, Action Research has the potential to offer a worthwhile and rich experience for those involved.

5.6 Considering time horizons

An important question to be asked in designing your research is, 'Do I want my research to be a "snapshot" taken at a particular time or do I want it to be more akin to a diary or a series of snapshots and be a representation of events over a given period?' This will, of course, depend on your research question. The 'snapshot' time horizon we call **cross-sectional**, while the 'diary' perspective we call **longitudinal**.

Cross-sectional studies

It is probable that your research will be cross-sectional, involving the study of a particular phenomenon (or phenomena) at a particular time. We say this because we recognise that

most research projects undertaken for academic courses are necessarily time constrained. However, the time horizons on many courses do allow sufficient time for a longitudinal study, provided, of course, that you start your research early!

Cross-sectional studies can use quantitative, qualitative and mixed methods research designs and a correspondingly wide variety of strategies. They may be using a survey strategy and a quantitative design to describe the incidence of a phenomenon such as the IT skills possessed by managers in one organisation at a given point in time. Alternatively, they may be using a case study strategy and both quantitative and qualitative data to examine how small UK businesses are trading with Europe post Brexit.

Longitudinal studies

The main strength of longitudinal research is its capacity to study change and development. This type of study may also provide you with a measure of control over some of the variables being studied. One of the best-known examples of this type of research comes from outside the world of business. It is the long-running UK television series *Seven Up*. This has charted the progress of a cohort of people every seven years of their life since 1964 (*63 Up*, 2019). Not only is this fascinating television, it has also provided social scientists with a rich source of data on which to test and develop theories of human development.

Longitudinal studies can use quantitative, qualitative and mixed methods research designs and a correspondingly wide variety of strategies. They may be using an experiment strategy and a quantitative design to establish the impact of a particular intervention. Alternatively, they may be using an ethnography strategy and qualitative data to examine how home working practices are adapting post pandemic.

Even with time constraints it is possible to introduce a longitudinal element to your research. There is a massive amount of published data collected over time just waiting to be reanalysed (as Section 8.2 indicates)! An example is the Edelman Trust Barometer, an annual trust and credibility survey undertaken every year since 2001 (Edelman, 2021). From these surveys you can gain valuable secondary data, which gives a global measurement of trust across the world and how it is changing with regard to government, businesses, media and non-governmental organisations (NGOs).

5.7 Anticipating potential ethical issues

Research ethics are a critical part of formulating your research design. While this is discussed in detail in Chapter 6, which focuses on issues associated with negotiating access and research ethics, it is crucial that you anticipate and address potential ethical issues at the design stage. In particular, Section 6.6 highlights ethical issues associated with specific aspects of the research process, and all these need to be considered now. Here we focus on two ethical issues that need to be considered at the start of designing your research:

- exposing your research subject(s) and yourself to harm;
- researching covertly.

Harm

Your choice of topic and how you collect you data will be governed by the need to minimise the risk of harm, embarrassment, pain or any other material disadvantage to those

involved in the research. You may be particularly interested to study the consumer decision to buy flower bouquets. Although this may provide some data collection challenges (who buys, for whom and why), there are not the same ethical difficulties as will be involved in studying, say, the funeral purchasing decision. Your research design in this case may have to concentrate on data collection from the undertaker and, ideally, the purchaser at a time as close to the death as delicacy permits; although the later would be considered insensitive. It is a matter of judgement as to whether the strategy and data collection method(s) suggested by ethical considerations will yield data that are valid.

Risk of harm does not only relate to potential research subjects. As a researcher you have a duty to protect yourself from harm. While it could be exciting to research the informal economy and use of the dark web for your research project, this could carry significant risks of harm to yourself, as well as the possibility of being involved in illegal activities. Alternatively, your research might necessitate collecting data in places considered unsafe. Most universities now require students to complete a risk assessment for their research projects and explain how risk would be minimised. Not surprisingly, projects where there is a significant risk of harm to you as researcher are unlikely to be approved. The general ethical issue here is that the research design should not subject those involved to the risk of embarrassment, pain, harm or any other material disadvantage.

Covert research

You may also need to consider whether you should collect data covertly, in other words where those you are researching are unaware they are the subject of research, and so have not consented. For example, if you plan to adopt an ethnographic strategy observing informants and intend not to reveal yourself as researcher you will need to explain ethically why this is necessary and appropriate. Beware, although covert research such as undertaking observation in a public place is usually considered acceptable, many university research ethics procedures preclude the use of any covert research. Circumstances related to the use of covert observation and issues related to privacy are considered rather in Sections 6.6 and 9.2.

5.8 Assessing the quality of research design

Underpinning our discussion of research design is the issue of research quality. This is neatly expressed by Raimond (1993: 55) when he subjects findings to the ‘how do I know?’ test, ‘Will the evidence and my conclusions stand up to the closest scrutiny?’ For example, how do you know that the advertising campaign for a new product has resulted in increased sales? How do you know that manual employees in an electronics factory have more negative feelings towards their employer than their clerical counterparts? The answer, of course, is that, in the literal sense of the question, you cannot know. All you can do is reduce the possibility of getting the answer wrong.

A key concern in designing your research will be to familiarise yourself with the criteria to be used to assess your research project. These assessment criteria might state that your research design and report have to consider issues of reliability or dependability, validity or credibility, or authenticity. Other assessment criteria will be generic, related to analytical and evaluative abilities, only implicitly recognising the need for reliable/dependable and valid/credible/authentic research in assessing your research design and outcomes. Familiarising yourself with the assessment criteria to be used will help you to decide how you should approach the way you describe and discuss the quality of your research.

A split often occurs at this point between quantitative and qualitative research designs and the associated philosophical underpinnings. For example, positivists use the ‘canons of scientific inquiry’ related to reliability and validity to assess the quality of research, while interpretivists adapt the terms ‘reliability’ and ‘validity’ to assess their research, or reject them as inappropriate (Lincoln et al. 2018). We briefly discuss each of these approaches to establish and assess research quality.

Scientific canons of inquiry: reliability and validity

Reliability and validity are central to judgements about the quality of research in the natural sciences and quantitative research in the social sciences. Their role in relation to qualitative research is contested, as we discuss later. **Reliability** is the extent a data collection procedure yields consistent findings. If a researcher is able to replicate an earlier research design and achieve the same findings, then that research would be considered reliable. **Validity** is the extent these procedures measure accurately what they are intended to measure, and the research findings are about what they profess to be about:

- 1** Do the measures being used in the research to assess the phenomenon being studied actually measure what they are intended to – are they appropriate for their intended purpose?
- 2** Are the analysis of the results and the relationships being advanced accurate?
- 3** What do the research findings represent: does the claim about how generalisable they are stand up?

This first of these is sometimes termed **measurement validity** and is associated with face validity, construct validity, content validity and predictive validity (Section 11.4). The second refers to internal validity and the third to external validity.

Reliability

When considering reliability, a distinction is made between internal reliability and external reliability. Internal reliability refers to ensuring consistency during a research project. This may be achieved, if permitted, by using more than one researcher within a research project. As researchers conduct interviews or observations and code and analyse data they will be able to evaluate they agree and their consistency of use. You can help promote consistency through the stages of your research project by writing memos and keeping detailed notes about how you have coded, analysed and interpreted your data. External reliability refers to whether your data collection procedures and analysis techniques would produce consistent findings if you repeated them on another occasion, or they were replicated by another researcher. Ensuring reliability can be difficult as there are a number of threats to reliability (Table 5.6). Research that is unreliable will also prove to be invalid since any error or bias in data collection and analysis can affect the results and subsequent interpretation, casting doubt on the research.

These threats emphasise the need for methodological rigour in your research. More specific advice appears in other chapters, but one key aspect is to ensure that your research process is systematic, being clearly thought through and evaluated and not containing ‘logic leaps and false assumptions’. To allow others to judge your work, you need to report and justify your methods fully and transparently.

Reliability is a key characteristic of research quality; however, while necessary, it is not sufficient by itself to ensure good-quality research. The quality of research depends also on its validity.

Table 5.6 Threats to reliability

| Threat | Definition and explanation |
|-------------------|--|
| Participant error | Factor altering adversely the way in which a participant or respondent answers or acts. For example, asking a respondent to complete a questionnaire just before a lunch break may affect the way they respond compared to choosing a less sensitive time (i.e. they may not take care and hurry to complete it) |
| Participant bias | Factor inducing a false response. For example, conducting an interview in a public space may lead participants to provide falsely positive answers where they fear they are being overheard, rather than retaining their anonymity |
| Researcher error | Factor altering the researcher's interpretation. For example, being tired or not sufficiently prepared, and misunderstanding some subtle meanings in participants' responses |
| Researcher bias | Factor inducing bias in recording or interpretation of participants' or respondents' responses. For example, allowing subjective views or dispositions to prevent fair and accurate recording and interpretation of participants' responses |

Validity

Internal validity is the extent your findings can be attributed to the intervention you are researching rather than to flaws in your research design. For example, in an experiment, internal validity would be established where an intervention can be shown statistically to lead to an outcome rather than this having been caused by some other confounding variable acting at the same time. In a survey strategy using questionnaires, **criterion validity**, is whether the questions are actually measuring what they are intended to measure, thereby allowing accurate statistical predictions to be made (Section 11.3). These concepts are associated with quantitative research designs. Forms of measurement validity are discussed in Section 11.4.

Your research findings are considered invalid if a finding has been arrived at falsely or when an inaccurate relationship reported and, the most frequent causes are outlined in Table 5.7. Research that produces invalid results and conclusions will also be unreliable as it is highly unlikely a subsequent study will find the same false results and statistical relationships.

External validity is the extent research findings from a particular study are generalisable to other relevant contexts. For example, a corporate manager may ask, 'Can the findings from the research study in one organisation in our corporation also be used to inform policy and practice in other organisations in the group?' The chief executive of a multinational organisation may ask, 'Are the findings from the survey in the Finance and Resources Department applicable to other departments in the organisation?' Just as researchers take great care when selecting a sample from within a population to make sure that it represents that population, researchers and their clients are often concerned to establish the generalisability of their findings to other contexts. Even in such cases, however, it will be necessary to replicate the study in that other context, or contexts, to establish generalisability.

Table 5.7 Threats to internal validity

| Threat | Definition and explanation |
|----------------------------------|---|
| Past or recent events | Event changing participants' or respondents' perceptions. For example, a vehicle manufacturer recalling cars for safety modifications affecting customers' views about product quality (unless the objective of the research is to find out about post-product recall opinions) |
| Testing | Impact of testing on participants' or respondents' views or actions. For example, informing participants about a research project may alter work behaviour or responses during the research, if they believe it might lead to future consequences for them |
| Instrumentation | Impact of changes to a research instrument during a research project affecting comparability of results. For example, in structured observational research on call centre operations, the definitions of behaviours being observed may be altered between stages, making comparison difficult |
| Mortality | Impact of participants or respondents withdrawing from a study. For example, participants or respondents leaving an organisation due to promotion during a study |
| Maturation | Impact of a change in participants or respondents outside of the influence of the study that affects their attitudes or behaviours etc. For example, management training may mean participants or respondents revise their responses during a subsequent research stage |
| Ambiguity about causal direction | Lack of clarity about cause and effect. For example, during a study, it was difficult to say whether poor performance ratings were caused by negative attitudes to appraisal or, negative attitudes to appraisal were caused by poor performance ratings |

Alternative criteria to assess the quality of research design

The types of measurement validity and reliability just discussed in relation to quantitative research based on positivist assumptions, are often considered as philosophically and technically inappropriate for qualitative research based on interpretive assumptions. If these concepts are applied rigidly, where reality is regarded as being socially constructed and multifaceted, there is a mismatch between the nature of the research and how it is being judged. It therefore becomes difficult to demonstrate that qualitative research designs are of high quality and credible.

Three types of response to this are evident:

- adaptation of existing conceptualisations of validity and reliability;
- use of parallel versions with distinct names;
- use of alternative authenticity criteria.

Adaptation of existing conceptualisations

Those who continue to use the concepts of reliability and validity, adapting them to qualitative research, believe generally that, since all research needs to be reliable and valid, using these terms is important to be able to demonstrate the quality and comparable status of qualitative research. Qualitative research is not necessarily intended to be replicable, as it reflects the socially constructed interpretations of participants in that particular setting at the time it is conducted (Section 10.4). However, rigorous description of the research design, context and methods help others assess how it was undertaken and enable them to conduct similar studies. Use of more than one interviewer, observer and data can, where possible, also improve internal reliability of the research. Adaptation of the concept of internal validity to qualitative research is generally not considered a problem since the in-depth nature of qualitative methods means findings can be shown to be well grounded in rich data. In contrast, adaptation of external validity has been questioned as small samples limit the generalisability of qualitative studies. Other forms of generalisability can, however, demonstrate the quality and value of qualitative research. For example, findings from one qualitative research setting may lead to generalisations across other settings, where, for example, characteristics of the research setting are similar, or where learning from the research setting can be applied in another setting.

Use of parallel versions

Parallel versions of reliability, internal validity and external validity have been developed with distinct names that recognise the nature of qualitative research. In this regard, Lincoln and Guba (1985) formulated ‘dependability’ for ‘reliability’, ‘credibility’ for ‘internal validity’ and ‘transferability’ for ‘external validity’ (Table 5.8).

Alternative criteria

Others have moved further away from the concepts of reliability and validity developing new concepts through which to ensure and judge the quality of qualitative research. In this regard, Guba and Lincoln (1989) and Lincoln et al. (2018) have developed ‘authenticity criteria’ as an alternative to validity for constructivist and interpretivist research (Table 5.9).

Validation

In our discussion about assessing quality and alternative criteria to evaluate it we have already referred to techniques of validation without using this term. **Validation** is the process of verifying research data, analysis and interpretation to establish their validity/credibility/authenticity. We now discuss two techniques that, incorporated into your research design, can help validate the quality of your research:

- triangulation;
- participant or member validation.

Triangulation

Triangulation involves using more than one source of data and method of collection to confirm the validity, credibility or authenticity of research data, analysis and interpretation. This necessitates using a multi-method quantitative study, multi-method qualitative study or a mixed methods study (Section 5.4). Two or more independent sources of data and methods of collection are used within one study to ensure that the data are telling

Table 5.8 Parallel quality criteria

| Criterion | Definition and techniques to achieve each |
|-----------------|--|
| Dependability | This is the parallel criterion to reliability. In interpretivist research, the focus is likely to be modified as research progresses. Dependability means recording all these changes to produce a reliable/dependable account of the emerging research focus that may be understood and evaluated by others |
| Credibility | This is the parallel criterion to internal validity. Emphasis is placed on ensuring that the representations of research participants' socially constructed realities match what participants intended. A range of techniques to ensure this include: <ul style="list-style-type: none"> • lengthy research involvement to build trust and rapport and to collect sufficient data; • reflection using a different person to discuss ideas and test out findings etc.; • developing a thorough analysis that accounts for negative cases, refining the analysis in order to produce the best possible explanation of the phenomenon being studied; • checking data, analysis and interpretations with participants; • ensuring the researchers' preconceived expectations regarding what the research will reveal are not privileged over social constructions of the participant by regularly recording these and challenging them during analysis of the data. |
| Transferability | This is the parallel criterion to external validity or generalisability. Providing a full description of the research questions, design, context, findings and interpretations, allows the reader to judge the transferability of the study to other settings |

Sources: Developed from Guba and Lincoln 1989; Lincoln et al. 2018

Table 5.9 Alternative quality criteria

| Criterion | Definition |
|--|---|
| Fairness | Balanced representation, so that all stakeholder views, perspectives, claims concerns and voices are represented. |
| Ontological and educative authenticity | Raising of the level of awareness about the research with participants and those with whom they come in contact. |
| Catalytic and tactical authenticities | Prompting action by the participants, and the involvement of the researcher(s) in training participants in social and/or political action |

Sources: Developed from Guba and Lincoln 1989; Lincoln et al. 2018

you what you think they are telling you. In a research study based on positivist assumptions, this will help reveal the 'reality' in the data. Interpretivist researchers challenge this outcome arguing that for participants, 'reality' is socially constructed and multifaceted. For

interpretivists, the value of triangulation is it adds depth, breadth, complexity and richness to their research (Denzin 2012; Denzin and Lincoln 2018).

Participant or member validation

Participant or member validation involves taking or sending research data back to participants, allowing them to confirm the accuracy by commenting on and correcting it. This includes showing them interview transcripts, observational or other notes, storied accounts as well as researcher interpretations of participants. Engaging participants in collaboration is increasingly important in all research designs and member validation can be incorporated into quantitative, qualitative mixed methods research. While the anonymous nature of some aspects of quantitative designs may preclude member validation, you may still find it useful to discuss the results from your quantitative analysis with a sample of respondents to help you to understand and interpret these data. Member validation may be problematic when a participant wishes to withdraw some of the data shared with you. You will need to differentiate between cases where participants correct your interpretation of the data they shared with you and cases where they simply change their attitude. The latter scenario relates to an ethical concern and you will need to reflect on the extent to which you should alter the original data (Sections 6.5 and 6.6).

Logic leaps and false assumptions

So far in this chapter we have outlined a range of research design decisions to help ensure that your research yields sufficient good-quality data. These decisions necessitate careful thought. Your research design will need to be logical and, along with any assumptions you make, to stand up to careful scrutiny. Raimond (1993: 128) advises you to ‘stand back from your research [design] and take a critical, objective view of it, as though you were a detached observer’. This will allow you to see your design as others might, examining the research steps you propose in relation to your research question and objectives to see if they are logical, systematic and will stand up to rigorous scrutiny.

We also considered the issue of false assumptions and claims in the Introduction to Chapter 2 as you thought about choosing your research topic and developing your research proposal. Concern about making false assumptions and claims will continue to be a major issue as you design your research and then undertake it. Looked at this way, establishing the quality of your research is not an abstract idea but a tangible one that you need to be concerned about throughout your research project.

5.9 Recognising your role as researcher

This chapter has considered the decisions you need to take to formulate your research design and the interdependency between your research question and objectives, your research philosophy and your research purpose. You need to choose between quantitative, qualitative or mixed methods; between research strategies; and between time frames. Each decision will have implications for your design. Each decision also has implications for the ways in which you seek to establish a quality research design that is ethical. As you have read through this chapter, you have probably been evaluating each of these decisions in relation to practical constraints as well as personal preferences. We have alluded

to practical constraints in a number of places in the chapter in terms of the way they may affect each choice. A crucial practical consideration in deciding how to formulate a research design is related to your role as researcher.

The external researcher role

If you are a full-time student, you are likely to adopt the role of an **external researcher**. Where you intend to undertake research in one or a few organisations you will need to negotiate access to the organisation(s) and to those from whom you would like to collect data. Having achieved this you will need to gain their trust so that they will participate meaningfully to allow you to collect these data. You will need to take these practical factors into account when formulating your research question and your research design. Sections 6.2 to 6.4 provide more detail about issues of access that you need to take into account as an external researcher before finalising your research design.

In doing this you will need to consider the extent to which you engage with those from whom you collect data. For some research projects your engagement ends once data are collected. For others your engagement will be throughout the research process, working with an organisation or group to co-produce knowledge. Such **engaged scholarship** is a participative form of research in which you work with the organisation or group and obtain the advice and perspectives of key stakeholders to understand a complex problem (Van de Ven 2007). The engagement process is therefore a collaboration to produce high-quality research, reciprocate, address identified community (organisation) needs, where needed cross disciplinary boundaries, and support the democratisation of knowledge (Beaulieu et al. 2018). Van de Ven (2007) conceptualises engagement as the processes linking four points of a diamond emphasizing the importance in engaged scholarship of engaging:

- those who experience and know the problem in problem formulation;
- knowledge experts in theory building;
- methods experts and those providing access and data in the research design;
- the intended audience to interpret meanings and uses in problem-solving.

The internal researcher (practitioner researcher) role

If you are currently working in an organisation, you may choose to undertake your research project within that organisation and adopt the role of an **internal researcher** or **practitioner researcher**. As a part-time student, you will be surrounded by numerous opportunities to pursue business and management research. Indeed, like many people in such a position, you may be asked to research a particular problem by your employer.

As an internal researcher, another advantage will be your knowledge of the organisation. However, this advantage carries with it a significant disadvantage. You need to become conscious of the assumptions and preconceptions that you normally take-for-granted in your workplace. This is an inevitable consequence of knowing the organisation well and may prevent you from exploring issues that would enrich the research.

Familiarity may create other problems for you as an internal researcher. When we were doing case study work in a manufacturing company, we found it very useful to ask ‘basic’ questions revealing our ignorance about the industry and the organisation. These ‘basic’ questions are ones that as a practitioner researcher you would be less likely to ask because

you, and your respondents, would feel that you should know the answers already. There is also the problem of status. If you are a junior employee, you may feel that working with more senior colleagues inhibits your interactions as researcher practitioner. The same may be true if you are more senior than your colleagues.

A more practical problem is that of time. Combining two roles at work is obviously very demanding, particularly as it may involve you in much data recording ‘after hours’. This activity is hidden from those who determine your workload and they may not appreciate the extra demands of your researcher role. There are no easy answers to these problems. All you can do is be aware of the possible impact on your research of being too close to your research setting.

Tietze (2012) offers some guidance for internal researchers. These include reflecting on your role as internal researcher so that you may recognise how this affects the way you design and conduct your research (where you have scope to influence what you are going to research). The research you undertake and the report you produce may have implications for those you work with and you will therefore need to consider the implications of how you research and what you report (Section 6.6). You will need to consider your emotions and to manage these during this process of being an internal researcher. The process of analysing, interpreting and theorising about the research data you collect may have the effect of making ‘strange the all-too-familiar’ (Tietze 2012: 68) and you will need to cope with the degree of detachment that this may produce as you re-evaluate the way in which you view your organisation.

5.10 Summary

- Research design is the way a research question and objectives are operationalised into a research project. The research design process involves a series of decisions that need to combine into a coherent research project.
- The focus of your research will be exploratory, descriptive, explanatory, evaluative or a combination of these.
- A methodological choice has to be made regarding using quantitative or qualitative methods, or both, in a mono-method, multi-method or mixed methods research design.
- Methodological choice will be underpinned by your research philosophy and it is important to recognise the associated assumptions and implications of these.
- A decision has to be made to use one or more research strategies, related to the nature of the research question and objectives and to ensure coherence with the research philosophy and other elements of the research design.
- Possible research strategies include: experiment; survey; ethnography; Grounded Theory; Narrative Inquiry, archival research, documentary research; case study; and Action Research.
- The choice of research strategy or strategies will be related to use of an appropriate time horizon.
- Research ethics play a critical part in formulating a research design.
- Establishing the quality of research is also a critical part of formulating a research design, using appropriate criteria to judge and ensure the quality of the research.
- Practical considerations will also affect research design, including the role of the researcher.

Self-check questions

Answers to these questions are available at the end of the chapter.

- 5.1** You wish to study the reasons why car owners join manufacturer-sponsored owners' clubs. You choose to use a qualitative methodology and narrative inquiry research strategy involving unstructured 'discussions' with some members of these owners' clubs. You are asked by a small group of marketing managers to explain why your chosen research design is as valid as using a quantitative methodology and survey strategy that uses a questionnaire. What would be your answer?
- 5.2** You are working in an organisation that has branches throughout the country. The managing director is mindful of the fact that managers of the branches need to talk over common problems on a regular basis. That is why there has always been monthly meetings. During the global pandemic monthly meetings were conducted online using cloud-based videoconferencing. However, it is now unclear whether it is cost-effective to return to monthly face-to-face meetings. Some managers feel that their time travelling would be better spent pursuing their principal job objectives. Other managers see face-to-face meetings as more helpful, and offer greater insights.
- The managing director has asked you to carry out some research on the format of the monthly meetings. You have defined the research question you are seeking to answer as 'What are the managers' opinions of face-to-face versus online monthly meetings?'
 - Your principal research strategy will be a survey using a questionnaire to all managers who attend the monthly meetings. However, you are keen to triangulate your findings. How might you do this?
- 5.3** You have started conducting online interviews in a university with the non-academic employees (such as administrative and other support staff). The research objective is to establish the extent to which these employees feel a sense of 'belonging' to the university. You have negotiated access to your interviewees through the head of each of the appropriate departments. In each case you have been presented with a list of interviewees.
- It soon becomes apparent to you that you are getting a rather rosier picture than you expected. The interviewees are all very positive about their jobs, their managers and the university. This makes you suspicious. Are all the non-academic staff as positive as this? Are you being given only the employees who can be relied on to tell the 'good news'? Have they been 'got at' by their manager?
 - There is a great risk that your results will not be valid. What can you do?
- 5.4** You are about to embark on a year-long study of customer service training for sales assistants in two national supermarket companies. The purpose of the research is to compare the way in which the training develops and its effectiveness. What measures would you need to take in the research design stage to ensure that your results were valid?

Review and discussion questions

5.5 Agree with a friend to watch the same television documentary.

- a** To what extent is the nature of the documentary exploratory, descriptive, explanatory, evaluative or a combination of these?
- b** Does the documentary use quantitative, qualitative or mixed methods?
- c** What other observations can you make about the research strategy or strategies the documentary makers have used in their programme?

Do not forget to make notes regarding your reasons for your answers to each of these questions and to discuss these answers with your friend.

5.6 Use the search facilities of an online database to search for scholarly (peer-reviewed) articles that have used first a case study, second Action Research and third experiment research strategy in an area of interest to you. Download a copy of each article. What reasons do the articles' authors give for the choice of strategy?

5.7 Visit the Internet gateway to the European Union website (<http://europa.eu/>) and click on the link in your own language. Discuss with a friend how you might use the data available via links from this web page in archival research. In particular, you should concentrate on the research questions you might be able to answer using these data.



Progressing your research project

Deciding on your research design

- Review your research question, research aim and research objectives and, based on this decide the purpose of your research.
- Do your research question, aim, objectives, purpose and philosophy suggest using a mono method (qualitative or quantitative) multi-method (qualitative or quantitative) or mixed methods design? Make notes as you undertake this evaluation. Reflect on your options and decide which methodological approach is most appropriate in relation to your research question, aim, objectives, and purpose.
- Based on the decision(s) you have made so far, either (a) choose the research strategy that is suitable for your research, or (b), where you possibly have a choice, including using a combination of strategies, create a shortlist of research strategies that may be appropriate to conduct your research, together with the advantages and disadvantages of each.

- *If you have decided (a)*, search for studies in the literature that are based on the use of your chosen research strategy. Evaluate how the authors of these studies have used this research strategy. Compare this to your proposed use of this research strategy. Identify learning points from these studies for your proposed research. Reflect on your choice of this strategy: confirm and justify that it is an appropriate choice for your research, or re-appraise your choice of research strategy.
- *If you have decided (b)*, set this shortlist aside and search for studies in the literature that are similar to your own. Use these to note which strategies have been used. What explanations do the researchers give for their choice of strategy? Evaluate your shortlist against the notes from your search of studies in the literature. Use this evaluation to decide which strategy or combination of strategies would be most appropriate for your own research.
- Decide on the time frame to conduct your proposed research.
- Ask yourself, 'What practical constraints may affect my choice of proposed research design?' Use this question to review your decisions above and if necessary make changes. Repeat this step

- until you are satisfied that your proposed research design is practical.
- Use your draft research design to list (a) potential threats to research quality and (b) ethical issues in your design and make notes about how you propose to deal with each. Where necessary, make further changes to the decisions in the steps above.

- Reflect on your own role in your research and, if needed, make further changes to the steps considered above until you are satisfied with your proposed research design.
- You should now be ready to discuss your proposed research design with your project tutor.
- Use the questions in Box 1.4 to guide your reflective diary entry.

References

- 63 Up (2019) [Blu-ray] London: Network.
- Beaulieu, M., Breton, M. and Brousseau, A. (2018) Conceptualizing 20 years of engaged scholarship: A scoping review. *PLoS ONE*, Vol. 13, No. 2, p. e0193201. Available at <https://doi.org/10.371/journal.pone.0193201> [Accessed 12 June 2021].
- Bryant, A. and Charmaz, K. (2007) *The Sage Handbook of Grounded Theory*. London: Sage.
- Bryman, A. (2006) 'Integrating quantitative and qualitative research: How is it done?', *Qualitative Research*, Vol. 6, No. 1, pp. 97–113.
- Charmaz, K. (2011) 'Grounded theory methods in social justice research', in N.K. Denzin and Y.S. Lincoln (eds) *The Sage Handbook of Qualitative Research* (4th edn). London: Sage, pp. 359–80.
- Charmaz, K. (2014) *Constructing Grounded Theory* (2nd edn). London: Sage.
- Chase, S.E. (2011) 'Narrative inquiry: Still a field in the making', in N.K. Denzin and Y.S. Lincoln (eds) *The Sage Handbook of Qualitative Research* (4th edn). London: Sage, pp. 421–34.
- Coffey, A. and Atkinson, P. (1996) *Making Sense of Qualitative Data*. London: Sage.
- Coghlan, D. (2011) 'Action Research: Exploring perspectives on a philosophy of practical knowing', *The Academy of Management Annals*, Vol. 5, No. 1, pp. 53–87.
- Coghlan, D. (2019) *Doing Action Research in Your Own Organisation* (5th edn). London: Sage.
- Corbin, J. and Strauss, A. (2008) *Basics of Qualitative Research* (3rd edn). London: Sage.
- Corley, K.G. (2015) 'A commentary on "what grounded theory is." Engaging a phenomenon from the perspective of those living it', *Organizational Research Methods*, Vol. 18, No. 4, pp. 600–5.
- Creswell, J.W. and Plano Clark, V.L. (2017) *Designing and Conducting Mixed Methods Research* (3rd edn). Thousand Oaks, CA: Sage.
- Cunliffe, A.L. (2010) 'Retelling tales of the field: In search of organisational ethnography 20 years on', *Organizational Research Methods*, Vol. 13, No. 2, pp. 224–39.
- Delamont, S. (2007) 'Ethnography and participant observation', in C. Seale, G. Gobo, J.F. Gubrium and D. Silverman (eds) *Qualitative Research Practice*. London: Sage, pp. 205–17.
- Denzin, N.K. (2012) 'Triangulation 2.0', *Journal of Mixed Methods Research*, Vol. 6, No. 2, pp. 80–8.
- Denzin, N.K. and Lincoln, Y.S. (2018) *The Sage Handbook of Qualitative Research* (5th edn). London: Sage.
- Dougherty, D. (2015) 'Reflecting on the reflective conversation', *Organizational Research Methods*, Vol. 18, No. 4, pp. 606–11.
- Edelman (2021) *Edelman Trust Barometer Archive*. Available at <https://www.edelman.com/trust/archive> [Accessed 10 June 2021].

- Eden, C. and Huxham, C. (1996) 'Action research for management research', *British Journal of Management*, Vol. 7, No. 1, pp. 75–86.
- Eisenhardt, K.M. and Graebner, M.E. (2007) 'Theory building from cases: Opportunities and challenges', *Academy of Management Journal*, Vol. 50, No. 1, pp. 25–32.
- Flyvbjerg, B. (2011) 'Case study', in N.K. Denzin and Y.S. Lincoln (eds) *The Sage Handbook of Qualitative Research* (4th edn). London: Sage, pp. 301–16.
- Gabriel, Y. (2018) 'Stories and narratives', in C. Cassell, A. Cunliffe and G. Grundy (eds) *Qualitative Business and Management Research Methods: Methods and Challenges*. London: Sage, pp. 63–81.
- Geertz, C. (1988) *Works and Lives: The Anthropologist as Author*. Stanford, CA: Stanford University Press.
- Glaser, B. and Strauss, A. (1967) *The Discovery of Grounded Theory*. Chicago, IL: Aldine.
- Glaser, B.G. (1978) *Theoretical Sensitivity: Advances in the Methodology of Grounded Theory*. Mill Valley, CA: Sociology Press.
- Glaser, B.G. (1992) *Basics of Grounded Theory*. Mill Valley, CA: Sociology Press.
- Greene, J.C. (2007) *Mixed Methods in Social Inquiry*. San Francisco, CA: Jossey-Bass.
- Greenwood, D.J. and Levin, M. (2007) *Introduction to Action Research* (2nd edn). London: Sage.
- Guba, E.G. and Lincoln, Y.S. (1989) *Fourth Generation Evaluation*. Newbury Park, CA: Sage.
- Hakim, C. (2000) *Research Design: Successful Designs for Social and Economic Research* (2nd edn). London: Routledge.
- Huopalainen, A.S. and Satama, S.T. (2019) 'Mothers and researchers in the making: Negotiating "new" motherhood within the new "academia"', *Human Relations*, Vol. 72, No. 1, pp. 98–121.
- Kenealy, G.J.J. (2012) 'Grounded theory: A theory building approach', in G. Symon and C. Cassell (eds) *Qualitative Organisational Research Core Methods and Current Challenges*. London: Sage, pp. 408–25.
- Lee, B. (2012) 'Using documents in organizational research', in G. Symon and C. Cassell (eds) *Qualitative Organisational Research Core Methods and Current Challenges*. London: Sage, pp. 389–407.
- Lee, B. and Saunders M. (2017) *Doing Case Study Research for Business and Management Students*. London: Sage.
- Lewin, K. (1946) 'Action research and minority problems', *Journal of Social Issues*, Vol. 24, No. 1, pp. 34–46.
- Lincoln, Y.S. and Guba, E.G. (1985) *Naturalistic Inquiry*. Beverly Hills, CA: Sage.
- Lincoln, Y.S., Lynham S.A. and Guba, E.G. (2018) 'Paradigmatic controversies, contradictions and emerging confluences revisited', in N.K. Denzin and Y.S. Lincoln (eds) *The Sage Handbook of Qualitative Research* (5th edn). Los Angeles, CA: Sage, pp. 108–50.
- Locke, K. (2015) 'Pragmatic reflections on a conversation about grounded theory in management and organization studies', *Organizational Research Methods*, Vol. 18, No. 4, pp. 612–19.
- Maitlis, S. (2012) 'Narrative analysis', in G. Symon and C. Cassell (eds) *Qualitative Organisational Research Core Methods and Current Challenges*. London: Sage, pp. 492–511.
- Molina-Azorin, J.F. (2011) 'The use and added value of mixed methods in management research', *Journal of Mixed Methods Research*, Vol. 5, No. 1, pp. 7–24.
- Morse, J.M., Stern, P.N., Corbin, J., Bowers, B., Charmaz, K. and Clarke, A.E. (2009) *Developing Grounded Theory: The Second Generation*. Walnut Creek, CA: Left Coast Press.
- Musson, G. (2004) 'Life histories', in C. Cassell and G. Symon (eds) *Essential Guide to Qualitative Methods in Organizational Research*. London: Sage, pp. 34–46.

- Naipaul, V.S. (1989) *A Turn in the South*. London: Penguin.
- Nastasi, B.K., Hitchcock, J.H. and Brown, L.M. (2010) 'An inclusive framework for conceptualising mixed methods typologies', in A. Tashakkori and C. Teddlie (eds) *The Sage Handbook of Mixed Methods in Social and Behavioural Research* (2nd edn). Thousand Oaks, CA: Sage, pp. 305–338.
- Nielson, B.B., Welch, C., Chidlow, A., Miller, S.R., Gardner, E., Karafyllia, M. and Pegoraro, D. (2020) 'Fifty years of methodological trends in JIBs: Why future IB research needs more triangulation', *Journal of International Business*. Vol. 51, No. 9, pp. 1478–99.
- Prior, L. (2007) 'Documents', in C. Seale, G. Gobo, J.F. Gubrium and D. Silverman (eds) *Qualitative Research Practice*. London: Sage, pp. 345–60.
- Raimond, P. (1993) *Management Projects*. London: Chapman & Hall.
- Reason, P. (2006) 'Choice and quality in action research practice', *Journal of Management Inquiry*, Vol. 15, No. 2, pp. 187–202.
- Ridder, H-G., Hoon, C. and McCandless Baluch, A. (2014) 'Entering a dialogue: Positioning case study findings towards theory', *British Journal of Management*, Vol. 25, No. 2, pp. 373–87.
- Saunders, M.N.K., Lewis, P. and Thornhill, A. (2019) *Research Methods for Business Students* (8th edition) Harlow: Pearson.
- Schein, E. (1999) *Process Consultation Revisited: Building the Helping Relationship*. Reading, MA: Addison-Wesley.
- Shani, A.B. and Pasmore, W.A. (1985) 'Organization inquiry: Towards a new model of the action research process', in D.D. Warrick (ed.) *Contemporary Organization Development*. Glenview, IL: Scott Foresman, pp. 438–48.
- Stake, R.E. (2005) 'Qualitative case studies', in N.K. Denzin and Y.S. Lincoln (eds) *The Sage Handbook of Qualitative Research* (3rd edn). London: Sage, pp. 443–65.
- Strauss, A. and Corbin, J. (1998) *Basics of Qualitative Research* (2nd edn). London: Sage.
- Suddaby, R. (2006) 'From the editors: What grounded theory is not', *Academy of Management Journal*, Vol. 49, No. 4, pp. 633–42.
- Sun, J.J., Bellezza, S., Paharia, N. (2021) 'Buy less, buy luxury: understanding and overcoming product durability neglect for sustainable consumption' *Journal of Marketing*, Vol. 85, No. 3., pp. 28–43.
- Tashakkori, A. and Teddlie, C. (eds) (2010) *The Sage Handbook of Mixed Methods in Social and Behavioural Research* (2nd edn). Thousand Oaks, CA: Sage.
- Teddlie, C. and Tashakkori, A. (2009) *Foundations of Mixed Methods Research: Integrating Quantitative and Qualitative Approaches in the Social and Behavioral Sciences*. Thousand Oaks, CA: Sage.
- Teddlie, C. and Tashakkori, A. (2011) 'Mixed methods research: Contemporary issues in an emerging field', in N.K. Denzin and Y.S. Lincoln (eds) *The Sage Handbook of Qualitative Research* (4th edn). London: Sage, pp. 285–99.
- Tedlock, B. (2005) 'The observation of participation and the emergence of public ethnography', in N.K. Denzin and Y.S. Lincoln (eds) *The Sage Handbook of Qualitative Research* (3rd edn). London: Sage, pp. 467–482.
- Tietze, S. (2012) 'Researching your own organization', in G. Symon and C. Cassell (eds) *Qualitative Organisational Research Core Methods and Current Challenges*. London: Sage, pp. 53–71.
- Van de Ven, A.H. (2007). *Engaged Scholarship: A Guide for Organizational and Social Research*. Oxford: Oxford University.
- Walsh, I., Holton, J.A., Bailyn, L., Fernandez, Levina, N. and Glaser, B. (2015a) 'What grounded theory is . . . a critically reflective conversation among scholars', *Organizational Research Methods*, Vol. 18, No. 4, pp. 581–99.

- Walsh, I., Holton, J.A., Bailyn, L., Fernandez, Levina, N. and Glaser, B. (2015b) 'Rejoinder: Moving the management field forward', *Organizational Research Methods*, Vol. 18, No. 4, pp. 620–8.
- Wasserstein, R.L. and Lazar, N.A. (2016) 'The ASA's statement on p-values: context, process and purpose', *The American Statistician*, Vol. 70, No. 2, pp. 129–33.
- Whyte, W.F. (1993) *Street Corner Society: The Social Structure of an Italian Slum* (4th edn). Chicago, IL: University of Chicago Press.
- Yin, R.K. (2018) *Case Study Research and Applications: Design and Methods* (6th edn). London: Sage.

Further reading

- Cassell, C., Cunliffe, A. and Grundy, G. (eds) (2018) *Qualitative Business and Management Research Methods*. London: Sage. Both volumes of this book are useful for several research strategies including Action Research, autoethnography, archival research, case study, documentary research, Grounded Theory, and Narrative Inquiry (stories and narratives).
- Charmaz, K. (2014) *Constructing Grounded Theory* (2nd edn). London: Sage. Useful for Grounded Theory strategy.
- Coghlan, D. (2019) *Doing Action Research in Your Own Organisation* (5th edn). London: Sage. Useful for Action Research strategy.
- Cunliffe, A.L. (2010) 'Retelling tales of the field: In search of organisational ethnography 20 years on', *Organizational Research Methods*, Vol. 13, No. 2, pp. 224–39. Useful for organisational ethnography.
- deVaus, D.A. (2014) *Surveys in Social Research* (6th edn). Abingdon: Routledge. Useful for survey strategy.
- Hakim, C. (2000) *Research Design: Successful Designs for Social and Economic Research* (2nd edn). London: Routledge. Chapter 9 is useful for experiment strategy.
- Yin, R.K. (2018) *Case Study Research and Applications: Design and Methods* (6th edn). London: Sage. Useful for case study strategy.

Theoretical Framework & Hypothesis Development

Sekaran & Bougie (2016; 71-94)

Theoretical framework and hypothesis development

LEARNING OBJECTIVES

After completing Chapter 5, you should be able to:

1. Discuss the need for a theoretical framework in deductive research.
2. Describe four main types of variables and identify and label variables associated with any given situation.
3. Develop a theoretical framework that includes all the relevant components.
4. Develop a set of hypotheses to be tested.
5. Demonstrate awareness of the role of the manager in the development of a theoretical framework.

INTRODUCTION

After a critical review of the literature you may be ready to develop a theoretical framework. A theoretical framework is the foundation of hypothetico-*deductive* research as it is the basis of the hypotheses that you will develop. Indeed, the development of a theoretical framework is crucial in *deductive, theory-testing, causal* research (but *not* in exploratory or descriptive research where one does not develop a theoretical framework to develop and test hypotheses).

This chapter highlights the importance of theory development in deductive research and explains how theory is developed. The chapter starts with a definition of a theoretical framework followed by a discussion on the need for a theoretical framework. It explains that a theoretical framework involves the identification of a network of relationships among variables considered important to the problem. Different types of variables on hypotheses development are reviewed at the end of this chapter.

As you proceed through this chapter, in various places you are instructed to work through certain exercises. Doing them at that time, before reading further, will help you in becoming adept at formulating theoretical frameworks in a logical manner without getting confused.

THE NEED FOR A THEORETICAL FRAMEWORK

A **theoretical framework** represents your beliefs on *how* certain phenomena (or variables or concepts) are related to each other (a model) and an explanation of *why* you believe that these variables are associated with each other (a theory). Both the model and the theory flow logically from the documentation of previous research in the problem area. Integrating your logical beliefs with published research, taking into consideration the boundaries and constraints governing the situation, is pivotal in developing a scientific basis for investigating the research problem.

The process of building a theoretical framework includes:

1. Introducing definitions of the concepts or variables in your model.
2. Developing a conceptual model that provides a descriptive representation of your theory.
3. Coming up with a theory that provides an explanation for relationships between the variables in your model.

From the theoretical framework, then, testable hypotheses can be developed to examine whether your theory is valid or not. The hypothesized relationships can thereafter be tested through appropriate statistical analyses. Hence, the entire *deductive* research project rests on the basis of the theoretical framework. Even if testable hypotheses are not necessarily generated (as in some applied research projects), developing a good theoretical framework is central to examining the problem under investigation.

Since a theoretical framework involves the identification of the network of relationships among the variables considered important to the study of any given problem situation, it is essential to understand what a variable means and what the different types of variables are.

VARIABLES

A **variable** is anything that can take on differing or varying values. The values can differ at various times for the same object or person, or at the same time for different objects or persons. Examples of variables are production units, absenteeism, and motivation.

EXAMPLE

Production units: One worker in the manufacturing department may produce one widget per minute, a second might produce two per minute, a third might produce five per minute. It is also possible that the same member might produce one widget the first minute and five the next minute. In both cases, the number of widgets produced has taken on different values, and is therefore a variable.

Absenteeism: Today, three members in the sales department may be absent; tomorrow, six members may not show up for work; the day after, there may be no one absent. The value can thus theoretically range

from “zero” to “all” being absent, on the absenteeism variable.

Motivation: The levels of motivation of members to learn in the class or in a work team might take on varying values ranging from “very low” to “very high.” An individual’s motivation to learn from different classes or in different work teams might also take on differing values. Now, how one *measures* the level of motivation is an entirely different matter. The factor called motivation has to be reduced from its level of abstraction and operationalized in such a way that it becomes measurable. We will discuss this in Chapter 11.

Four main types of variables are discussed in this chapter¹:

1. The dependent variable (also known as the criterion variable).
2. The independent variable (also known as the predictor variable).
3. The moderating variable.
4. The mediating variable.

Each of these variables can be discrete (e.g., male/female) or continuous (e.g., the age of an individual). Associated scale levels of variables are discussed in Chapter 12.

Dependent variable

The **dependent variable** is the variable of primary interest to the researcher. The researcher's goal is to understand and describe the dependent variable, or to explain its variability, or predict it. In other words, it is the main variable that lends itself for investigation as a viable factor. Through the analysis of the dependent variable (i.e., finding what variables influence it), it is possible to find answers or solutions to the problem. For this purpose, the researcher will be interested in quantifying and measuring the dependent variable, as well as the other variables that influence this variable.

EXAMPLE

A manager is concerned that the sales of a new product, introduced after test marketing it, do not meet with his expectations. The dependent variable here is "sales." Since the sales of the product can vary – they can be low, medium, or high – it is a variable; since sales is the main focus of interest to the manager, it is the dependent variable.

A basic researcher is interested in investigating the debt-to-equity ratio of manufacturing companies in southern Germany. Here, the dependent variable is the ratio of debt to equity.

A vice president is concerned that the employees are not loyal to the organization and, in fact, seem to

switch their loyalty to other institutions. The dependent variable in this case is "organizational loyalty." Here again, there is variance found in the levels of organizational loyalty of employees. The vice president might want to know what accounts for the variance in the loyalty of organizational members with a view to controlling it. If he finds that increased pay levels would ensure their loyalty and retention, he can then offer inducement to employees by way of pay rises, which will help control the variability in organizational loyalty and keep them in the organization.

It is possible to have more than one dependent variable in a study. For example, there is always a tussle between quality and volume of output, low-cost production and customer satisfaction, and so on. In such cases, the manager is interested to know the factors that influence all the dependent variables of interest and how some of them might differ in regard to different dependent variables. These investigations may call for multivariate statistical analyses.

Now do Exercise 5.1 and Exercise 5.2.

¹Extraneous variables that confound cause-and-effect relationships are discussed in Chapter 10 on Experimental Designs.

EXERCISE 5.1

Research in behavioral finance has shown that overconfidence can cause investors to underreact to new information.

What is the dependent variable in this case?

EXERCISE 5.2

A marketing manager believes that limiting the availability of a product increases product desirability.

What is the dependent variable here?

Independent variable

It is generally conjectured that an **independent variable** is one that influences the dependent variable in either a positive or negative way. That is, when the independent variable is present, the dependent variable is also present, and with each unit of increase in the independent variable, there is an increase or decrease in the dependent variable. In other words, the variance in the dependent variable is accounted for by the independent variable. To establish that a change in the independent variable *causes* a change in the dependent variable, *all four* of the following conditions should be met:

1. The independent and the dependent variable should covary: in other words, a change in the dependent variable should be associated with a change in the independent variable.
2. The independent variable (the presumed causal factor) should precede the dependent variable. In other words, there must be a time sequence in which the two occur: the cause must occur before the effect.
3. No other factor should be a possible cause of the change in the dependent variable. Hence, the researcher should *control for* the effects of other variables.
4. A logical explanation (a theory) is needed and it must explain why the independent variable affects the dependent variable.

Because of the time sequence condition, experimental designs, described in Chapter 10, are often used to establish causal relationships.

EXAMPLE

Research studies indicate that successful new product development has an influence on the stock market price of the company. That is, the more successful the new product turns out to be, the higher will be the stock market price of that firm. Therefore, the “success of the new product” is the

independent variable, and “stock market price” the *dependent variable*. The degree of perceived success of the new product developed will explain the variance in the stock market price of the company. This relationship and the labeling of the variables are illustrated in Figure 5.1.

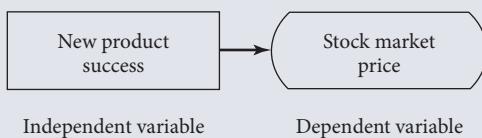
**FIGURE 5.1**

Diagram of the relationship between the independent variable (new product success) and the dependent variable (stock market price)

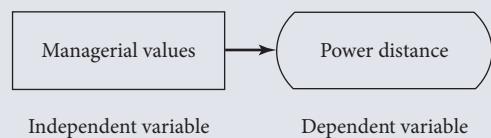
**FIGURE 5.2**

Diagram of the relationship between the independent variable (managerial values) and the dependent variable (power distance)

Cross-cultural research indicates that managerial values govern the power distance between superiors and subordinates. Here, power distance (i.e., egalitarian interactions between the boss and the employee, versus the high-power superior in limited interaction

with the low-power subordinate) is the subject of interest and hence the dependent variable. Managerial values that explain the variance in power distance comprise the independent variable. This relationship is illustrated in Figure 5.2.

Now do Exercise 5.3 and Exercise 5.4. List the variables in these two exercises individually, and label them as dependent or independent, explaining why they are so labeled. Create diagrams to illustrate the relationships.

EXERCISE 5.3

An investor believes that more information increases the accuracy of his forecasts.

EXERCISE 5.4

A marketing manager believes that selecting physically attractive spokespersons and models to endorse their products increases the persuasiveness of a message.

Moderating variable

Visit the companion website at www.wiley.com/college/sekaran for Author Video: The moderating variable.

The **moderating variable** is one that has a strong *contingent* effect on the independent variable-dependent variable relationship. That is, the presence of a third variable (the moderating variable) modifies the original relationship between the independent and the dependent variables. This becomes clear through the following examples.

EXAMPLE

It has been found that there is a relationship between the availability of reference manuals that manufacturing employees have access to and the product rejects. That is, when workers follow the procedures laid down in the manual, they are able to manufacture products that are flawless. This relationship is illustrated in Figure 5.3(a). Although this relationship can be said to hold true generally for all workers, it is nevertheless contingent on the inclination or urge of the employees

to look in the manual every time a new procedure is to be adopted. In other words, only those who have the interest and urge to refer to the manual every time a new process is adopted will produce flawless products. Others who do not consult the manual will not benefit and will continue to produce defective products. This influence of the attributes of the worker on the relationship between the independent and the dependent variables can be illustrated as shown in Figure 5.3(b).

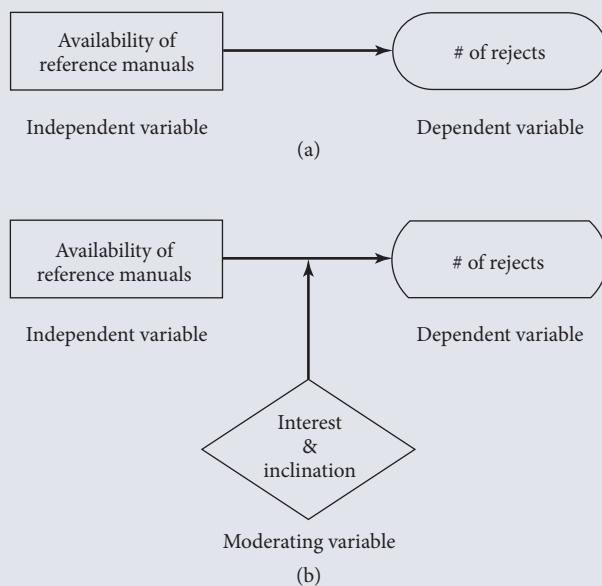


FIGURE 5.3

(a) Diagram of the relationship between the independent variable (availability of reference manuals) and the dependent variable (rejects); (b) diagram of the relationship between the independent variable (availability of reference materials) and the dependent variable (rejects) as moderated by the moderating variable (interest and inclination)

As in the above case, whenever the relationship between the independent variable and the dependent variable becomes contingent or dependent on another variable, we say that the third variable has a moderating effect on the independent variable–dependent variable relationship. The variable that moderates the relationship is known as the moderating variable.

EXAMPLE

Let us take another example of a moderating variable. A prevalent theory is that the diversity of the workforce (comprising people of different ethnic origins, races, and nationalities) contributes more to organizational effectiveness because each group brings its own special expertise and skills to the workplace. This synergy can be exploited, however, only if managers know how to harness the special talents of the diverse work group; otherwise they will remain untapped. In the above scenario, organizational effectiveness is the dependent variable, which is positively influenced by workforce diversity – the independent variable. However, to harness the potential, managers must know how to encourage and coordinate the talents of the various groups to make things work. If not, the synergy will not be tapped. In other words, the effective utilization of different talents, perspectives, and eclectic problem-solving capabilities for enhanced organizational effectiveness is contingent on the skill of the managers in

acting as catalysts. This managerial expertise then becomes the moderating variable. These relationships can be depicted as in Figure 5.4.

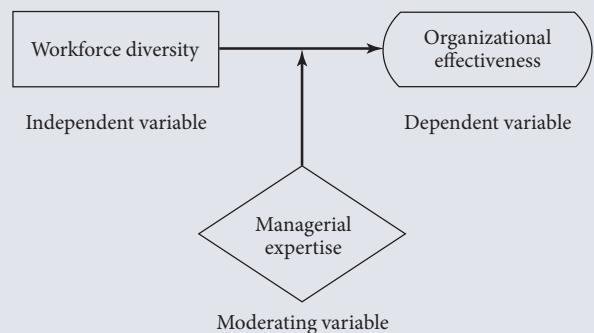


FIGURE 5.4

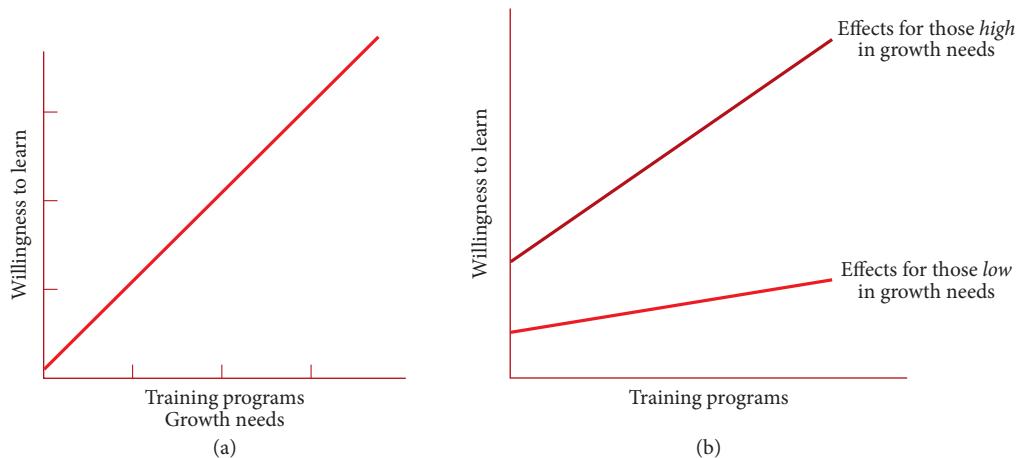
Diagram of the relationship among the three variables: workforce diversity, organizational effectiveness, and managerial expertise

The distinction between an independent variable and a moderating variable

At times, confusion is likely to arise as to when a variable is to be treated as an independent variable and when it becomes a moderating variable. For instance, there may be two situations as follows:

1. A research study indicates that the better the quality of the training programs in an organization and the greater the growth needs of the employees (i.e., where the need to develop and grow on the job is strong), the greater is their willingness to learn new ways of doing things.
2. Another research study indicates that the willingness of the employees to learn new ways of doing things is *not* influenced by the quality of the training programs offered by the organizations to *all* people without any distinction. Only those with high growth needs seem to have the yearning to learn to do new things through specialized training.

In the above two situations, we have the same three variables. In the first case, the training programs and growth need strength are the independent variables that influence employees' willingness to learn, this latter being the dependent variable. In the second case, however, the quality of the training program is the independent variable, and while the dependent variable remains the same, growth need strength becomes a moderating variable. In other words, only those with high growth needs show a greater willingness and adaptability to learn to do new things when the quality of the training program is improved. Thus, the relationship between the independent and dependent variables has now become contingent on the existence of a moderator.

**FIGURE 5.5**

(a) Illustration of the influence of independent variables on the dependent variable when no moderating variable operates in the situation; (b) illustration of the influence of independent variables on the dependent variable when a moderating variable is operating in the situation

The above illustration makes it clear that even though the variables used are the same, the decision as to whether to label them dependent, independent, or moderating depends on how they affect one another. The differences between the effects of the independent and the moderating variables may be visually depicted as in Figures 5.5(a) and 5.5(b). Note the steep incline of the top line and the relative flatness of the bottom line in Figure 5.5(b).

Now do Exercise 5.5 and Exercise 5.6. List and label the variables in these two exercises and explain and illustrate by means of diagrams the relationships among the variables.

EXERCISE 5.5

A manager finds that off-the-job classroom training has a great impact on the productivity of the employees in her department. However, she also observes that employees over 60 years of age do not seem to derive much benefit and do not improve with such training.

EXERCISE 5.6

A manager of an insurance company finds that “fear appeals” in commercials are positively associated with consumers’ behavioral intentions to insure their house. This effect is particularly strong for people with a high inherent level of anxiety.

Mediating variable

Visit the companion website at www.wiley.com/college/sekaran for Author Video: The mediating variable.

A **mediating variable** (or **intervening variable**) is one that surfaces between the time the independent variables start operating to influence the dependent variable and the time their impact is felt on it. There is thus a temporal quality or time dimension to the mediating variable. In other words, bringing a mediating variable into play helps you to model a *process*. The mediating variable surfaces as a function of the independent variable(s) operating in any situation, and helps to conceptualize and explain the influence of the independent variable(s) on the dependent variable. The following example illustrates this point.

EXAMPLE

In the previous example, where the independent variable (workforce diversity) influences the dependent variable (organizational effectiveness), the mediating variable that surfaces as a function of the diversity in the workforce is “creative synergy.” This creative synergy results from a multiethnic, multiracial, and multinational (i.e., diverse) workforce interacting and bringing together their multifaceted expertise in problem solving. This helps us to understand how organi-

zational effectiveness can result from having diversity in the workforce. Note that creative synergy, the mediating variable, surfaces at time t_2 , as a function of workforce diversity, which was in place at time t_1 , to bring about organizational effectiveness in time t_3 . The mediating variable of creative synergy helps us to conceptualize and understand how workforce diversity brings about organizational effectiveness. The dynamics of these relationships are illustrated in Figure 5.6.

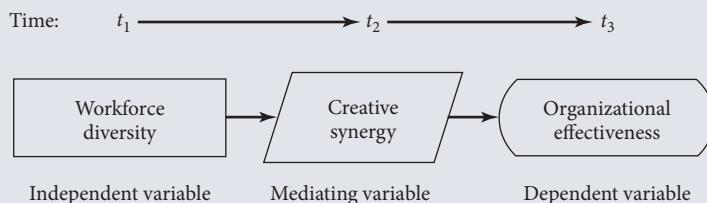


FIGURE 5.6

Diagram of the relationship among the independent, mediating, and dependent variables

It would be interesting to see how the inclusion of the moderating variable, “managerial expertise” in the foregoing example, would change the model or affect the relationships. The new set of relationships that would emerge in the presence of the moderator is depicted in Figure 5.7. As can be seen, managerial expertise moderates the relationship between workforce diversity and creative synergy. In other words, creative synergy will not

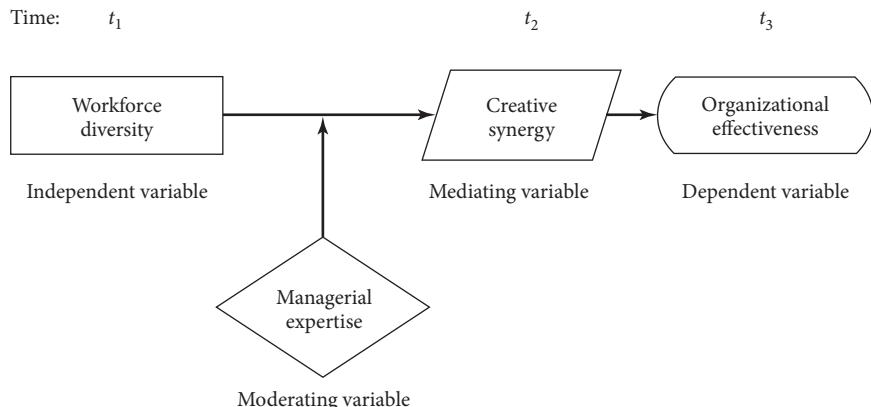
**FIGURE 5.7**

Diagram of the relationship among the independent, mediating, moderating, and dependent variables

result from the multifaceted problem-solving skills of the diverse workforce unless the manager is capable of harnessing that synergy by creatively coordinating the different skills. If the manager lacks the expertise to perform this role, then no matter how many different problem-solving skills the diverse workforce might have, synergy will just not surface. Instead of functioning effectively, the organization might just remain static, or even deteriorate.

It is now easy to see what the differences are among an independent variable, a mediating variable, and a moderating variable. The **independent variable** helps to *explain* the variance in the dependent variable; the **mediating variable** *surfaces at time t_2* as a function of the independent variable, which also helps us to conceptualize the relationship between the independent and dependent variables; and the **moderating variable** has a *contingent effect* on the relationship between two variables. To put it differently, while the independent variable explains the variance in the dependent variable, the mediating variable does not add to the variance already explained by the independent variable, whereas the moderating variable has an interaction effect with the independent variable in explaining the variance. That is, unless the moderating variable is present, the theorized relationship between the other two variables considered will not hold.

Whether a variable is an independent variable, a dependent variable, a mediating variable, or a moderating variable should be determined by a careful reading of the dynamics operating in any given situation. For instance, a variable such as motivation to work could be a dependent variable, an independent variable, a mediating variable, or a moderating variable, depending on the theoretical model that is being advanced.

Now do Exercise 5.7, Exercise 5.8, and Exercise 5.9.

EXERCISE 5.7

Make up three different situations in which motivation to work would be an independent variable, a mediating variable, and a moderating variable.

EXERCISE 5.8

Failure to follow accounting principles causes immense confusion, which in turn creates a number of problems for the organization. Those with vast experience in bookkeeping, however, are able to avert the problems by taking timely corrective action. List and label the variables in this situation, explain the relationships among the variables, and illustrate these by means of diagrams.

EXERCISE 5.9

A store manager observes that the morale of employees in her supermarket is low. She thinks that if their working conditions are improved, pay scales raised, and the vacation benefits made more attractive, the morale will be boosted. She doubts, however, if an increase in pay scales would raise the morale of all employees. Her conjecture is that those who have supplemental incomes will just not be “turned on” by higher pay, and only those without side incomes will be happy with increased pay, with a resultant boost in morale. List and label the variables in this situation. Explain the relationships among the variables and illustrate them by means of diagrams. What might be the problem statement or problem definition for the situation?

HOW THEORY IS GENERATED

Having examined the different kinds of variables that can operate in a situation and how the relationships among these can be established, it is now possible to see how we can develop the theoretical framework for our research.

The theoretical framework is the foundation on which the entire deductive research project is based. It is a logically developed, described, and elaborated network of associations among the variables deemed relevant to the problem situation and identified through such processes as interviews, observations, and literature review. Experience and intuition also guide the development of the theoretical framework.

It becomes evident at this stage that, to arrive at good solutions to the problem, one should first correctly identify the problem, and then the variables that contribute to it. The importance of doing a thorough literature review and conducting exploratory and inductive research now becomes clear. After identifying the appropriate variables, the next step is to elaborate the network of associations among the variables, so that relevant hypotheses can be developed and subsequently tested. Based on the results of hypothesis testing (which indicate whether or not the hypotheses have been supported), the extent to which the problem can be solved becomes evident. The theoretical framework is thus an important step in the research process.

The relationship between the literature review and the theoretical framework is that the former provides a solid foundation for developing the latter. That is, the literature review identifies the variables that might be important, as determined by previous research findings. This, in addition to other logical connections that can be conceptualized, forms the basis for the theoretical model. The theoretical framework represents and elaborates the relationships among the variables, explains the theory underlying these relations, and describes the nature and direction of the relationships. Just as the literature review sets the stage for a good theoretical framework, this in turn provides the logical base for developing testable hypotheses.

The components of the theoretical framework

A good theoretical framework identifies and defines the important variables in the situation that are relevant to the problem and subsequently describes and explains the interconnections among these variables. The relationships among the independent variables, the dependent variable(s), and, if applicable, the moderating and mediating variables are elaborated. Should there be any moderating variable(s), it is important to explain how and what specific relationships they moderate. An explanation of why they operate as moderators should also be offered. If there are any mediating variables, a discussion on how or why they are treated as mediating variables is necessary. Any interrelationships among the independent variables themselves, or among the dependent variables themselves (in case there are two or more dependent variables), should also be clearly spelled out and adequately explained. Note that a good theoretical framework is not necessarily a complex framework.

Earlier in this chapter, we have already explained that there are three basic features that should be incorporated in any theoretical framework:

1. The variables considered relevant to the study should be clearly defined.
2. A conceptual model that describes the relationships between the variables in the model should be given.
3. There should be a clear explanation of why we expect these relationships to exist.

It is not always easy to come up with generally agreed-upon *definitions* of the relevant variables. More often than not, there are many definitions available in the literature (for instance, there are literally dozens of definitions of “brand image,” “customer satisfaction,” and “service quality” available in the marketing literature). Still, well-chosen guiding definitions of concepts are needed, because they will help you to provide an explanation for the relationships between the variables in your model. What’s more, they will also serve as a basis for the operationalization or measurement of your concepts in the data collection stage of the research process. Hence, you will have to choose a useful definition from the literature (do not use dictionary definitions, they are usually too general). It is also important that you explain why you have chosen a particular definition as your guiding definition.

A *conceptual model* helps you to structure your discussion of the literature. A conceptual model describes your ideas about how the concepts (variables) in your model are related to each other. A schematic diagram of the conceptual model helps the reader to visualize the theorized relationships between the variables in your model and thus to obtain a quick idea about how you think that the management problem can be solved. Hence, conceptual models are often expressed in this form. However, relationships between variables can also be adequately expressed in words. Both a schematic diagram of the conceptual model and a description of the relationships between the variables in words should be given, so that the reader can see and easily comprehend the theorized relationships. This facilitates and stimulates discussion about the relationships between the variables in your model. It is therefore important that your model is based on a sound theory.

A theory or a clear explanation for the relationships in your model is the last component of the theoretical framework. A theory attempts to explain relationships between the variables in your model: an explanation should be provided for all the important relationships that are theorized to exist among the variables. If the nature and direction of the relationships can be theorized on the basis of the findings of previous research and/or your own ideas on the subject, then there should also be an indication as to whether the relationships should be positive or negative and linear or nonlinear. From the theoretical framework, then, testable hypotheses can be developed to examine whether the theory formulated is valid or not.

Note that you do not necessarily have to “invent” a new theory every time you are undertaking a research project. In an applied research context you apply existing theories to a specific context. This means that arguments can be drawn from previous research. However, in a basic research context you will make some contribution to existing theories and models. In such a case, it is not (always) possible to use existing theories or explanations for relationships between variables. As a result, you will have to rely on your own insights and ideas.

Now do Exercise 5.10.

EXERCISE 5.10

Avatars are virtual characters that can be used as representatives of a company that is using the Internet as a distribution channel. For instance, avatars can be used as shopping assistants, website guides, or as identification figures. A manager of an online company believes that avatar-mediated communication will have a positive effect on satisfaction with her company and on purchase intentions of consumers, because avatars enhance the value of information provided on the website and increase the pleasure of the shopping experience. She also believes that the positive effect of the perceived information value on satisfaction with the company and purchase intentions is stronger when customers are highly involved. Develop a theoretical framework for this situation after stating what the problem definition of the researcher would be in this case.

HYPOTHESIS DEVELOPMENT

Once we have identified the important variables in a situation and established the relationships among them through logical reasoning in the theoretical framework, we are in a position to test whether the relationships that have been theorized do, in fact, hold true. By testing these relationships scientifically through appropriate statistical analyses, or through negative case analysis in qualitative research (described later in the chapter), we are able to obtain reliable information on what kinds of relationships exist among the variables operating in the problem situation. The results of these tests offer us some clues as to what could be changed in the situation to solve the problem. Formulating such testable statements is called *hypothesis development*.

Definition of a hypothesis

A **hypothesis** can be defined as a tentative, yet testable, statement, which predicts what you expect to find in your empirical data. Hypotheses are derived from the theory on which your conceptual model is based and are often relational in nature. Along these lines, hypotheses can be defined as logically conjectured relationships between two or more variables expressed in the form of testable statements. By testing the hypotheses and confirming the conjectured relationships, it is expected that solutions can be found to correct the problem encountered.

EXAMPLE

Several testable statements or hypotheses can be drawn from the example depicted in Figure 5.4. One of them might be:

Workforce diversity has a positive effect on organizational effectiveness.

The above is a testable statement. By measuring the extent of workforce diversity and organizational effectiveness, we can statistically examine the relation-

ship between these two variables to see if there is a significant (positive) correlation between the two. If we do find this to be the case, then the hypothesis is substantiated. If a significant correlation is not found, then the hypothesis has not been substantiated. By convention in the social sciences, to call a relationship “statistically significant,” we should be confident that 95 times out of 100 the observed relationship will hold true. There should be only a 5% chance that the relationship will not be detected.

Statement of hypotheses: formats

If–then statements

As already stated, a hypothesis can be defined as a testable statement of the relationship among variables. A hypothesis can also test whether there are differences between two groups (or among several groups) with respect to any variable or variables. To examine whether or not the conjectured relationships or differences exist, these hypotheses can be set either as propositions or in the form of *if–then statements*. The two formats can be seen in the following two examples.

Young women will be more likely to express dissatisfaction with their body weight, when they are more frequently exposed to images of thin models in advertisements.

If young women are more frequently exposed to images of thin models in advertisements, then they will be more likely to express dissatisfaction with their body weight.

Directional and nondirectional hypotheses

If, in stating the relationship between two variables or comparing two groups, terms such as *positive*, *negative*, *more than*, *less than*, and the like are used, then these are **directional hypotheses** because the direction of the relationship between the variables (positive/negative) is indicated, as in the first example below, or the nature of the difference between two groups on a variable (more than/less than) is postulated, as in the second example.

The greater the stress experienced in the job, the lower the job satisfaction of employees.

Women are more motivated than men.

On the other hand, **nondirectional hypotheses** are those that do postulate a relationship or difference, but offer no indication of the direction of these relationships or differences. In other words, though it may be conjectured that there is a significant relationship between two variables, we may not be able to say whether the relationship is positive or negative, as in the first example below. Likewise, even if we can conjecture that there will be differences between two groups on a particular variable, we may not be able to say which group will be more and which less on that variable, as in the second example.

There is a relation between arousal-seeking tendency and consumer preferences for complex product designs.

There is a difference between the work ethic values of American and Asian employees.

Nondirectional hypotheses are formulated either because the relationships or differences have never been explored, and hence there is no basis for indicating the direction, or because there have been conflicting findings in previous research studies on the variables. In some studies a positive relationship might have been found, while in others a negative relationship might have been traced. Hence, the current researcher might only be able to hypothesize that there is a significant relationship, but the direction may not be clear. In such cases, the hypotheses can be stated nondirectionally. Note that in the first example there is no clue as to whether

arousal-seeking tendency and preferences for complex product designs are positively or negatively correlated, and in the second example we do not know whether the work ethic values are stronger in Americans or in Asians. However, it would have been possible to state that arousal-seeking tendency and preferences for complex product designs are positively correlated, since previous research has indicated such a relationship. Whenever the direction of the relationship is known, it is better to develop directional hypotheses for reasons that will become clear in our discussions in a later chapter.

Null and alternate hypotheses

The hypothetico-deductive method requires that hypotheses are falsifiable: they must be written in such a way that other researchers can show them to be false. For this reason, hypotheses are sometimes accompanied by null hypotheses. A **null hypothesis** (H_0) is a hypothesis set up to be rejected in order to support an alternate hypothesis, labeled H_A . When used, the null hypothesis is presumed true until statistical evidence, in the form of a hypothesis test, indicates otherwise. For instance, the null hypothesis may state that advertising does not affect sales, or that women and men buy equal amounts of shoes. In more general terms, the null hypothesis may state that the correlation between two variables is equal to zero or that the difference in the means of two groups in the population is equal to zero (or some other *definite* number). Typically, the null statement is expressed in terms of there being no (*significant*) relationship between two variables or no (*significant*) difference between two groups. The **alternate hypothesis**, which is the opposite of the null, is a statement expressing a relationship between two variables or indicating differences between groups.

To explain further, in setting up the null hypothesis, we are stating that there is no difference between what we might find in the population characteristics (i.e., the total group we are interested in knowing something about) and the sample we are studying (i.e., a limited number representative of the total population or group that we have chosen to study). Since we do not know the true state of affairs in the population, all we can do is to draw inferences based on what we find in our sample. What we imply through the null hypothesis is that any differences found between two sample groups or any relationships found between two variables based on our sample are simply due to random sampling fluctuations and not due to any “true” differences between the two population groups (say, men and women), or relationships between two variables (say, sales and profits). The null hypothesis is thus formulated so that it can be tested for possible rejection. If we reject the null hypothesis, then all permissible alternate hypotheses relating to the particular relationship tested could be supported. It is the theory that allows us to have faith in the alternate hypothesis that is generated in the particular research investigation. This is one more reason why the theoretical framework should be grounded on sound, defendable logic to start with. Otherwise, other researchers are likely to refute and postulate other defensible explanations through different alternate hypotheses.

The *null hypothesis* in respect of group differences stated in the example “*Women are more motivated than men*” would be:

$$H_0: \mu_M = \mu_W$$

or

$$H_0: \mu_M - \mu_W = 0$$

where H_0 represents the null hypothesis, μ_M is the mean motivational level of the men, and μ_W is the mean motivational level of the women.

The *alternate* for the above example would statistically be set as follows:

$$H_A: \mu_M < \mu_W$$

which is the same as

$$H_A: \mu_W > \mu_M$$

where H_A represents the alternate hypothesis and μ_M and μ_W are the mean motivation levels of men and women, respectively.

For the nondirectional hypothesis of mean group differences in work ethic values in the example “*There is a difference between the work ethic values of American and Asian employees*,” the null hypothesis would be:

$$H_0: \mu_{AM} = \mu_{AS}$$

or

$$H_0: \mu_{AM} - \mu_{AS} = 0$$

where H_0 represents the null hypothesis, μ_{AM} is the mean work ethic value of Americans and μ_{AS} is the mean work ethic value of Asians.

The alternate hypothesis for the above example would statistically be set as:

$$H_A: \mu_{AM} \neq \mu_{AS}$$

where H_A represents the alternate hypothesis and μ_{AM} and μ_{AS} are the mean work ethic values of Americans and Asians, respectively.

The null hypothesis for the relationship between the two variables in the example “*The greater the stress experienced in the job, the lower the job satisfaction of employees*,” would be H_0 : There is no relationship between stress experienced on the job and the job satisfaction of employees. This would be statistically expressed by:

$$H_0: \rho = 0$$

where ρ represents the correlation between stress and job satisfaction, which in this case is equal to 0 (i.e., no correlation).

The alternate hypothesis for the above null, which has been expressed directionally, can be statistically expressed as:

$$H_A: \rho < 0 \text{ (The correlation is negative.)}$$

For the example “*There is a relationship between age and job satisfaction*,” which has been stated nondirectionally, the null hypothesis would be statistically expressed as:

$$H_0: \rho = 0$$

whereas the alternate hypothesis would be expressed as:

$$H_A: \rho \neq 0$$

Having formulated the null and alternate hypotheses, the appropriate statistical tests (*t*-tests, *F*-tests) can then be applied, which indicate whether or not support has been found for the alternate hypothesis – that is, that there is a significant difference between groups or that there is a significant relationship between variables, as hypothesized.

The steps to be followed in hypothesis testing are:

1. State the null and the alternate hypotheses.
2. Choose the appropriate statistical test depending on whether the data collected are parametric or nonparametric.
3. Determine the level of significance desired ($p = 0.05$, or more, or less).
4. See if the output results from computer analysis indicate that the significance level is met. If, as in the case of Pearson correlation analysis in Excel software, the significance level is not indicated in the print-out, look up the critical values that define the regions of acceptance on the appropriate table (i.e., (*t*, *F*, χ^2) – see the statistical tables at the end of this book). This critical value demarcates the region of rejection from that of acceptance of the null hypothesis. When the resultant value is larger than the critical value, the null hypothesis is rejected, and the alternate accepted. If the calculated value is less than the critical value, the null is accepted and the alternate rejected.

Note that null hypotheses are rarely presented in research reports or journal articles.

Now do Exercise 5.11, Exercise 5.12, and Exercise 5.13.

EXERCISE 5.11

Create a diagram to illustrate the relationships between the relevant variables in Exercise 5.9 and develop five different hypotheses.

EXERCISE 5.12

A production manager is concerned about the low output levels of his employees. The articles that he has read on job performance frequently mention four variables as being important to job performance: (1) skills required for the job, (2) rewards, (3) motivation, and (4) satisfaction. In several of the articles it was also indicated that only if the rewards were (attractive) to the recipients did motivation, satisfaction, and job performance increase, not otherwise. Given this situation:

1. Define the problem.
2. Create a diagram.
3. Develop at least six hypotheses.

EXERCISE 5.13

A recent study has investigated the effect of corporate social responsibility (CSR) on the market value of the firm. This study developed and tested a conceptual framework, which posits that (1) customer satisfaction mediates the relationship between CSR and the market value of the firm, and (2) two firm factors (“innovativeness capability” and “product quality”) moderate the relationship between CSR and customer satisfaction. For this situation, define the problem, draw a schematic diagram, and formulate the hypotheses.

Hypothesis testing is strongly associated with designing experiments and the collection of quantitative data. However, as exemplified by Box 5.1, hypotheses can also be tested with qualitative data.

BOX 5.1**HYPOTHESIS TESTING WITH QUALITATIVE RESEARCH: NEGATIVE CASE ANALYSIS**

Hypotheses can also be tested with qualitative data. For example, let us say that, after extensive interviews, a researcher has developed the theoretical framework that unethical practices by employees are a function of their inability to discriminate between right and wrong, or due to a dire need for more money, or the organization's indifference to such practices. To test the hypothesis that these three factors are the primary ones that influence unethical practices, the researcher should look for data to refute the hypothesis. When even a single case does not support the hypothesis, the theory needs revision. Let us say that the researcher finds one case where an individual is deliberately engaged in the unethical practice of accepting kickbacks (despite the fact that he is knowledgeable enough to discriminate right from wrong, is not in need of money, and knows that the organization will not be indifferent to his behavior), simply because he wants to “get back” at the system, which “will not listen to his advice.” This new discovery, through disconfirmation of the original hypothesis, known as *the negative case method*, enables the researcher to revise the theory and the hypothesis until such time as the theory becomes robust.

We have thus far seen how a critical literature review is done, theoretical frameworks are formulated, and hypotheses developed. Let us now illustrate this logical sequence through a small example where a researcher wants to examine the organizational factors influencing women's progress to top management positions. The literature review and the number of variables are deliberately kept small, since the purpose is merely to illustrate how a theoretical framework is developed from the literature review, and how hypotheses are developed based on the theoretical framework.

EXAMPLE

Literature review, theoretical framework, and hypothesis development

Introduction

“Fewer large companies are run by women than by men named John, a sure indicator that the glass ceiling remains firmly in place in corporate America” (Wolfers, 2015). Despite the spectacular increase in the number of managerial women during the last decades, the number of women in top management positions continues to be very small and static, suggesting a glass-ceiling effect that women still face (Lückerath-Rovers, 2013; Morrison, White & Vura, 1999; O’Neil, Hopkins & Bilimoria, 2008; Van Velsor, 2000). Given the demographics of the workplace, which projects that more and more women will enter the workforce in the future, it becomes important to examine the factors that might facilitate the advancement of women to top executive positions. This study is an effort to identify the factors that currently impede women’s advancement to the top in organizations.

A brief literature review

It is often declared that since women have only recently embarked on careers and entered the managerial ranks, it will take more time for them to rise to top executive positions. However, many women in higher middle-management positions feel that there are at least two major stumbling blocks to their advancement: gender role stereotypes and inadequate access to critical information (Daniel, 1998; Koenig *et al.*, 2011; Schein, 2007; Welch, 2001).

Gender stereotypes, or sex-role stereotypes as they are also known, are societal beliefs that men are better suited for taking on leadership roles and positions of authority and power, whereas women are more suited for taking on nurturing and helping roles (DeArmond *et al.*, 2006; Eagly, 1989; Kahn & Crosby, 1998; Smith, 1999). These beliefs influence the posi-

tions that are assigned to organizational members. Whereas capable men are given line positions and developed to take on higher responsibilities and executive roles in the course of time, capable women are assigned to staff positions and dead-end jobs. With little exposure to management of budgets and opportunities for significant decision making, women are seldom groomed for top-level positions.

Women are also excluded from the “old boys” network because of their gender. Information exchange, development of career strategies, clues regarding access to resources, and such important information vital to upward mobility are thus lost to women (*The Chronicle*, 2000). While many other factors impinge on women’s upward mobility, the two variables of gender-role stereotypes and exclusion from critical information are particularly detrimental to women’s advancement to senior level positions.

Theoretical framework

The dependent variable of advancement of women to top management positions is influenced by gender-role stereotyping and access to critical information. These two variables are also interrelated as explained below.

Gender-role stereotypes adversely impact on women’s career progress. Since women are perceived as ineffective leaders but good nurturers, they are not assigned line positions in their early careers but offered staff responsibilities. It is only in line positions that managers make significant decisions, control budgets, and interact with top-level executives who have an impact on their future careers. These opportunities to learn, grow and develop on the job, and gain visibility in the system help managers to advance to top-level positions. However, since women in staff positions do not gain these experiences or have the visibility to be identified as key

people in the organization with the potential to be successful top managers, their advancement to top-level positions is never considered by the system and they are always overlooked. Thus, gender-role stereotypes hinder the progress of women to the top.

Gender-role stereotypes also hinder *access to information*. If women are not considered to be decision makers and leaders, but are perceived merely as support personnel, they will not be apprised of critical information essential for organizational advancement, since this is not seen as relevant for them. Exclusion from the networks where men informally interact with one another (golf courses, bars, and so on) precludes women from gaining access to crucial information and resources vital for their advancement. For example, many of the significant organizational changes and current events are discussed informally among men outside the work setting. Women are generally unaware of the most recent developments since they are not a part of the informal group that interacts and exchanges information away from the workplace. This definitely is a handicap. For example, knowledge of an impending vacancy for an executive

position enables one to strategize to occupy that position. One can become a key contender by procuring critical information relevant to the position, get prepared to present the appropriate credentials to the right people at the right time, and thus pave the way for success. Thus, access to critical information is important for the progress of all, including women. When women do not have the critical information that is shared in informal networks, their chances of advancement to top positions also get severely restricted.

The foregoing relationships are shown schematically in Figure 5.8.

Hypotheses

- 1. The greater the extent of gender-role stereotyping in organizations, the fewer will be the number of women at the top.*
- 2. The effect of gender-role stereotyping on advancement of women to the top is partially mediated by access to information.*

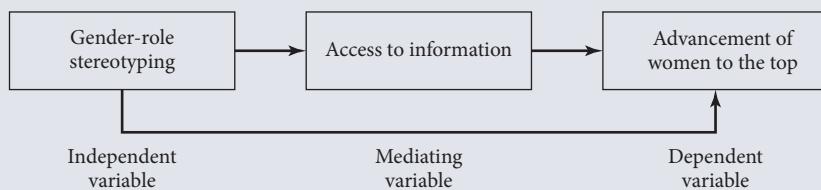


FIGURE 5.8

Schematic diagram of the example relating to women in managerial positions

MANAGERIAL IMPLICATIONS

Knowledge of how and for what purpose the theoretical framework is developed, and the hypotheses are generated, enables the manager to be an intelligent judge of the research report submitted by the consultant. At this juncture, it becomes clear that once the problem is defined, a good grasp of the concepts “independent variable” and “dependent variable” broadens the understanding of managers as to how multiple factors (the independent

variables in the model) may provide possible solutions to the problem (the dependent variable in the model). An understanding of the concept “moderating variable” may allow the manager to understand that some of the proposed solutions may not solve the problem for everybody or in every situation. Likewise, knowledge of what significance means, and why a given hypothesis is either accepted or rejected, helps the manager to persist in or desist from following hunches, which, while making good sense, do not work. If such knowledge is absent, many of the findings through research will not make much sense to the manager and decision making will bristle with confusion.

SUMMARY

- **Learning objective 1: Discuss the need for a theoretical framework in deductive research.**

A theoretical framework represents a researcher's beliefs on how certain phenomena (or variables or concepts) are related to each other (a model) and an explanation of why he or she believes that these variables are associated with each other (a theory). From a theoretical framework, testable hypotheses can be developed to examine whether a theory is valid or not. The entire deductive research project rests on the basis of the theoretical framework.

- **Learning objective 2: Describe four main types of variables and identify and label variables associated with any given situation.**

Since a theoretical framework involves the identification of a network of relationships among the variables considered important to the study of any given problem situation, it is essential to understand what a variable means and what the different types of variables are. A variable is anything that can take on differing or varying values. Four main types of variables discussed in this chapter are: (1) the dependent variable; (2) the independent variable; (3) the moderating variable; and (4) the mediating variable.

- **Learning objective 3: Develop a theoretical framework that includes all the relevant components.**

Three basic features that should be incorporated in any theoretical framework:

- The variables considered relevant to the study should be clearly defined.
- A conceptual model that describes the relationships between the variables in the model should be given.
- There should be a clear explanation of why we expect these relationships to exist.

Just as the literature review sets the stage for a good theoretical framework, this in turn provides the logical base for developing testable hypotheses.

- **Learning objective 4: Develop a set of hypotheses to be tested.**

Hypotheses are derived from the theory on which a conceptual model is based. They are often relational in nature. Along these lines, hypotheses can be defined as logically conjectured relationships between two or more variables expressed in the form of testable statements. By testing the hypotheses and confirming the conjectured relationships, it is expected that solutions can be found to correct the problem encountered.

● **Learning objective 5: Demonstrate awareness of the role of the manager in the development of a theoretical framework.**

Knowledge of how and for what purpose the theoretical framework is developed and the hypotheses are generated enables the manager to be an intelligent judge of the research report submitted by the researcher.

In the next chapter we will examine a number of basic research design issues.

Visit the companion website at www.wiley.com/college/sekaran for **Case Study: The Social Network.**

DISCUSSION QUESTIONS

1. “Because literature review is a time-consuming exercise, a good, in-depth interview should suffice to develop a theoretical framework.” Discuss this statement.
2. “Good models are complex. What’s more, a good model should include both moderating and mediating variables.” Discuss this statement.
3. “Academic researchers usually develop more complex and elaborate models than applied researchers.” Discuss this statement.
4. “In an applied research context you do not need to explain the relationships between the variables in your conceptual model.” Discuss this statement.
5. There is an advantage in stating the hypothesis both in the null and in the alternate; it adds clarity to our thinking of what we are testing. Explain.
6. It is advantageous to develop a directional hypothesis whenever we are sure of the predicted direction. How will you justify this statement?
7. In recent decades, many service markets have been liberalized. For this reason, incumbent service firms are facing new competitors and must address customer switching. You are discussing the determinants of customer switching with a service firm manager. She believes that product quality, relationship quality, and switching costs are important determinants of customer switching. You agree with the contention that product quality and relationship quality are important determinants of switching. However, you believe that switching costs *moderate* the relationships between product quality, relationship quality, and customer switching. Provide arguments for this contention.

8. For the following case:

- a. Identify the problem.
- b. Develop a diagram representing the conceptual model.
- c. Develop the hypotheses.

Concerned about her current customer base, manager Andersen started to think of factors that might affect the attractiveness of an auditing firm. Of course, the service quality provided and the fees charged by the auditor seem two important factors. Next, she decides that the reputation of the auditing firm also needs to be included in the framework as an independent variable. As illustrated by the dramatic effects of recent auditing scandals, reputation seems especially important for large auditors (i.e., auditing firms that are large in size). Finally, manager Andersen also thinks that the proximity of the auditing firm to the customer is another variable to be included as an independent variable. Proximity very likely affects the possibility for the client to personally meet with the auditors on a regular basis and she knows from her own contact with customers that they perceive personal interactions as quite important.

9. Develop a conceptual model for the scenario below.

Incidence of smoking in movies has started to increase again, after having declined for several decades. According to the National Cancer Institute, smoking is seen in at least three out of four contemporary box-office hits. What's more, identifiable cigarette brands appeared in about one-third of all movies in 2008. Exposure to smoking in movies is an important predictor of adolescent smoking initiation: smoking in movies has been shown to affect adolescents' intentions to start smoking. In turn, the intentions to start smoking are determined by a more positive attitude toward smoking after seeing a film character smoke. Recent research has revealed that the relationship between seeing a film character smoke and the attitude toward smoking is stronger when a person's identification with a film character increases. These findings are consistent with social learning theory, which predicts that attitudes and behaviors are modeled by observing the behaviors of others.

10. Develop a conceptual model for the following case.

Once given, bonuses are extraordinarily hard to take away without undermining employee morale. The adverse effects of these cuts far outweigh the anticipated savings in dollars. Research has shown that when the reason behind the cuts is explained to employees, morale does not drop.

11. Product placement is a form of advertising in which a company's products and name are intentionally positioned in motion pictures, television programs, radio broadcasts, and the like. Product placement can take many forms: verbal mentions in dialogue; actual use by a character; or visual displays (for instance, a company logo on a vehicle or billboard). Develop a theoretical framework on this issue, based on a review of the current literature. This framework should include:

- a. a specification and definition of an appropriate dependent variable;
- b. a conceptual model that describes the relationships between the dependent variable, at least one independent variable, and either a moderating or a mediating variable;
- c. a theory on why you would expect these relationships to exist;
- d. an appropriate number of testable hypotheses.

PRACTICE PROJECT

For the topic you chose to work on for the project in Chapter 4, do the following:

- Go through the computer-generated bibliography again.
- Define a problem statement that, in your opinion, would be most useful for researchers to investigate.
- Carry out a literature review that would seem to offer the greatest potential for developing a good theoretical framework, using about five to seven references.
- Develop the theoretical framework incorporating its three basic features, as discussed in the chapter.
- Generate a set of testable hypotheses based on the theoretical framework.

Selecting Samples

Saunders et al. (2023; 291-334)

Chapter 7



Selecting samples

Learning outcomes

By the end of this chapter you should be able to:

- explain why samples are necessary in business and management research;
- recognise the differences between and utility of probability and non-probability sampling procedures;
- select suitable sampling frames for a variety of research scenarios when using probability sampling;
- calculate actual sample sizes, response rates and levels of confidence when using probability sampling;
- justify your choices of probability sampling procedures for different research questions;
- be aware of a range of probability and non-probability sampling techniques and the possible need to combine techniques within a research project;
- assess the statistical representativeness of your sample when using probability sampling procedures;
- determine a suitable sample size for non-probability sampling;
- justify your choices of non-probability sampling procedures for different research questions;
- recognise the possible need to combine sampling procedures within a research project;
- apply the knowledge, skills and understanding gained to your own research project.



7.1 Introduction

Whatever your research question(s) and objectives, you will need to consider whether you need to select one or more samples. Occasionally, it may be possible to collect and analyse data from every possible case or group member; this is termed a **census**. However, for many research questions and objectives, it will be impossible for you either to collect or to analyse all the potential data available to you, owing to restrictions of time, money and often access. This means you will need to select data for a subgroup or **sample** of all possible cases. Sampling procedures enable you to reduce the amount of data you need to collect by considering only data from a subgroup rather than all possible cases or **elements**. Some research questions will require sample data that can allow you to generalise statistically about all the cases from which your sample has been selected. In the opening vignette you will see how advertisers are expected to be able to substantiate claims about consumers' views by selecting a robust and representative sample for their survey.

Other research questions may not involve such statistical generalisations. To gain an understanding of how people manage their careers, you may select a small sample of company chief executives. For such research your sample selection would be based on the premise that, as these people have reached executive level they have been very successful in managing their own careers, and so will be most likely to be able to offer insights from which you can build understanding. Alternatively, you may adopt a case study strategy using one large organisation and collect your data from a number of employees and managers using unstructured interviews. For this research you will still need to select your case study (sample) organisation and a group (sample) of employees and managers to interview. Consequently, whatever your research question, an understanding of techniques for selecting samples is likely to be very important.

The full set of cases or elements from which a sample is taken is called the **population**. In sampling, the term 'population' is not used in its normal sense, as the full set of cases need not necessarily be people. For research to discover the level of service at Indian restaurants throughout a country, the population from which you would select your sample would be all Indian restaurants in that country. Alternatively, you might need to establish the normal 'range' in miles that can be travelled by electric cars in everyday use produced by a particular manufacturer. Here the population would be all the electric cars in everyday use produced by that manufacturer.

When selecting a sample to study, it should represent the population from which it is taken in a way that is meaningful and which we can justify in relation to answering our research question and meeting our objectives (Becker 1998). If we are using our sample data to infer statistically something about a population, it is important that our sample is sufficiently large to allow such statistical inferences to be made with an acceptable margin of error. In the opening vignette we see how the requirement for advertisers to ensure the claims conveyed in their advertisements are reasonable and can be substantiated, necessitates careful consideration of sample size. We also see (in the UK) the expectation that qualifying text is included in the advertisements to allow claims made to be assessed where the findings may not be statistically significant due to the small sample size.

Interpreting advertisers' claims

Advertisers are expected to substantiate claims made in advertisements. If, for example, an advertiser claims 87 per cent of a sample of users of a skin cream said it reduced wrinkles, you might infer that 87 per cent of all that skin cream's users thought the same. Yet, whether the data on which this claim is based are sufficiently robust and representative to allow this (statistical) generalisation depends on the number of consumers from whom data were collected and how that sample of consumers had been selected.

Often, these claims are based on data collected from a sample of consumers using some form of questionnaire. When interpreting these claims, like most consumers, we usually assume the claim made from the sample is applicable to all consumers of that product or service. Not surprising advertisers are expected, often through self-regulation, to ensure there is a reasonable basis for these claims and that they are made on the basis of objective evidence. To support such self-regulation the associated industry bodies in many countries have developed codes of practice. These set out what is considered a reasonable basis and what is objective evidence, some offering guidance about sample requirements.

The UK's advertising self-regulatory system is set out in two advertising codes: the *BCAP Code* for broadcast advertising and the *CAP Code* for non-broadcasting advertising (Committee of Advertising Practice 2010;



milkos/123RF

2014). These are administered by the Advertising Standards Authority and referred to as rules or regulations. Like many systems around the world, these two sets of 'rules' require that advertisers' claims, which are likely to be regarded by consumers as objective, can be substantiated. However, although there is no easily accessible advice on sample size or associated margin of error and level of certainty, a subsequent 'quick guide' (Committee of Advertising Practice 2021) offers useful advice regarding sample size. If the sample size is relatively small so that the findings may not be statistically significant, the guide suggests it is best to include details about the sample in the advertisement. As a consequence, advertisements in the UK containing claims based on a small sample size such as '87 per cent of consumers say it reduces wrinkles*' usually include a 'small print' statement such as '*59 out of 67 consumers surveyed, August 2021'.

In this chapter we start by considering whether there is a need to sample (Section 7.2). Next we offer an overview of the two forms of sampling procedures used: probability and non-probability (Section 7.3). Our discussion of probability sampling considers a range of aspects including the sampling frame suitability (Section 7.4), sample size (Section 7.5), sample selection procedures (Section 7.6) and representativeness (Section 7.7). This is followed by a discussion of non-probability sampling where we consider aspects including sample size (Section 7.8) and sample section procedures (Section 7.9). Although each procedure is discussed separately, for many research projects you will need to use a combination of sampling procedures, some projects involving both probability and non-probability sampling techniques. Where you use a combination of two or more discrete samples, each

selected using either probability or non-probability procedures, this is known as **mixed sampling**. This along with sampling designs that have two or more successive stages using either probability, non-probability or both types of sample selection techniques are known as **multi-stage sampling**, which is discussed in Section 7.10.

7.2 The need to sample

The utility of samples

For some research questions it is possible to collect data from an entire population as it is of a manageable size. However, you should not assume that a census would necessarily provide more useful results than collecting data from a sample. Sampling provides a valid alternative to a census when:

- it would be impracticable for you to survey the entire population due to large size or inaccessibility;
- the act of collecting the data destroys the element for its intended use (such as in crash testing a car);
- your budget constraints prevent you from surveying the entire population;
- your time constraints prevent you from surveying the entire population.

For all research questions where it would be impracticable for you to collect data from the entire population, you need to select a sample. This is equally important whether you are planning to use interviews, questionnaires, observation or some other data collection technique. You might be able to obtain permission to collect data from only two or three organisations. Alternatively, testing an entire population of products to destruction, such as to establish the actual duration of long-life batteries, would be impractical for any manufacturer.

With other research questions it might be theoretically possible for you to collect data from the entire population, but the overall cost would prevent it. It is obviously cheaper for you to collect, prepare for analysis and check data from 300 customers than from 3,000, even though the cost per case for your study (in this example, customer) is likely to be higher than with a census. Your costs will be made up of new costs such as sample selection and the fact that overhead costs such as the questionnaire, interview or observation schedule design and general preparation of data for analysis are spread over a smaller number of cases. Sampling also saves time, an important consideration when you have tight deadlines. The organisation of data collection is more manageable as fewer cases are involved. As you have less data to prepare for analysis or check and then to analyse, the results will be available more quickly.

Although some claim using sampling makes possible a higher overall accuracy than a census, this appears most likely to happen where a researcher using a population fails to consider the implications of non-response, dealing with difficult to reach respondents or their population has been poorly identified. When selecting a sample, such issues are more likely to be considered carefully and addressed. In addition, the smaller number of cases for which you need to collect data means that more time can be spent designing and piloting the means of collecting these data. Collecting data from fewer cases also means that you can collect information that is more detailed. If you are employing people to collect the data (perhaps as interviewers) you can afford higher-quality staff. You can also devote more time to trying to obtain data from more difficult to reach cases so that

bias caused by potential respondents' non-response bias is minimised. Once your data have been collected, proportionally more time can be devoted to checking and testing the data for accuracy prior to analysis. However, one point remains crucial when selecting a sample: it must enable you to answer your research question!

The importance of defining the research population clearly

The sample you select should be related to the population that is highlighted in your research question and objectives. This means that if a research question is about all owners of tablets, then the population is all owners of tablet computers, and the sample selected should be a subset of all those owners. This sample, providing it is selected carefully, will allow conclusions to be drawn about all tablet owners. However, such a population may be difficult to research as it is unlikely all elements or cases will be known to you or easy to access. Consequently, you may redefine the population as something more manageable. This is often a subset of the population and is called the **target population** (Figure 7.1) and is the actual focus or target of your research. For example, rather than defining your population as all owners of tablet computers, you may redefine your target population as all owners of tablets who are studying for a business and management degree at your university. However, business and management students at one university are unlikely to be the same as all tablet owners, and even students from other universities may differ! Consequently, using a sample drawn from this target population of students to find out about all owners of a brand of tablet computer may result in biased or incorrect conclusions. In selecting your sample from this target population, you have narrowed the focus of your research to business and management students at a particular university who own that brand of tablet computer. We discuss this further in subsequent sections.

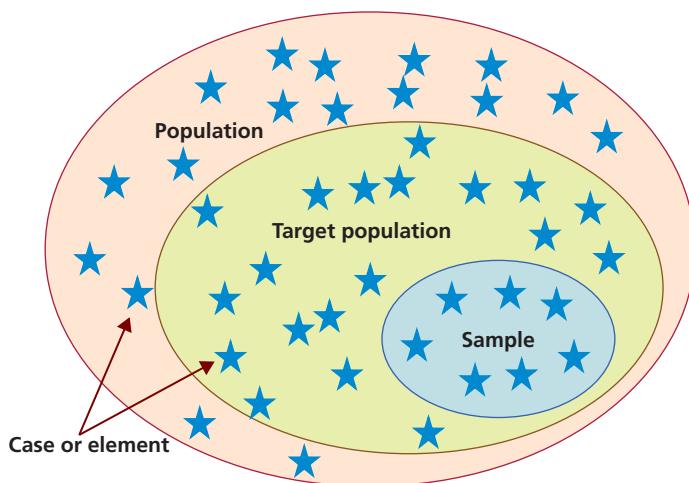


Figure 7.1 Population, target population, sample and individual cases

7.3 An overview of sampling procedures

Sampling techniques available to you can be divided into two types (Figure 7.2):

- probability or representative sampling;
- non-probability sampling.

Your choice depends, in part, on how you wish to answer your research question. We illustrate this by looking at two questions that could be answered using either type of sample: ‘What job attributes attract people to jobs?’ or ‘How are tourism companies adapting their services in response to the post-pandemic new norm?’. For either question, if you wish to offer statistical explanations, make statistical estimates or inferences about the target population from your sample or test a theory, you will choose probability sampling. However, if you wish to reveal understandings and insights by offering information rich rather than statistical explanations, where appropriate using reasoned judgements to generalise to theory, you will choose non-probability sampling (Saunders and Townsend 2018). This highlights how non-probability sampling can be used to develop theoretical generalisations based on analytic generalisability, while probability sampling can be used to generalise statistically about a target population.

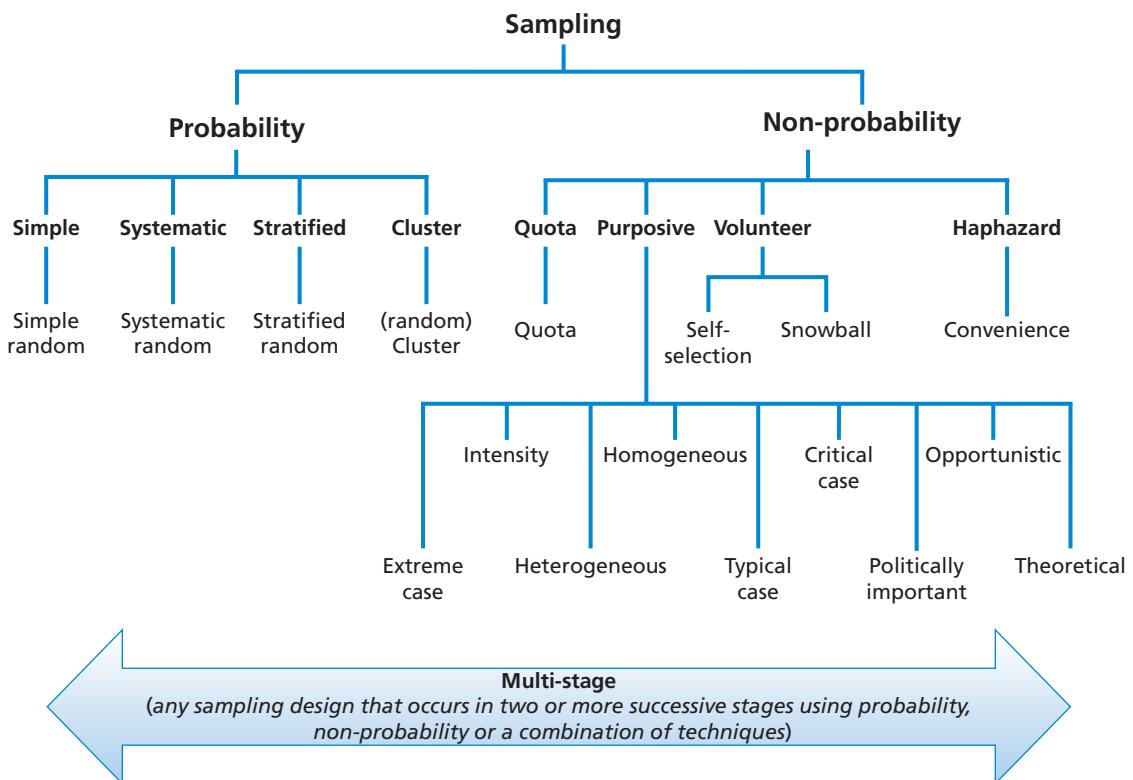


Figure 7.2 Sampling procedures

Probability samples

With **probability sampling** the chance, or probability, of each case being selected from the target population is known and is usually equal for all cases. This means it is possible to answer research questions and to achieve objectives that require you to estimate statistically the characteristics of the target population from the sample. Consequently, probability sampling is often associated with survey and experiment research strategies (Section 5.5).

Probability sampling requires a **sampling frame**, that is a complete list of all the cases in the target population from which your sample will be drawn. Without a sampling frame, you will not be able to select a probability sample and so will have to consider using non-probability sampling. If your research question or objective is concerned with members of a student society, your sampling frame will be the complete membership list for that society. If your research question or objective is concerned with registered childminders in a local area, your sampling frame will be the directory of all registered childminders in this area. Alternatively, if your research question is concerned with organisations in a particular sector, you might create a sampling frame from an existing database of companies available at your university, such as Fame or Amadeus. You would then select your sample from your list.

Probability sampling (or **representative sampling**) is associated most commonly with survey research strategies where you need to make statistical inferences from your sample about a population to answer your research question(s) and to meet your objectives. Once the decision has been taken to use probability sampling, the process can be divided into four stages:

- 1 Identify a suitable sampling frame (Section 7.4).
- 2 Decide on a suitable sample size (Section 7.5).
- 3 Select the most appropriate sampling procedure and select the sample (Section 7.6).
- 4 Assess the sample's representativeness of the target population. (Section 7.7)

However, for target populations of fewer than 50 cases, Henry (1990) advises against probability sampling. He argues that you should collect data on the entire target population, as the influence of a single extreme case on subsequent statistical analyses is more pronounced than for larger samples.

Non-probability samples

Where you do not have a sampling frame or probability sampling is not appropriate, non-probability sampling is used. With **non-probability sampling**, the probability of each case being selected from the target population is not known and it is impossible to answer research questions or to address objectives that require you to make statistical inferences about the characteristics of the population. You may still be able to generalise to other situations using non-probability samples but will need to make reasoned judgements about the transferability of the findings to other situations.

Non-probability sampling (or **non-random sampling**) provides a range of procedures to select samples, when you do not have a sampling frame or probability sampling may not be appropriate to answering your research question. It is therefore suitable to answering any research question where the focus is on gaining insights and understanding rather than statistical inference. The majority of non-probability sampling procedures (Figure 7.2) include an element of subjective judgement. In the exploratory stages of some research projects, such as a pilot testing a questionnaire, a non-probability sample may be the

most practical, although it will not allow the extent of the problem to be determined. Subsequent to this, probability sampling techniques may be used. Non-probability samples have become far more prevalent with the rapid growth of online questionnaires. For these a likely source of potential respondents is an online panel (Section 7.9) recruited in advance of the research (Baker et al. 2013). For other research projects your research question(s), objectives and choice of research strategy (Sections 2.4, 2.5 and 5.5) may dictate non-probability sampling. To answer your research question(s) and to meet your objectives you may need to undertake an in-depth study that focuses on a small number of cases, perhaps one, selected for a particular purpose. This sample would provide you with an information-rich case study in which you explore your research question and gain particular or theoretical insights.

Once a decision has been made to use non-probability sampling, the process can be divided into overlapping stages:

- 1** Consider a likely suitable sample size and review as data are collected and analysed (Section 7.8).
- 2** Select the most appropriate sampling procedure and select the sample (Section 7.9).

7.4 Probability sampling – sampling frame

Obtaining a sampling frame is essential if you are going to use probability sampling. However, as highlighted in research by Edwards et al. (2007), and more recently by Mark and colleagues (Saunders et al. 2017), you need to be aware of the possible problems of using existing databases for your sampling frame. In their work on multinationals in Britain, they found that:

- individual databases are often incomplete;
- the information held about organisations in databases is sometimes inaccurate;
- the information held in databases soon becomes out of date.

This emphasises the importance of ensuring your sampling frame is as complete, accurate and up to date as possible. An incomplete or inaccurate list means that some cases will have been excluded and so it will be impossible for every case in the target population to have a chance of selection. If this is the case you need to state it clearly.

Compiling your own sampling frame

Where no suitable list exists, and you wish to use probability sampling, you will have to compile your own sampling frame (perhaps drawing upon existing lists). It is important to ensure that your sampling frame is valid. You might decide to combine a number of online business directory to create your sampling frame from which to select a sample of typical businesses. However, each business directory will cover only subscribers who pay to be listed, often in one geographical area. Each directory will therefore be biased towards businesses that have chosen to subscribe and not a valid representation of all businesses. If any of your chosen directories are only updated annually, they will be increasingly out of date ('non-current') as the year progresses. In addition, you will need to remove duplicates caused by businesses subscribing to multiple lists! Your final sampling frame will comprise businesses that choose to subscribe when the directories were compiled.

The way you define your sampling frame also has implications regarding the extent to which you can generalise from your sample. As we have already discussed, sampling is

used when it is impracticable or unnecessary to collect data from the entire population. Within probability sampling, by defining the sampling frame you are defining the target population about which you want to generalise. This means that if your sampling frame is a list of all customers of an organisation, strictly speaking you can only generalise, that is, apply statistically the findings based upon your sample, to that target population. Similarly, if your sampling frame is all employees of an organisation (the list being the organisation's payroll) you can only generalise statistically to employees of that organisation. This can create problems, as often we hope that our findings have wider applicability than the target population from which our sample was selected. However, even if your probability sample has been selected from one large multinational organisation, you should not claim that what you have found would also occur in similar organisations. In other words, you should not generalise statistically beyond your sampling frame. Despite this, researchers often do make such claims, rather than placing clear limits on the generalisability of the findings.

Using purchased 'database' lists

An increasing number of organisations specialise in selling databases comprising electronic lists of names, addresses and email addresses. These databases list a wide range of people such as company directors, chief executives, marketing managers, production managers and human resource managers, for public, private and non-profit-making organisations, and can be merged into standard email letters such as those requesting completion of online questionnaires (Section 11.4). Because you pay for the list by the case (named individual), the organisations that provide them usually select your sample. It is therefore important to establish precisely how they will select your sample as well as obtaining an indication of the database's completeness, accuracy and currency (Saunders et al. 2017). For example, when obtaining a list of email addresses don't forget that some people change their Internet service provider and their email address regularly. This means the sampling frame is likely to under-represent this group. More generally, you need to ensure your intended sampling frame is relevant to your target population. Box 7.1 provides a checklist against which to check your sampling frame.



Box 7.1 Checklist

Selecting your sampling frame

- ✓ Are cases listed in the sampling frame relevant to your research topic, in other words does your target population enable you to answer your research question and meet your objectives?

- ✓ How recently was the sampling frame compiled; is it up to date?
- ✓ Does the sampling frame include all cases in the target population; is it complete?
- ✓ Does the sampling frame contain the correct information; is it accurate?
- ✓ Does the sampling frame exclude irrelevant cases; is it precise?
- ✓ For purchased database lists, can you establish and control precisely how the sample will be selected?

7.5 Probability sampling – sample size

Generalisations about target populations from data collected using any probability samples are based on statistical probability. The larger your sample's size the lower the likely error in generalising to the target population. Probability sampling is therefore a compromise between the accuracy of your findings and the amount of time and money you invest in collecting, checking and analysing the data. Your choice of sample size within this compromise is governed by:

- the confidence you need to have in your data – that is, the level of certainty that the characteristics of the data collected will represent the characteristics of the target population;
- the margin of error that you can tolerate – that is, the accuracy you require for any estimates made from your sample;
- the size of the target population from which your sample is being drawn;
- the (statistical) analyses you are going to undertake – in particular, the number of categories into which you wish to subdivide your data, as many statistical techniques have a minimum threshold of data cases (e.g. chi square, Section 12.5).

Given these competing influences, it is not surprising that the final sample size is almost always a matter of judgement as well as of calculation. This is particularly so for cross-national comparative research where samples selection needs to be contextually feasible in different countries and yet still yield usable, valid and comparable data (Parry et al. 2021). In addition, as we discuss in Section 12.5, if your sample is extremely large you may find that while relationships are statistically significant, the practical implications (effect size) of this difference are small (Ellis 2010).

Statistical inference – confidence and margin of error

It is likely that, if you are undertaking statistical analyses on your sample, you will be drawing conclusions from these analyses about the target population from which your sample was selected. This process of coming up with conclusions about a population on the basis of data describing the sample is called **statistical inference** and allows you to calculate how probable it is that your result, given your sample size, could have been obtained by chance. Such probabilities are usually calculated automatically by statistical analysis software. However, it is worth remembering that, providing they are not biased, samples of larger absolute size are more likely to be representative of the target population from which they are drawn than smaller samples and, in particular, the mean (average) calculated for the sample is more likely to equal the mean for the target population. This is known as the **law of large numbers**.

Researchers normally work to a 95 per cent level of certainty. This means that if your sample was selected 100 times, at least 95 of these samples would be certain to represent the characteristics of the target population. The confidence level states the precision of your estimates of the target population as the percentage that is within a certain range or margin of error (Box 7.2). Table 7.1 provides a guide to the different minimum sample sizes required from different sizes of target population given a 95 per cent confidence

Table 7.1 Sample sizes for different sizes of target population at a 95 per cent confidence level (assuming data are collected from all cases in the sample)

| Target population size | Sample size for margin of error of ... | | | |
|------------------------|--|------|------|------|
| | 5% | 3% | 2% | 1% |
| 50 | 44 | 48 | 49 | 50 |
| 100 | 79 | 91 | 96 | 99 |
| 150 | 108 | 132 | 141 | 148 |
| 200 | 132 | 168 | 185 | 196 |
| 250 | 151 | 203 | 226 | 244 |
| 300 | 168 | 234 | 267 | 291 |
| 400 | 196 | 291 | 343 | 384 |
| 500 | 217 | 340 | 414 | 475 |
| 750 | 254 | 440 | 571 | 696 |
| 1000 | 278 | 516 | 706 | 906 |
| 2000 | 322 | 696 | 1091 | 1655 |
| 5000 | 357 | 879 | 1622 | 3288 |
| 10 000 | 370 | 964 | 1936 | 4899 |
| 100 000 | 383 | 1056 | 2345 | 8762 |
| 1 000 000 | 384 | 1066 | 2395 | 9513 |
| 10 000 000 | 384 | 1067 | 2400 | 9595 |

level for different margins of error. It assumes that data are collected from all cases in the sample (details of the calculation for minimum sample size and adjusted minimum sample size are given in Appendix 2). For most business and management research, researchers are content to estimate the target population's characteristics at 95 per cent certainty to within plus or minus 3 to 5 per cent of its true values. This means that if 45 per cent of your sample are in a particular category then you will be 95 per cent certain that your estimate for the target population, within the same category, will be 45 per cent plus or minus the margin of error – somewhere between 42 and 48 per cent for a 3 per cent margin of error.

As you can see from Table 7.1, the smaller the absolute size of the sample and, to a far lesser extent, the smaller the relative proportion of the target population sampled, the greater the margin of error. Within this, the impact of absolute sample size on the margin of error decreases for larger sample sizes. De Vaus (2014) argues that it is for this reason that many market research companies limit their samples' sizes to approximately 2,000. Unfortunately, from many samples, a 100 per cent response rate is unlikely and so your sample will need to be larger to ensure sufficient responses for the margin of error you require.

Statistical analyses

For many research questions and objectives, the specific statistical analyses (Section 12.5) you need to undertake will determine the threshold sample size for individual categories. In particular, an examination of virtually any statistics textbook (or Sections 12.3 and 12.5) will highlight that, in order to ensure spurious results do not occur, the data analysed



Box 7.2 Focus on research in the news

Coronavirus infection survey to be expanded across UK

ONS to increase testing in England from 28,000 people a fortnight to 150,000 by October

By Clive Cookson and Laura Hughes

England's largest coronavirus infection survey will expand more than fivefold in the next two months to test 150,000 people every fortnight as health experts prepare for a surge of cases with the onset of autumn.

The Office for National Statistics said on Wednesday the initial boost to the sample size from the current level of 28,000 would be in place by October but it planned eventually to involve 400,000 people in England. The statistical agency, which runs the survey with Oxford University, will also extend the project to Scotland, Wales and Northern Ireland.

So far, the ONS survey has been too small to give strong statistical confidence in its findings, beyond a general trend downward in infections from April to June, followed by a slight increase during July and a levelling off so far in August.

Last week the agency said the number of people who had the virus in England was in a range of 19,000 to 40,700 with 95 per cent confidence. This estimate was based on just 58 people in the ONS sample testing positive for the virus over the previous six weeks.

The survey is designed to establish community infection levels, outside care homes and hospitals, and the ONS said the expansion would enable it to make far more accurate estimates, particularly at local level when flare-ups occur.

'Vigilance is key to containing this pandemic and the extra data on the spread of infections and antibodies at local level will be invaluable to the planning of effective local responses,' said Ian Diamond, the UK's national statistician.

Letters are being sent to tens of thousands of households inviting their participation, with north-west England a particular regional priority. 'If you've been approached to take part then please do so,' Sir Ian said. 'You will be helping us all to contain this terrible virus and get on with our lives.'

Participants provide samples from self-administered nose and throat swabs and answer a few short questions during a home visit by a health worker. The swab tests show whether or not people have the virus.

In addition, 20 per cent of participants aged 16 and over provide a blood sample. These tests help determine what proportion of the population has developed antibodies to the virus. Participants will be asked to give further blood monthly for the next year.

If the survey carries on for two years at the planned scale, it will cost around £750m, said Katherine Kent, ONS lead analyst. But the study will be assessed periodically and will be discontinued if it is found no longer to be useful.



Source: Extract from Cookson, C. and Hughes, L. (2020) 'Coronavirus infection survey to be expanded across UK', FT.com, 19 August. Copyright © 2020 The Financial Times.

must be normally distributed. While the normal distribution is discussed in Chapter 12, its implications for sample size need to be considered here. Statisticians have proved that the larger the absolute size of a sample, the closer its distribution will be to the normal distribution and thus the more robust it will be. This relationship, known as the **central limit theorem**, occurs even if the population from which the sample is drawn is not normally distributed. Statisticians have also shown that a sample size of 30 or more will usually result in a sampling distribution for the mean that is very close to a normal distribution. For this reason, Stutely's (2014) advice of a minimum number of 30 for statistical analyses provides a useful rule of thumb for the smallest number of cases in each category within your overall sample. Where the population in the category is less than 30, and you wish to undertake your analysis at this level of detail, you should normally collect data from all cases in that category.

Response rates

The most important aspect of a probability sample is that it represents the target population. A perfect **representative sample** is one that exactly represents the target population from which it is taken. If 60 per cent of your sample were small service sector companies then, provided the sample was representative, you would expect 60 per cent of the target population to be small service sector companies. You therefore need to obtain as high a response rate as possible to reduce the risk of non-response bias and ensure your sample is representative (Groves and Peytcheva 2008). This is not to say that a low response rate will necessarily result in your sample being biased, just that it is more likely!

In reality, you are likely to have non-responses. Non-respondents are different from the rest of the target population because they are unable or unwilling to be involved in your research for whatever reason. Consequently, your respondents will not be representative of the target population and the data you collect may be biased. Bias resulting from respondents differing in meaningful ways from non-respondents is known as **non-response bias**. In addition, each non-response will necessitate an extra respondent being found to reach the required sample size, increasing the cost of your data collection.

You should therefore collect data on refusals to respond to both individual questions and entire questionnaires or interview schedules to check for non-response bias (Section 12.2) and report this briefly in your project report. For returned questionnaires or structured interviews, four levels of non-response can be reported with regard to the proportion of applicable questions that have been answered (American Association for Public Opinion Research 2016):

- **complete refusal:** none of the questions answered;
- **break-off:** less than 50 per cent of all questions answered other than by a refusal or no answer (this therefore includes complete refusal);
- **partial response:** 50 per cent to 80 per cent of all questions answered other than by a refusal or no answer;
- **complete response:** over 80 per cent of all questions answered other than by a refusal or no answer.

Non-response is due to four interrelated problems:

- refusal to respond;
- ineligibility to respond;
- inability to locate respondent;
- respondent located but unable to make contact.

The most common reason for non-response is your respondent refuses to answer all the questions or be involved in your research but does not give a reason. Such non-response can be minimised by paying careful attention to the methods used to collect your data (Chapters 9, 10 and 11). Alternatively, some selected respondents may not meet your research requirements and so will be **ineligible** to respond. Non-location and non-contact create further problems; the fact that these respondents are **unreachable** means they will not be represented in the data you collect.

As part of your research report, you will need to include your **response rate**. Neuman (2014) suggests that when you calculate this you should include all eligible respondents:

$$\text{total response rate} = \frac{\text{total number of responses}}{\text{total number in sample} - \text{ineligible}}$$

This he calls the **total response rate**. A more common way of doing this excludes ineligible respondents and those who, despite repeated attempts (Sections 10.7 and 11.8), were unreachable. This is known as the **active response rate**:

$$\text{active response rate} = \frac{\text{total number of responses}}{\text{total number in sample} - (\text{ineligible} + \text{unreachable})}$$

Examples of calculations of the total response rate and the active response rate are given in Box 7.3.

Even after ineligible and unreachable respondents have been excluded, it is probable that you will still have some non-responses. You therefore need to be able to assess how representative your data are and to allow for the impact of non-response in your calculations of sample size. These issues are explored in subsequent sections.



Box 7.3 Focus on student research

Calculation of total and active response rates

Ming had decided to collect data from people who had left his company's employment over the past five years by using a web questionnaire. He obtained a list of the 1,034 people who had left over this period (the total population) and selected a 50 per cent sample. Unfortunately, he

could obtain current email addresses for only 311 of the 517 ex-employees who made up his total sample. Of these 311 people who were potentially reachable, he obtained a response from 147. In addition, his list of people who had left his company was inaccurate, and nine of those he contacted were ineligible to respond, having left the company over five years earlier.

$$\text{His total response rate} = \frac{147}{517 - 9} = \frac{147}{508} = 28.9\%$$

$$\text{His active response rate} = \frac{147}{311 - 9} = \frac{147}{302} = 48.7\%$$

Estimating response rates and sample size

With all probability samples, it is important that your sample size is large enough to provide you with the necessary confidence in your data. The margin of error must be within acceptable limits, and you must ensure that you will be able to undertake your analysis at the level of detail required. You therefore need to estimate the likely response rate – that is, the proportion of cases from your sample who will respond or from which data will be collected – and increase the sample size accordingly. Once you have an estimate of the likely response rate and the minimum or the adjusted minimum sample size, the actual sample size you require can be calculated using the following formula:

$$n^a = \frac{n \times 100}{re\%}$$

where n^a is the actual sample size required,
 n is the minimum (or adjusted minimum) sample size (see Table 7.1 or Appendix 2),
 $re\%$ is the estimated response rate expressed as a percentage.

This calculation is shown in Box 7.4.

If you are collecting your sample data from a secondary source (Section 8.2) within an organisation that has already granted you access, for example a database recording customer complaints, your response rate should be virtually 100 per cent. Your actual sample size will therefore be the same as your minimum sample size.

In contrast, estimating the likely response rate from a sample to which you will be sending a questionnaire or interviewing is more difficult. One way of obtaining this estimate is to consider the response rates achieved for similar surveys already undertaken, basing your estimate on these. Alternatively, you can err on the side of caution. For general business and management studies involving individuals or organisations' representatives, response rates of approximately 50 per cent and 35 to 40 per cent respectively are reasonable (Baruch and Holtom 2008). However more recent research suggests that overall response rates are declining (Chidlow et al. 2015).



Box 7.4 Focus on student research

Calculation of actual sample size

Jan was a part-time student employed by a large manufacturing company. He had decided to email a questionnaire to the company's customers and calculated that an adjusted minimum sample size of 439 was required. From previous questionnaires that his company had used to collect data from customers, Jan knew the likely response rate would be approximately

30 per cent. Using these data he could calculate his actual sample size:

$$\begin{aligned} n^a &= \frac{439 \times 100}{30} \\ &= \frac{43900}{30} \\ &= 1463 \end{aligned}$$

Jan's actual sample, therefore, needed to be 1,463 customers. The likelihood of 70 per cent non-response meant that Jan needed to include a means of checking his sample was representative when he designed his questionnaire.

Response rates can vary considerably when collecting primary data. Reviewing literature on response rates for questionnaires, Mellahi and Harris (2016) noted wide variation and no consensus as to what was acceptable. Noting response rates of between 1 per cent and 100 per cent in published Business and Management research they offer general guidelines dependent upon discipline suggesting mean response rates of 35 per cent for International Business and Marketing, while Human Resource Management and General Management typically achieve 50 per cent. More recent work by Pielsticker and Hiebl (2020) reveals that for family businesses survey research response rates are, at approximately 20 per cent, even lower.

Looking at mode of questionnaire delivery, Neuman (2014) suggests response rates of between 10 and 50 per cent for postal questionnaire surveys and up to 90 per cent for face-to-face interviews. Our examination of response rates to recent business surveys reveals rates as low as 10–20 per cent for web and postal questionnaires, an implication being that respondents' questionnaire fatigue was a contributory factor! With regard to telephone questionnaires, response rates have fallen from 36 per cent to less than 9 per cent, due in part to people using answering services to screen calls (Dillman et al. 2014). Fortunately, a number of different interventions, depending on your data collection method, can be used to enhance your response rate. These are discussed with the data collection method in the appropriate sections (Sections 10.3 and 11.5).

Reporting sample selection and response rates

General suggestions regarding reporting sample selection in your project report are offered in the checklist in Box 14.7. When reporting response rates it is helpful to state whether the questionnaire or structured interview was administered to respondents (answered as part of their job, role or studies) or truly voluntary, as well as providing sufficient detail regarding how the sample were selected and the questionnaire distributed and returned. The aspects to include when reporting questionnaire response rates are summarised in Box 7.5 as a checklist.



Box 7.5 Checklist

Reporting questionnaire response rates

- ✓ Was the questionnaire administered or truly voluntary?
- ✓ How many respondents were sent the questionnaire?
- ✓ How was the questionnaire distributed?
- ✓ How was the questionnaire returned?
- ✓ Was prior consent obtained from respondents?

- ✓ How many questionnaires were returned?
- ✓ Of those returned, how many were complete responses, and how many were partial responses?
- ✓ What were the reasons (if known) for non-response?
- ✓ Where different populations received a questionnaire, what were there differences (if any) in response rates?
- ✓ What interventions (if any) were used to increase response rates?
- ✓ Where response rates differ from likely norms, what are possible reasons for this?

Sources: Baruch and Holtom (2008); American Association for Public Opinion Research 2016.

7.6 Probability sampling – procedures

Having chosen a suitable sampling frame and established the actual sample size required, you need to select the most appropriate sampling procedure to obtain a representative sample. Four main procedures can be used when selecting a probability sample (Figure 7.3):

- simple random;
- systematic random;
- stratified random;
- cluster.

Your choice of probability sampling procedure depends on your research question(s) and your objectives. Subsequently, your need for face-to-face contact with respondents, and the geographical area over which the population is spread, further influence your choice of probability sampling procedure (Figure 7.3). The structure of the sampling frame, the size of sample you need and, if you are using a research assistant, the ease with which the procedure may be explained will also influence your decision. The impact of each of these is summarised for probability sampling procedures in Table 7.2.

Simple random sampling

Simple random sampling (often called just **random sampling**) involves you selecting the sample at random from the sampling frame using a spreadsheet's random number generator function or random number tables. To do this you:

- 1 Number each of the cases in your sampling frame with a unique number. The first case is numbered 1, the second 2 and so on.
- 2 Select cases using random numbers such as those generated by a spreadsheet (Table 7.3) until your actual sample size is reached.

Starting with your first random number, you use this and subsequent random numbers in the order they were generated to select the cases (elements) until your sample size is reached. If the same random number is generated more than once it must be disregarded as you need different cases. This means that you are not putting each case's number back into the sampling frame after it has been selected and is termed 'sampling without replacement'. If a number is selected that is outside the range of those in your sampling frame, you simply ignore it and continue reading off numbers until your sample size is reached (Box 7.6).

Random numbers allow you to select your sample without bias. The sample selected, therefore, can be said to be representative of the target population. However, it is not a perfect miniature replica of this population, since it still possesses sampling error. In addition, the selection that simple random sampling provides is more evenly dispersed throughout the target population for samples of more than a few hundred cases. The first few hundred cases selected using simple random sampling normally consist of groups of cases whose numbers are close together followed by a gap and then a further grouping. For more than a few hundred cases, this pattern occurs far less frequently. Because of the technique's random nature it is possible that a chance occurrence of such patterns will result in certain parts of a population being over- or under-represented.

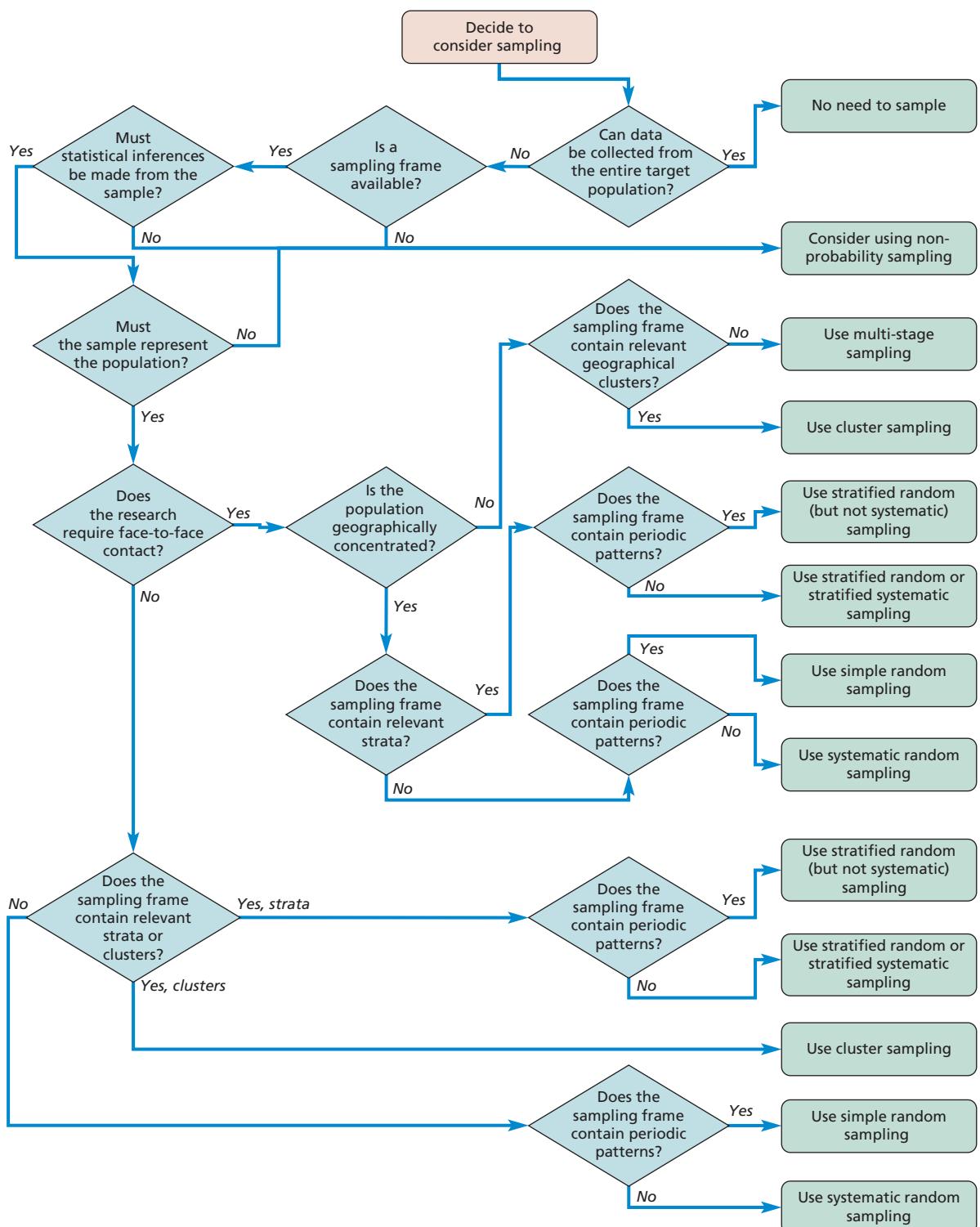


Figure 7.3 Selecting a probability sampling procedure

Note: Simple random sampling ideally requires a sample size of over a few hundred.

Table 7.2 Impact of various factors on choice of probability sampling procedures

| | | Factor | | | | |
|-------------------|---|---|---|--|--|---|
| Sample technique | Sampling frame required | Size of sample needed | Geographical area to which suited | Relative cost | Easy to explain to support workers? | Advantages compared with simple random |
| Simple random | Accurate and easily accessible | Better with over a few hundred | Concentrated if face-to-face contact required, otherwise does not matter | High if large sample size or sampling frame not computerised | Relatively difficult to explain | |
| Systematic random | Accurate, easily accessible and not containing periodic patterns. Actual list not always needed | Suitable for all sizes | Concentrated if face-to-face contact required, otherwise does not matter | Low | Relatively easy to explain | Normally no difference |
| Stratified random | Accurate, easily accessible, divisible into relevant strata (see comments for simple random and systematic random as appropriate) | See comments for simple random and systematic random as appropriate | Concentrated if face-to-face contact required, otherwise does not matter | Low, provided that lists of relevant strata available | Relatively difficult to explain (once strata decided, see comments for simple random and systematic random as appropriate) | Better comparison and hence representation across strata. Differential response rates may necessitate reweighting |
| Cluster | Accurate, easily accessible, relates to relevant clusters, not individual population members | As large as practicable | Dispersed if face-to-face contact required and geographically based clusters used | Low, provided that lists of relevant clusters available | Relatively difficult to explain until clusters selected | Quick but reduced precision |

Source: © Mark Saunders, Philip Lewis and Adrian Thornhill 2021

Table 7.3 Extract of spreadsheet generated random numbers between 1 and 5011

| | | | | | | | | | |
|-------------|------|------|------|------|------|------|------|------|------|
| 4306 | 1966 | 1878 | 4428 | 3571 | 62 | 838 | 4881 | 3045 | 4192 |
| 4582 | 4543 | 457 | 4151 | 1208 | 2014 | 3891 | 111 | 4197 | 1455 |
| 1303 | 2463 | 151 | 1236 | 2822 | 4539 | 1970 | 3788 | 3070 | 967 |
| 1547 | 139 | 3175 | 3773 | 3883 | 2161 | 209 | 2364 | 2324 | 3849 |
| 2009 | 4352 | 4685 | 4820 | 1386 | 4990 | 786 | 4516 | 2851 | 2571 |
| 3589 | 539 | 2809 | 2065 | 1548 | 661 | 4506 | 788 | 4082 | 1450 |
| 4311 | 4827 | 3137 | 3000 | 69 | 1925 | 393 | 42 | 1032 | 3450 |
| 1605 | 1105 | 4949 | 1791 | 2761 | 879 | 709 | 221 | 2894 | 3232 |

Simple random sampling is best used when you have an accurate and easily accessible sampling frame that lists the target population, preferably in electronic format. While you can often obtain these for employees within organisations or members of clubs or societies, adequate lists are less likely to be available for organisations. If your population covers a large geographical area, random selection means that selected cases are likely to be dispersed throughout the area. Consequently, this form of sampling is not suitable if collecting data over a large geographical area using a method that requires face-to-face contact, owing to the associated high travel costs. Simple random sampling would still be suitable for a geographically dispersed area if you used an alternative procedure of collecting data such as Internet or postal questionnaires or telephone or Internet mediated interviewing (Chapter 11).



Box 7.6 Focus on student research

Simple random sampling

Jemma was undertaking her work placement at a large supermarket, where 5,011 of the supermarket's customers used the supermarket's online shopping and home delivery scheme. She was asked to interview customers and find out what aspects they liked and disliked. As there was insufficient time to interview all of them she decided to use a structured telephone interview. Her calculations revealed that to obtain acceptable levels of confidence and accuracy she needed an actual sample size of approximately 360 customers. Assuming a response rate of 30 per cent, this meant she needed to contact 1,200 Internet customers. She decided to select them using simple random sampling.

Having obtained a list of Internet customers and their telephone numbers, Jemma gave each of the cases (customers) in this sampling frame a unique number starting with 1 through to 5,011.

Using her spreadsheet's random number generator function, Jemma generated a series of random numbers between 1 and 5,011. The first random number generated was 4,306 (shown in bold and shaded in Table 7.3). Starting with this number she used the random numbers in the order they were generated (in this example continuing along the line) to select her cases:

4306 1966 1878 4428 3571 62 838 4881 3045
...

She continued in this manner until 1,200 different cases had been selected, ensuring that where a random number was repeated, the associated case was disregarded and the cases selected therefore all different. These 1,200 cases selected formed her random sample, of whom she expected 360 to participate.

Sampling frames used for computer-aided telephone interviewing (CATI) have, in the main, been replaced by random digital dialling. Although selecting particular within-country area dialling codes for land-line telephone numbers provides a chance to reach any household within that area represented by that code that has a landline telephone, regardless of whether or not the number is ex-directory, care must be taken. Such a sample excludes households who use only mobile telephones as their dialling codes are network operator rather than geographical area specific (Tucker and Lepkowski 2008).

Systematic random sampling

Systematic random sampling (often called just **systematic sampling**) involves you selecting the sample at regular intervals from the sampling frame. To do this you:

- 1 Number each of the cases in your sampling frame with a unique number. The first case is numbered 1, the second 2 and so on.
- 2 Select the first case using a random number.
- 3 Calculate the sampling fraction.
- 4 Select subsequent cases systematically using the sampling fraction to determine the frequency of selection.

To calculate the **sampling fraction** – that is, the proportion of the target population that you need to select – you use the formula:

$$\text{Sampling fraction} = \frac{\text{actual sample size}}{\text{total population}}$$

If your sampling fraction is 1/3 you need to select one in every three cases – that is, every third case from the sampling frame. Unfortunately, your calculation will usually result in a more complicated fraction. In these instances it is normally acceptable to round your population down to the nearest 10 (or 100) and to increase your minimum sample size until a simpler sampling fraction can be calculated.

On its own, selecting one in every three would not be random as every third case would be bound to be selected, whereas those between would have no chance of selection. To overcome this, a random number is used to decide where to start on the sampling frame. If your sampling fraction is 1/3 the starting point must be one of the first three cases. You therefore generate a random number (in this example a one-digit random number between 1 and 3) as described earlier and use this as the starting point. Once you have selected your first case at random you then select, in this example, every third case until you have gone right through your sampling frame (Box 7.7).

In some instances it is not necessary to actually construct a list for your sampling frame. For Internet questionnaires, such as pop-up questionnaires that appear in a window on the computer screen, there is no need to create an actual list if an invitation to participate is triggered at random. For systematic random sampling, a random selection could be triggered by a mechanism such as every tenth visitor to the website over a specified time period (Bradley 1999).

Despite the advantages, you must be careful when using existing lists as sampling frames. You need to ensure that the lists do not contain periodic patterns. Let us assume a high street bank needs you to administer a questionnaire to a sample of customers with joint bank accounts. A sampling fraction of 1/2 means that you will need to select every second customer on the list. The names on the customer lists, which you intend to use



Box 7.7 Focus on student research

Systematic random sampling

Stefan worked as a receptionist in a dental surgery with approximately 1,500 patients. He wished to find out their attitudes to the new automated appointments scheme. As there was insufficient time and money to collect data from all patients using a questionnaire he decided to email a hyperlink for the web questionnaire to a sample. The calculation of sample size revealed that to obtain acceptable levels of confidence and accuracy he needed an actual sample size of approximately 300 patients. Having obtained ethical approval he generated an alphabetical list of all registered patients from the patient record system for his sampling frame and selected his sample systematically.

First, he calculated the sampling fraction:

$$\frac{300}{1500} = \frac{1}{5}$$

This meant that he needed to select every fifth patient from the sampling frame. Next, he used a random number to decide where to start on his sampling frame. As the sampling fraction was 1/5, the starting point had to be one of the first five patients. He therefore selected a one-digit random number between 1 and 5.

Once he had selected his first patient at random he continued to select every fifth patient until he had gone right through his sampling frame (the list of patients). As the random number Stefan selected was 2, the selected the following patient numbers:

2 7 12 17 22 27 32 37 ...

and so on until 300 patients had been selected.

as the sampling frame, are arranged alphabetically by joint account, with predominantly males followed by females (Table 7.4). If you start with a male customer, the majority of those in your sample will be male. Conversely, if you start with a female customer, the majority of those in your sample will be female. Consequently, your sample will be biased (Table 7.4). Systematic random sampling is therefore not suitable without reordering or stratifying the sampling frame (discussed later).

Unlike simple random sampling, systematic random sampling works equally well with a small or large number of cases. However, if your target population covers a large geographical area, the random selection means sample cases are likely to be dispersed throughout the area. Consequently, systematic random sampling is suitable for geographically dispersed cases only if you do not require face-to-face contact when collecting your data.

Table 7.4 The impact of periodic patterns on systematic random sampling

| Number | Customer | Sample | Number | Customer | Sample |
|--------|-----------------|--------|--------|----------------|--------|
| 1 | Mr J. Lewis | ✓ | 7 | Mr J. Smith | ✓ |
| 2 | Mrs P. Lewis | * | 8 | Mrs K. Smith | * |
| 3 | Mr T. Penny | ✓ | 9 | Mr R. Thompson | ✓ |
| 4 | Mrs J. Penny | * | 10 | Ms M. Wroot | * |
| 5 | Mr A. Saunders | ✓ | 11 | Mr J. Whalley | ✓ |
| 6 | Mrs C. Saunders | * | 12 | Mr C. Simon | * |

✓ Sample selected if you start with 1. * Sample selected if you start with 2.

Stratified random sampling

Stratified random sampling is a modification of random sampling in which you divide the target population into two or more relevant and significant strata based on one or a number of attributes relevant to your research question and objectives. In effect, your sampling frame is divided into a number of subsets. A random sample (simple or systematic) is then drawn from each of the strata. Consequently, stratified random sampling shares many of the advantages and disadvantages of simple random or systematic random sampling.

Dividing the population into a series of relevant strata means that the sample is more likely to be representative, as you can ensure that each of the strata is represented proportionally within your sample. However, it is only possible to do this if you are aware of, and can easily distinguish, significant strata in your sampling frame. In addition, the extra stage in the sampling procedure means that it is likely to take longer, to be more expensive and to be more difficult to explain than simple random or systematic random sampling.

In some instances, as pointed out by De Vaus (2014), your sampling frame will already be divided into strata. A sampling frame of employee names that is in alphabetical order will automatically ensure that, if systematic random sampling is used (discussed earlier), employees will be sampled in the correct proportion to the letter with which their name begins. Similarly, membership lists that are ordered by date of joining will automatically result in stratification by length of membership if systematic random sampling is used. However, if you are using simple random sampling or your sampling frame contains periodic patterns, you will need to stratify it. To do this you:

- 1** Choose the stratification variable or variables.
- 2** Divide the sampling frame into the discrete strata.
- 3** Number each of the cases within each stratum with a unique number, as outlined earlier.
- 4** Select your sample using either simple random or systematic random sampling, as outlined earlier.

The stratification variable (or variables) chosen should represent the discrete characteristic (or characteristics) for which you want to ensure correct representation within the sample (Box 7.8).

Samples can be stratified using more than one characteristic. You may wish to stratify a sample of an organisation's employees by both department and salary grade. To do this you would:

- 1** Divide the sampling frame into the discrete departments.
- 2** Within each department divide the sampling frame into discrete salary grades.
- 3** Number each of the cases within each salary grade within each department with a unique number, as outlined earlier.
- 4** Select your sample using either simple random or systematic random sampling, as outlined earlier.

In some instances the relative sizes of different strata mean that, in order to have sufficient data for analysis, you need to select larger samples from the strata with smaller target populations. Here the different sample sizes must be taken into account when aggregating data from each of the strata to obtain an overall picture. More sophisticated statistical analysis software packages enable you to do this by differentially weighting the responses for each stratum (Section 12.2).



Box 7.8 Focus on student research

Stratified random sampling

Dilek worked for a major supplier of office supplies to public and private organisations. As part of her research into her organisation's customers, she needed to ensure that both public- and private-sector organisations were represented correctly. An important

stratum was, therefore, the sector of the organisation. Her sampling frame was therefore divided into two discrete strata: public sector and private sector. Within each stratum, the individual cases were then numbered (see below).

She decided to select a systematic random sample. A sampling fraction of 1/4 meant that she needed to select every fourth customer on the list. As indicated by the ticks (✓), random numbers were generated to select the first case in the public sector (2) and private sector (4) strata. Subsequently, every fourth customer in each stratum was selected.

| Public sector stratum | | | Private sector stratum | | |
|-----------------------|---------------------------------|----------|------------------------|----------------------------------|----------|
| Number | Customer | Selected | Number | Customer | Selected |
| 1 | Anyshire County Council | | 1 | ABC Automotive manufacturer | |
| 2 | Anyshire Hospital Trust | ✓ | 2 | Anytown printers and bookbinders | |
| 3 | Newshire Army Training Barracks | | 3 | Ben Toy Company | |
| 4 | Newshire Police Force | | 4 | Jane's Internet Flower Shop | ✓ |
| 5 | Newshire Housing | | 5 | Multimedia Productions | |
| 6 | St Peter's Secondary School | ✓ | 6 | Roger's Consulting | |
| 7 | University of Anytown | | 7 | The Paperless Office | |
| 8 | West Anyshire Council | | 8 | U-need-us Ltd | ✓ |

Cluster sampling

Cluster sampling (sometimes known as **one-stage cluster sampling**) is, on the surface, similar to stratified random sampling as you need to divide the target population into discrete groups prior to sampling (Latpate et al. 2021). The groups are termed clusters in this form of sampling and can be based on any naturally occurring grouping. For example, you could group your data by type of manufacturing firm or geographical area (Box 7.9).

For cluster sampling, your sampling frame is the complete list of clusters rather than a complete list of individual cases within the population. You then select a few clusters, normally using simple random sampling. In one-stage cluster sampling, data are collected from every case within the selected clusters. For two-stage cluster sampling, some other form of sampling is used to select cases from those clusters selected at stage one. One-stage cluster-sampling has three main steps:



Box 7.9 **Focus on student research**

Cluster sampling

Ceri needed to select a sample of firms from which to collect data using an interviewer completed face-to-face questionnaire about the workspace adaptations to protect workers due to the pandemic. As she had limited resources with which to pay for travel and other associated data collection costs, she decided

to collect data from firms in four geographical areas selected from a cluster grouping of local administrative areas. A list of all local administrative areas formed her sampling frame. Each of the local administrative areas (clusters) was given a unique number, the first being 1, the second 2 and so on. The four sample clusters were selected from this sampling frame of local administrative areas using simple random sampling.

Ceri's sample was all firms within the selected clusters. She decided that the appropriate directories could probably provide a suitable list of all firms in each cluster.

- 1** Choose the cluster grouping for your sampling frame.
- 2** Number each of the clusters with a unique number. The first cluster is numbered 1, the second 2 and so on.
- 3** Select your sample of clusters using some form of random sampling, as discussed earlier.

Selecting clusters randomly makes cluster sampling a probability sampling technique. Despite this, the technique normally results in a sample that represents the target population less accurately than stratified random sampling. Restricting the sample to a few relatively compact geographical sub-areas (clusters) maximises the amount of data you can collect using face-to-face methods within the resources available. However, it may also reduce the representativeness of your sample. For this reason you need to maximise the number of sub-areas to allow for variations in the target population within the available resources. Your choice is between a large sample from a few discrete subgroups and a smaller sample distributed over the whole group. It is a trade-off between the amount of precision lost by using a few subgroups and the amount gained from a larger sample size.

7.7 Probability sampling – representativeness

It is often possible to compare data you collect from your sample with data from another source for the population, such as data contained in an 'archival' database. For example, you can compare data on the age and socioeconomic characteristics of respondents in a marketing survey with these characteristics for the population in that country as recorded by the latest national census of population. If there is no statistically significant difference, then the sample is representative with respect to these characteristics.

When working within an organisation, comparisons can also be made. In a questionnaire Mark sent to a sample of employees in a large UK organisation, he asked closed questions about salary grade, gender, length of service and main place of work. Possible responses to each question were designed to provide sufficient detail to compare the characteristics of the sample with the characteristics of the entire population of employees as recorded by the organisation's Human Resources (HR) database. At the same time he kept the categories sufficiently broad to preserve, and to be seen to preserve, the anonymity of individual respondents. The two questions on length of service and salary grade from a questionnaire he developed illustrate this:

37 How long have you worked for organisation's name?

less than 1 year 1 year to less than 3 years 3 or more years

38 Which one of the following best describes your salary grade?

| | | | |
|---------------------------|--------------------------|----------------------------------|--------------------------|
| Clerical (grades 1–3) | <input type="checkbox"/> | Management (grades 9–11) | <input type="checkbox"/> |
| Supervisory (grades 4–5) | <input type="checkbox"/> | Senior management (grades 12–14) | <input type="checkbox"/> |
| Professional (grades 6–8) | <input type="checkbox"/> | Other (please say) | <input type="checkbox"/> |

Using the Kolmogorov test (Section 12.5), Mark found there was no statistically significant difference between the proportions of respondents in each of the length of service groups and the data obtained from the organisation's HR database for all employees. This meant that the sample of respondents was representative of all employees with respect to length of service. However, those responding were (statistically) significantly more likely to be in professional and managerial salary grades than in technical, administrative or supervisory salary grades. It is therefore important to discuss the representativeness of your findings when reporting how the sample was selected (Box 14.7).

Representativeness of samples can be assessed in a variety of other ways (Rogelberg and Stanton 2007). Those our students have used most often, in order of quality of assessment of possible bias, include:

- replicating your findings using a new sample selected using different sampling techniques, referred to as 'demonstrate generalisability';
- resurveying non-respondents, the 'follow-up approach';
- analysing whether non-response was due to refusal, ineligibility or some other reason through interviews with non-respondents, known as 'active non-response analysis';
- comparing late respondents' responses with those from early respondents, known as 'wave analysis'.

In relation to this list, the quality of the assessment of bias provided by archival analysis, as outlined earlier, is similar to that provided by the follow-up approach and active non-response analysis.

7.8 Non-probability sampling – sample size

For all non-probability sampling procedures, other than for quota samples (which we discuss later), the issue of sample size is not so much about how many cases need to be selected as which cases need to be selected (Hammersley 2015). Crucial is the logical relationship between your sample selection procedure and the purpose and focus of your research (Figure 7.4); the sample selected being used, for example, to illustrate a particular aspect or to make generalisations to theory rather than about a population (Box 7.10).

Often a case study strategy uses a sample of one or two case studies to explore a particular phenomenon or institution in depth (Lee and Saunders 2017). Data are then collected from all participants or from some form of sample of participants for these case studies. Your sample size is therefore dependent on your research question(s) and objectives – in particular, what you need to find out, what will be useful, what will have credibility and what can be done within your available resources (Patton 2015). This is particularly so where you are intending to collect qualitative data using participant observation, semi-structured or unstructured interviews (Chapters 9 and 10). The understanding and insights that you will gain from your data will be more to do with whether it enables you to develop



Box 7.10 Focus on management research

Volunteer sampling

In their 2019 *Journal of Industrial Relations* article Goods and Colleagues (Goods et al. 2019) used a qualitative industry case study design to explore job quality in the food-delivery sector of the Australian gig economy. Within this their aim was to make sense of workers' experiences in the wider debates concerning the future of work, rather than identifying the size and scale of such work. A sampling decision was made to interview riders operating on the Deliveroo and

UberEATS platforms in Melbourne and Perth, these sites being the two locations where the research team were located.

Fifty-eight participants were recruited using a combination of methods including street intercepts, riders were 'on-call' and available for work but not actively engaged being interviewed while they waited for work; online recruitment initiatives, through which participants self-selected; and snowball sampling, for example those recruited through street intercepts providing further referrals.

Drawing on these workers' accounts of economic security, autonomy and enjoyment to assess job quality they Goods et al. (2019: 502) argue 'the gig economy is a new juncture in capitalist production, the consequences of which need to be taken seriously'.

a fruitful analytic argument (Hammersley 2015), and your data collection and analysis skills than with the size of your sample (Patton 2015).

In addressing 'how many are likely to be needed', many research textbooks simply recommend continuing to collect qualitative data, such as by conducting additional interviews, until **data saturation** is reached: in other words, until the additional data collected provide little, if any, new information or suggest new themes. In relation to this, Francis et al. (2010), suggest continuing to collect data to for three more interviews to ensure saturation has definitely been reached. However, while some consider saturation to be crucial (Guest et al. 2006) to establishing how many interviews or observations are required; others note that not reaching saturation only means the phenomenon has still to be fully explored and that the findings are still valid (O'Reilly and Parker 2013). Saturation is also inappropriate for some research questions such as, for example, where research is to establish whether something is possible. Not surprisingly Sim et al. (2018) note determining the sample size for interpretative qualitative research is problematic.

Despite this, we believe it is possible to offer guidance regarding likely sample sizes to help in planning your data collection when using observations or interviews. Mark (Saunders 2012) summarises the limited guidance available as between four and 12 participants for a homogenous and 12 and 30 participants for a heterogeneous group. This he notes differs between groups, research strategies and complexities and is dependent upon the research question. His more recent research on practices in published organisation and workplace research (Saunders and Townsend 2016), while recognising that for some research purposes a sample of one can be sufficient, offers guidance on credible sample sizes when planning qualitative interviews. This is summarised in Table 7.5.

Reporting data saturation

Few of the academic research papers you read will mention data saturation and, of those that do, few actually provide evidence of when data saturation was reached. One way of providing evidence is to record for each additional case or element from which data were

Table 7.5 Non-probability sample size norms when using qualitative interviews

| Purpose | Sample size norm |
|---|------------------|
| Planning research where participants are from a single organisation or will be analysed as a single group | 30 |
| Planning research where participants are from multiple organisations or will be analysed in multiple groups | 50 |
| Overall number likely to be considered sufficient | 15–60 |

Source: Developed from Saunders and Townsend (2016)

Table 7.6 Demonstrating data saturation

| Case number | Number of new themes coded |
|-------------|----------------------------|
| 1 | 10 |
| 2 | 13 |
| 3 | 1 |
| 4 | 0 |
| 5 | 2 |
| : | : |
| 27 | 1 |
| 28 | 0 |
| 29 | 0 |
| 30 | 0 |

collected, whether any new information was provided or new themes or insights suggested. To show this you can create a table listing for each case how many new themes were generated when the data were coded. In Table 7.6 data collected from the first interviewee (Case 1) revealed 10 new themes. An additional 13 new themes were revealed by the second interviewee (Case 2). Subsequent interviewees revealed either one, two or no new themes. However, it was not until the 30th interview that there had been three interviews in a row from which no new themes were suggested.

Reporting sample selection

General suggestions regarding reporting sample selection in your project report are offered in the Checklist in Box 14.7. However, when discussing the final non-probability sample from which your data were collected, it can be helpful to provide details about the sample and outline how their characteristics relate to your research question and objectives. Increasingly such details are supported by a table or appendix in the project report, listing each case or element in the sample and providing sufficient detail to show how they relate to the research question while still preserving participants' anonymity (Box 7.11).



Box 7.11 **Focus on student research**

Reporting non-probability sample selection

Idris's research project was concerned with the development of collaborative relationships between university business schools and small and medium-sized enterprises. In his project report methodology he outlined his sample selection:

After obtaining ethical approval I selected purposively a maximum variation sample comprising three small and three medium-sized service sector

enterprises involved in collaborative research projects with the university. Subsequently 10 SME and 10 university knowledgeable participants were selected from those involved in these collaborations to provide diverse perspectives, 8 SME and all university participants agreeing to take part. In table 1 (below) informants are identified by an alphanumeric code. For SMEs this comprises the letter S (e.g. S1 for the first SME) followed by whether their involvement was as a director (d) or manager (m). University participants involved in these and other collaborative projects are identified by an alphanumeric code comprising the letter U and whether the involvement was as a student (s), academic (a) or professional (p) staff member; and where needed a number distinguished between them.

Table 1 Research participants*

| Code | No. of employees | SME's engagement with university/Informant's engagement with SMEs | Informant's role |
|-------------|-------------------------|--|-------------------------|
| S1d | 220 | Student consultancy projects | Managing Director |
| S1m | 220 | Student consultancy projects | Operations Director |
| S2d | 30 | Product development | Managing Director |
| : | : | : | : |
| Ua | n/a | Consultancy project module leader, supervise projects | Lecturer |
| Us | n/a | Group consultancy project member | Undergraduate student |
| Us | n/a | Group consultancy project member | Master's student |
| : | : | : | : |

*In Idris's actual project report all participants were listed in this table

7.9 Non-probability sampling – procedures

Alongside considering the likely sample size, you also need to select the most appropriate sampling procedure to enable you to answer your research question from the range of non-probability sampling procedures available (Figure 7.4). At one end of this range is quota sampling, which, like probability samples, tries to represent the total population. At the other end of this range is haphazard sampling, based on the need to obtain a sample as

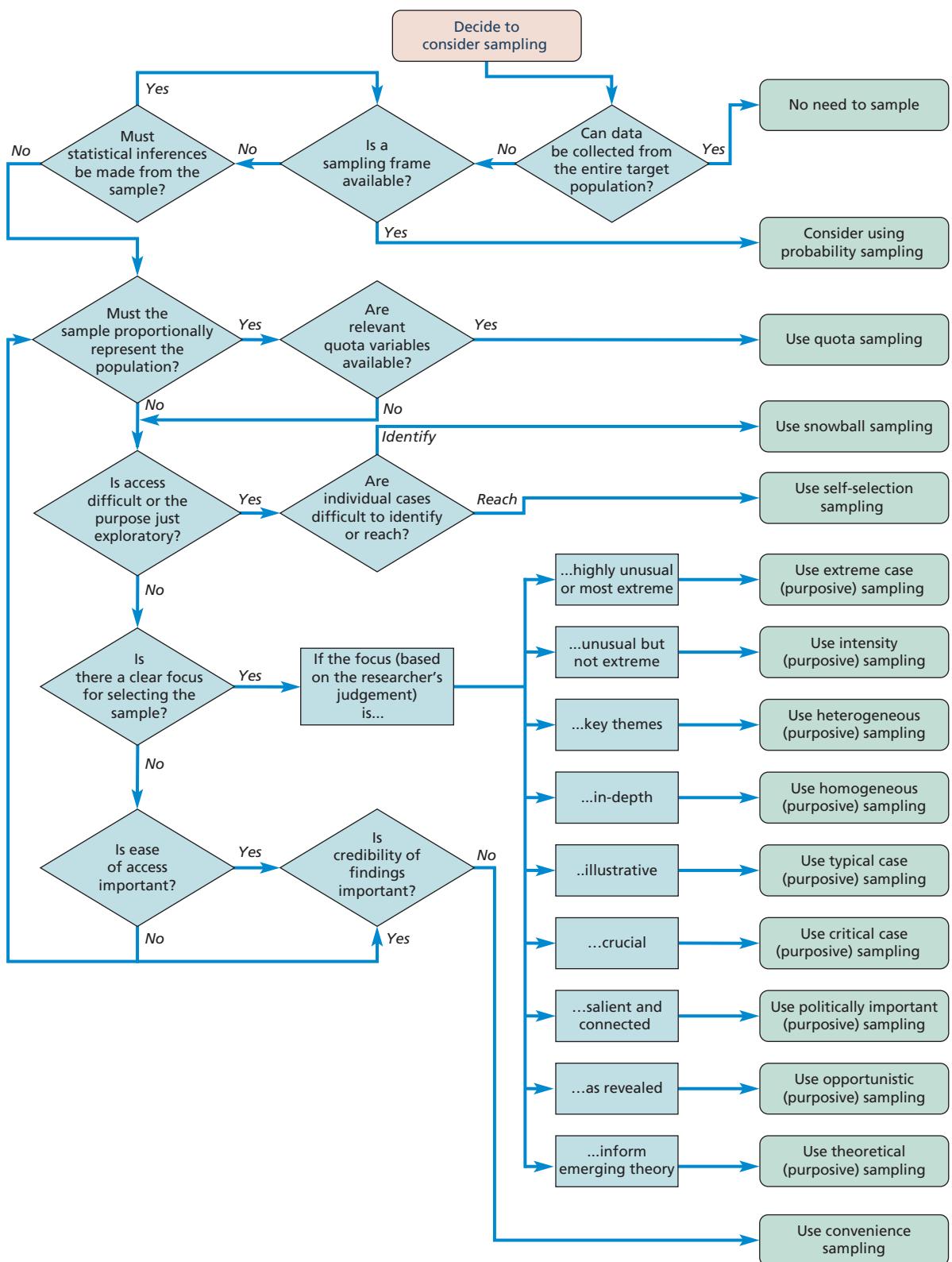


Figure 7.4 Selecting a non-probability sampling procedure

quickly as possible. With this technique you have virtually no control over the cases that will be included in your sample. Purposive sampling and volunteer sampling techniques lie between these extremes (Table 7.7). Where your research question has been derived from the literature and is reasonably well defined, the characteristics of potentially suitable participants are likely to be relatively easy to establish. However, where your research question is more emergent the sample characteristics will, at least in the early stages of data collection, be more broadly defined (Saunders and Townsend 2018).

Quota sampling

Quota sampling is entirely non-random and is often used as an alternative to probability sampling for Internet and interviewer completed questionnaires as part of a survey strategy where there is no sampling frame available. It is based on the premise that your sample will represent the target population as the variability in your sample for various quota variables is the same as that in the target population. However, this depends on the appropriateness of the assumptions on which the quota are based and having high quality data (Baker et al. 2013). Quota sampling has similar requirements for sample size as probabilistic sampling techniques (Section 7.5). To select a quota sample you:

- 1 Divide the population into specific groups.
- 2 Calculate a quota for each group based on relevant and available data.
- 3 Either:

Table 7.7 Impact of various factors on choice of non-probability sampling procedures

| Group | Procedure | Likelihood of sample being representative | Types of research in which useful | Relative costs | Control over sample contents |
|--------------------|---|--|--|--|---|
| Quota | Quota | Reasonable to high, although dependent on selection of quota variables | Where costs constrained or data needed very quickly so an alternative to probability sampling needed | Moderately high to reasonable | Specifies quota criteria |
| Purpo- sive | Extreme case | Low | Unusual or special to offer more revealing insights to explain the more typical | Reasonable | Specifies what is unusual or extreme |
| | Intensity | Low | Manifest phenomenon unusual but not highly unusual | Reasonable | Specifies what is unusual but not extreme |
| Heteroge- neous | Low, although dependent on researcher's choices | Reveal/illumi-nate key themes | Reasonable | Specifies criteria for maximum diversity | |

| Group | Procedure | Likelihood of sample being representative | Types of research in which useful | Relative costs | Control over sample contents |
|-----------|-----------------------|---|---|----------------|--|
| | Homogeneous | Low | In-depth exploration and reveal minor differences | Reasonable | Specifies criteria to identify particular group |
| | Typical case | Low, although dependent on researcher's choices | Illustrative | Reasonable | Specifies what is 'normal' |
| | Critical case | Low | Where focus is on importance | Reasonable | Specifies criteria as to what is important |
| | Politically important | Low | Where focus is on salience and connections | Reasonable | Specifies criteria re political importance |
| | Opportunistic | Low | Where unexpected occurs during research | Reasonable | Recognises and decides whether to take opportunity |
| | Theoretical | Low | Inform emerging theory | Reasonable | Specifies where to select initial participants and subsequent choice to inform emerging theory |
| Volunteer | Snowball | Low, but cases likely to have characteristics desired | Where cases difficult to identify | Reasonable | Selects only initial participant |
| | Self-selection | Low, as cases self-selected | Where access difficult, research exploratory | Reasonable | Offers general invitation |
| Haphazard | Convenience | Very low (often lacks credibility) | Ease of access | Low | Haphazard |

Sources: Developed from Patton (2015); Saunders and Townsend (2018)

- a (for Internet questionnaires) contract an online panel company specifying the number of cases in each quota from which completed questionnaires must be obtained; or
 - b (for interviewer-completed questionnaires) ensure that where multiple interviewers are used each interviewer has an ‘assignment’, which states the number of cases in each quota from which they must collect data.
- 4 Where necessary, combine the data collected to provide the full sample.

Quota sampling has a number of advantages over the probability sampling techniques. In particular, it is less costly and can be set up very quickly and does not require a sampling frame. If, as with television audience research surveys or opinion polls, your data collection needs to be undertaken very quickly then quota sampling may be the only possibility. In addition, it is relatively straightforward to specify quotas if collecting data using a panel company’s pre-screened volunteer panel. Quota sampling is normally used for large target populations. Decisions on sample size are governed by the need to have sufficient responses in each quota to enable subsequent statistical analyses to be undertaken. This often necessitates a sample size of between 2,000 and 5,000.

Calculations of quotas are based on relevant and available data and are usually relative to the proportions in which they occur in the population (Box 7.12). Without sensible and relevant quotas, data collected may be biased. For many market research projects, quotas are derived from census data. Your choice of quota is dependent on two main factors:

- usefulness as a means of stratifying the data;
- ability to overcome likely variations between groups in their availability for interview.

Where people who are retired are likely to have different opinions from those in work, a quota that does not ensure that these differences are captured may result in the data being biased as it would probably be easier to collect the data from those people who are



Box 7.12 Focus on student research

Devising a quota sample

Paolo was undertaking the data collection for his project as part of his full-time employment. For his research his employer had agreed to pay an online panel company to distribute his web questionnaire to a sample of people representing those aged 16–74 who were either economically active or inactive. No sampling frame was available. Once the data had been collected, he was going to disaggregate his findings into subgroups dependent on gender and whether they were economically active or economically inactive. Previous research had suggested that whether or not people were retired would also have an impact on responses and so he needed to make sure that those

surveyed in each group also reflected these people. Fortunately, his country’s national census of population contained a breakdown of the number of people who were economically active and inactive, their employment status and gender. These formed the basis of the categories for his quotas:

| | | | | |
|--------|---|-------------------|---|---|
| Gender | × | economic activity | × | employment status |
| male, | | active, | | part-time employee, |
| female | | inactive | | full-time employee, self-employed, unemployed, full-time student, retired, student, looking after home or family, long-term sick or disabled, other |

As he was going to analyse the data for economic activity and gender, it was important that each of these four groups (male and economically active, male and economically inactive, female and economically active, female and economically inactive) had sufficient

respondents (at least 30) to enable meaningful statistical analyses. Paolo calculated that a 0.00001 per cent quota (1 in 100,000) would provide sufficient numbers in each of these four groups. This gave him the following quotas:

| Gender | Economic activity | Employment status | Population | Quota |
|--------|-------------------|------------------------------|------------|-------|
| Male | Active | Part-time employee | 1 175 518 | 12 |
| | | Full-time employee | 9 013 615 | 90 |
| | | Self-employed | 2 670 662 | 27 |
| | | Unemployed | 1 015 551 | 10 |
| | | Full-time student | 619 267 | 6 |
| | Inactive | Retired | 2 270 916 | 22 |
| | | Student | 1 148 356 | 11 |
| | | Looking after home or family | 156 757 | 2 |
| | | Long-term sick or disabled | 823 553 | 8 |
| | | Other | 385 357 | 4 |
| Female | Active | Part-time employee | 4 158 750 | 42 |
| | | Full-time employee | 6 002 949 | 60 |
| | | Self-employed | 1 122 970 | 11 |
| | | Unemployed | 687 296 | 7 |
| | | Full-time student | 717 556 | 7 |
| | Inactive | Retired | 3 049 775 | 30 |
| | | Student | 1 107 475 | 11 |
| | | Looking after home or family | 1 538 377 | 15 |
| | | Long-term sick or disabled | 750 581 | 8 |
| | | Other | 467 093 | 5 |
| Total | | | 38 882 374 | 388 |

These were specified to the online panel company who were paid for each completed questionnaire received up to the number in each quota group.

retired. Quotas used in market research and opinion polls usually include measures of age, gender and economic activity or social class. These may be supplemented by additional quotas, dictated by the research question(s) and objectives (Box 7.12).

When you provide an online panel company the quota specification, they deliver your questionnaire to a ‘volunteer panel’ of potential respondents they have selected to meet your quota criteria. For online panel company data it is important to establish whether or not the online panel company offers panel members an incentive to encourage response and the likely implications of this for the characteristics of the respondents and, consequently, their responses (Section 11.2). Despite this being a non-probability sample, the number invited to complete a particular questionnaire and the number who do so are both

known. It is therefore possible to calculate a **participation rate** (American Association for Public Opinion Research, 2016):

$$\text{Participation rate} = \frac{\text{Number of respondents providing a usable response}}{\text{Number of respondents invited to participate}}$$

For data collected using more than one interviewer, assignments (completed questionnaires) from each interviewer are combined to provide the full sample. Because the interviewer can choose within quota boundaries whom they interview, your quota sample may be subject to bias. Interviewers tend to choose respondents who are easily accessible and who appear willing to answer the questions. Clear controls may therefore be needed. In addition, it has been known for interviewers to fill in quotas incorrectly. This is not to say that your quota sample will not produce good results; they can and often do! However, you cannot measure the level of certainty or margins of error as the sample is not probability based.

Purposive sampling

With **purposive sampling** you need to use your judgement to select cases that will best enable you to answer your research question(s) and to meet your objectives. For this reason it is sometimes known as **judgemental sampling**. You therefore need to think carefully about the impact of your decision to include or exclude cases on the research when selecting a sample in this way. Purposive sampling is often used when working with very small samples such as in case study research and when you wish to select cases that are particularly informative. A particular form of purposive sampling, theoretical sampling, is used by researchers adopting the grounded theory strategy (Section 5.5).

Purposive samples cannot be considered to be statistically representative of the target population. The logic on which you base your strategy for selecting cases for a purposive sample should be dependent on your research question(s) and objectives. Patton (2015) emphasises this point by contrasting the need to select information-rich cases in purposive sampling with the need to be statistically representative in probability sampling. The more common purposive sampling strategies were outlined in Table 7.7.

Extreme case sampling

Extreme case or **deviant sampling** focuses on highly unusual or special cases on the basis that the data collected about these highly unusual or extreme outcomes will enable you to learn the most and to answer your research question(s) and meet your objectives most effectively. This is often based on the premise that findings from extreme cases will be relevant in understanding or explaining more typical cases (Patton 2015).

Intensity sampling

Intensity sampling uses a similar justification to extreme case sampling; more can be learned from the unusual than from the normal. The sample contains those cases that richly reveal the phenomenon of interest but are not the most extreme or highly unusual. You might, for example, having explored the variation in sales staff performance, select those who have consistently performed well, rather than those with the best overall performance.

Heterogeneous sampling

Heterogeneous or **maximum variation sampling** uses your judgement to choose participants with sufficiently diverse characteristics to provide the maximum variation possible in the data collected. It enables you to collect data to describe and explain the key themes that

can be observed. Although this might appear a contradiction, as a small sample may contain cases that are completely different, Patton (2015) argues that this is in fact a strength. Any patterns that do emerge are likely to be of particular interest and value and represent the key themes. In addition, the data collected should enable you to document uniqueness. To ensure maximum variation within a sample, Patton (2015) suggests you identify your diverse characteristics (sample selection criteria) prior to selecting your sample.

Homogenous sampling

In direct contrast to heterogeneous sampling, **homogeneous sampling** focuses on one particular subgroup in which all the sample members are similar, such as a particular occupation or level in an organisation's hierarchy. Characteristics of the selected participants are similar, allowing them to be explored in greater depth and minor differences to be more apparent.

Typical case sampling

Typical case sampling is usually used as part of a research project to provide an illustrative profile using a representative case. Such a sample enables you to provide an illustration of what is 'typical' to those who will be reading your research report and may be unfamiliar with the subject matter. It is not intended to be definitive.

Critical case sampling

In contrast, **critical case sampling** selects critical cases on the basis that they can make a point dramatically or because they are important. The focus of data collection is to understand what is happening in each critical case so that logical generalisations can be made. Patton (2015) outlines a number of clues that suggest critical cases. These can be summarised by the questions such as:

- If it happens there, will it happen everywhere?
- If they are having problems, can you be sure that everyone will have problems?
- If they cannot understand the process, is it likely that no one will be able to understand the process?

Politically important sampling

Politically important sampling relies on your judgement regarding anticipated politically sensitive issues and associated outcomes when deciding whether to include one or a number of prominent potential participants. Consequently, you choose to include (or exclude) participants on the basis of their connections with politically sensitive issues (Miles et al. 2019).

Opportunistic sampling

Opportunistic sampling acknowledges how, particularly within qualitative research involving inductive theory building, unforeseen opportunities can occur. For example, new potential research participants may emerge requiring an on-the-spot decision about their fit with the research and their inclusion. As such it relies on you using your judgment as to recognise such opportunities and assess whether or not to take them (Miles et al. 2019).

Theoretical sampling

Theoretical sampling is a special case of purposive sampling, being particularly associated with grounded theory and analytic induction (Morse and Clark 2019) (Sections 13.9 and 13.8). Initially, you need to have some idea of where to sample, although not

necessarily what to sample for, participants being chosen as they are needed. Subsequent sample selection is dictated by the needs of the emerging theory and the evolving storyline, your participants being chosen purposively to inform this. A theoretical sample is therefore cumulatively chosen according to developing categories and emerging theory based upon your simultaneous collecting, coding and analysis of the data.

Volunteer sampling

Snowball sampling

Snowball sampling is the first of two procedures we look at where participants volunteer to be part of the research rather than being chosen. It is used commonly when it is difficult to identify members of the desired population (Box 7.10); for example, people who are working while claiming unemployment benefit. You, therefore, need to:

- 1 Make contact with one or two cases.
- 2 Ask these cases to identify further cases.
- 3 Ask these new cases to identify further new cases (and so on).
- 4 Stop when either no new cases are given or the sample is as large as is manageable or data saturation has been reached.

The main problem is making initial contact. Once you have done this, these cases identify further members of the population, who then identify further members, and so the sample grows like a snowball being rolled in snow. For such samples the problems of bias are huge, as respondents are most likely to identify other potential respondents who are similar to themselves, resulting in a homogeneous sample (Lee 2000). The next problem is to find these new cases. However, for populations that are difficult to identify, snowball sampling may provide the only possibility.

A development of snowball sampling is **respondent driven sampling (RDS)**. This combines snowball sampling with the use of coupons or some other method to track the identification of further cases and statistical modelling to compensate for the sample being collected in a non-random way. This can enable researchers to make unbiased estimates of their target population (Baker et al. 2013).

Self-selection sampling

Self-selection sampling is the second of the volunteer sampling procedures we look at. It occurs when you allow each case, usually individuals, to identify their desire to take part in the research (Box 5.6, 7.10). You therefore:

- 1 Publicise your need for cases, either by advertising through appropriate media or by asking them to take part.
- 2 Collect data from those who respond.

Publicity for volunteer samples can take many forms. These include articles and advertisements in magazines that the population are likely to read, postings on appropriate online newsgroups and discussion groups, hyperlinks from other websites as well as letters, emails or tweets of invitation to colleagues and friends (Box 7.13). Cases that self-select often do so because of their strong feelings or opinions about the research question(s) or stated objectives. In some instances, this is exactly what the researcher requires to answer her or his research question and meet the objectives.



Box 7.13 Focus on student research

Self-selection sampling

Siân's research was concerned with the impact of student loans on studying habits. She had decided to

distribute her questionnaire using the Internet. She publicised her research on Facebook in a number of groups' pages, using the associated description to invite people to self-select and click on the link to the questionnaire. Those who self-selected by clicking on the hyperlink were automatically taken to the web questionnaire she had developed using the SurveyMonkey.com online survey software.

Haphazard sampling

Haphazard sampling occurs when sample cases are selected without any obvious principles of organisation in relation to your research question, the most common form being **convenience sampling** (also known as **availability sampling**). This involves selecting cases haphazardly only because they are easily available (or most convenient) to obtain for your sample, such as the person interviewed at random in a shopping centre for a television programme 'vox pop'. Although convenience sampling is used widely (for example, Facebook polls or questions), it is prone to many sources of bias and influences that are beyond your control. Cases appear in the sample only because of the ease of obtaining them; consequently, all you can do is make some statement about the people who felt strongly enough about the subject of your question to answer it (and were using Facebook) during the period your poll was available!

Not surprisingly, as emphasised in Table 7.7, findings from convenience samples may be given very little credibility. Despite this, samples ostensibly chosen for convenience often meet purposive sample selection criteria that are relevant to the research aim (Saunders and Townsend 2018). It may be that an organisation you intend to use as a case study is 'convenient' because you have been able to negotiate access through existing contacts. Where this organisation also represents an 'extreme' case, it can also offer insights about the unusual or extreme, providing justification regarding its purpose when addressing the research aim. Alternatively, while a sample of operatives in another division of an organisation for which you work might be easy to obtain and consequently 'convenient', the fact that such participants allow you to address a research aim necessitating an in-depth focus on a particular homogenous group is more crucial.

Where the reasons for using a convenience sample have little, if any, relevance to the research aim, participants appear in the sample only because of the ease of obtaining them. While this may not be problematic if there is little variation in the target population, where the target population is more varied it can result in participants that are of limited use in relation to the research question. Often a sample is intended to represent more than the target population, for example, managers taking a part-time MBA course as a surrogate for all managers. In such instances the selection of individual cases may introduce bias to the sample, meaning that subsequent interpretations must be treated with caution.

7.10 Mixed and multi-stage sampling designs

Both mixed and multi-stage sampling designs use different sampling procedures within a research project.



Box 7.14 Focus on management research

Mixing different sampling procedures

Rinken and colleagues (Rinken et al. 2020) undertook a web survey to collect data on the COVID pandemic's social dimensions using mixed sampling. Their paper in *Survey Research Methods* focuses on the two complementary sampling procedures adopted.

The first sample was generated by sending random number text-messaging (SMS) invitations to potential Spanish respondents' mobile phones. This sampling frame comprised a possible number of over 93 million, of which 47.1 per cent were active at the time of the research and they argue emulates probability sampling (using simple random sampling). A total of 51,046

SMS messages were sent resulting in a response rate of 5.6 per cent, nearly double that expected, and a sample size of 1,379. Within this there was under-representation in some age groups and of people with low levels of formal education.

The second sample was recruited subsequently using paid advertisements on Facebook and Instagram, focusing entirely on people with basic education (low levels of formal education) and aged either under 30 years or over 65 years. This was a non-probability volunteer sample in which respondents self-selected. The advertisement was displayed 1,337,856 times to 1,187,580 people being clicked on by 3,752 different individuals, 944 who completed the questionnaire. Although the overall response rate was ultra-low (0.08 per cent) the click to response rate was 26 per cent.

Rinken et al. (2020) note their two-pronged sampling procedure provided less biased data than either of the two components would have done on a standalone basis.

Mixed sampling

In a **mixed sampling** design, data are collected from two or more discrete samples, each discrete collection using either probability or non-probability sample selection procedures. You might, for example, combine data in a single research project from two different samples; one generated using probability sampling procedure such as random sampling, and the other being a non-probability procedure such as volunteer sampling (Box 7.14). Alternatively, you might use two data collected from two discrete probability samples from different organisations suppliers, both collected using systematic random sampling.

Multi-stage sampling

In contrast, **multi-stage sampling** refers to any sampling design that occurs in two or more successive stages using either probability, non-probability or both types of sample selection procedures. In the first stage you might select two organisations using critical case purposive sampling. Subsequently, in the second stage, you might select a sample of employees from each organisation using stratified random sampling, thereby combining non-probability with probability sampling. Alternatively, you may first select a large sample of customers or organisations using quota sampling. Subsequently, based on analysis of the data collected from these customers or organisations, you may select a smaller heterogeneous purposive sample to illustrate the key themes. Multi-stage sampling can also use cluster sampling to overcome problems associated with a geographically dispersed population when face-to-face contact is needed, or when it is expensive and time consuming to construct a sampling frame for a large geographical area (Box 7.15).



Box 7.15 Focus on student research

Multi-stage sampling

Laura worked for a market research organisation that needed her to interview a sample of 400 households in England and Wales face-to-face. She decided to use the electoral register as a sampling frame. Laura knew that selecting 400 households using either systematic or simple random sampling was likely to result in these 400 households being dispersed throughout England and Wales, resulting in considerable amounts of time spent travelling between interviewees as well as high travel costs. By using multi-stage sampling Laura felt these problems could be overcome.

In her first stage the geographical area (England and Wales) was split into discrete sub-areas (counties). These formed her sampling frame. After numbering all the counties, Laura selected a small number

of counties at random using cluster sampling. Since each case (household) was located in a county, each had an equal chance of being selected for the final sample.

As the counties selected were still too geographically large, each was subdivided into smaller geographically discrete areas (electoral wards). These formed the next sampling frame (stage 2). Laura selected another sample at random. This time she selected a larger number of wards using simple random sampling to allow for likely important variations in the nature of households between wards.

A sampling frame of the households in each of these wards was then generated. Laura purchased copies of the edited electoral register from the relevant local authorities. These contained the names and addresses of people who had registered to vote and had not 'opted out' of allowing their details to be made widely available for others to use. Laura finally selected the actual cases (households) that she would interview using systematic random sampling.

Where multi-stage sampling uses one or more probability sampling procedures, you need to ensure that the sampling frames are appropriate and available. In order to minimise the impact of selecting smaller and smaller subgroups on the representativeness of your sample, you can apply stratified random sampling procedures (Section 7.6). This can be further refined to take account of the relative size of the subgroups by adjusting the sample size for each subgroup. As you have selected your sub-areas using different sampling frames, you only need a sampling frame that lists all the members of the population for those subgroups you finally select (Box 7.15). This provides considerable savings in time and money.

7.11 Summary

- Sampling is used when it is not feasible or sensible to collect data from every possible case or group member to answer your research question(s) and to address your objectives.
- Choice of sampling procedure or procedures is dependent on your research question(s) and objectives:
 - Research question(s) and objectives that need you to estimate statistically the characteristics of the target population from a sample nearly always require probability samples. (When considering probability sampling for a population of 50 or fewer, it is usually sensible to collect data from the entire population.)
 - Research question(s) and objectives that do not require such statistical generalisations can, alternatively, make use of non-probability sampling procedures.

- Probability sampling procedures all necessitate some form of sampling frame.
- Where it is not possible to obtain or devise a sampling frame you will need to use non-probability sampling procedures.
- Non-probability sampling procedures provide the opportunity to select your sample purposively and to also reach difficult-to-identify members of the target population.
- The size of probability samples selected to address research questions that require statistical estimation should be calculated. It is dependent upon the target population and the margin of error and confidence level required. Statistical analyses usually require a minimum sample size of 30.
- The size for non-probability samples selected to address research questions that do not require statistical estimation is dependent upon the research question and objectives, what will allow you to develop a fruitful analytic argument and what will be credible. While guidance suggests between 15 and 60 interviews is likely to be sufficient to reach saturation and be credible, for some research purposes a sample of one can be sufficient and credible.
- Sample size and the procedure used are also influenced by the availability of resources, in particular financial support and time available to select the sample and to collect, input and analyse the data.
- For some research projects you will need to combine multiple discrete samples or use multi-stage sampling, perhaps combining probability and non-probability procedures.
- Your sampling choices will, for some research questions, be dependent on your ability to gain access to organisations or groups. The considerations summarised earlier must therefore be tempered with an understanding of what is practically possible.

Self-check questions

Help with these questions is available at the end of the chapter.

- 7.1** Identify a suitable sampling frame for each of the following research questions.
 - a** How do company directors of manufacturing firms of over 500 employees think a specified piece of legislation will affect their companies?
 - b** Which factors are important in accountants' decisions regarding working overseas?
 - c** How do employees at Best Plumbing Ltd feel about the company's proposal to make a current Covid-19 vaccination certificate compulsory for all employees?
- 7.2** Lisa has emailed her tutor with the following query regarding sampling and dealing with non-response. Imagine you are Lisa's tutor. Draft a reply to answer her query.

HELP!!!! Sampling non-response

Hi,

I hope you are well.

I interviewed someone yesterday and I (almost) failed to get him to say anything useful for my research project. This was strange as he had appeared to have really useful background. I was unable to get him to reflect on the issue of inhibitors of spin out companies in the *light of his own experiences or provide actual examples*. He clearly wanted to talk about what he wanted to and had decided this before the interview and I asked my questions 😢. He obviously thought there were right and wrong answers and he was telling me what he thought he should say rather than giving me the actual examples I asked for. This meant he did not talk about feedback loops, linkages or ideas, which you know is what my research is about. My attempts to get the conversation onto my research were gently, but firmly, put aside.

My question is: **Can I just exclude this interview from my sample?** He was a great person and I really enjoyed meeting him. However, because I could not get him to answer my questions, the interview did not yield any insights. What should I do?

With best wishes

Lisa



- 7.3** You have been asked to select a sample of manufacturing firms using the sampling frame below. This also lists the value of their annual output in tens of thousands of pounds over the past year. To help you in selecting your sample the firms have been numbered from 1 to 100.
- a Select two simple random samples, each of 20 firms, and mark those firms selected for each sample on the sampling frame.
 - b Describe and compare the pattern on the sampling frame of each of the samples selected.
 - c Calculate the average (mean) annual output in tens of thousands of pounds over the past year for each of the samples selected.
 - d Given that the true average annual output is £6,608,900, is there any bias in either of the samples selected?

| | Output | | Output | | Output | | Output | | Output | |
|----|---------------|----|---------------|----|---------------|----|---------------|-----|---------------|--|
| 1 | 10 | 21 | 7 | 41 | 29 | 61 | 39 | 81 | 55 | |
| 2 | 57 | 22 | 92 | 42 | 84 | 62 | 73 | 82 | 66 | |
| 3 | 149 | 23 | 105 | 43 | 97 | 63 | 161 | 83 | 165 | |
| 4 | 205 | 24 | 157 | 44 | 265 | 64 | 275 | 84 | 301 | |
| 5 | 163 | 25 | 214 | 45 | 187 | 65 | 170 | 85 | 161 | |
| 6 | 1359 | 26 | 1440 | 46 | 1872 | 66 | 1598 | 86 | 1341 | |
| 7 | 330 | 27 | 390 | 47 | 454 | 67 | 378 | 87 | 431 | |
| 8 | 2097 | 28 | 1935 | 48 | 1822 | 68 | 1634 | 88 | 1756 | |
| 9 | 1059 | 29 | 998 | 49 | 1091 | 69 | 1101 | 89 | 907 | |
| 10 | 1037 | 30 | 1298 | 50 | 1251 | 70 | 1070 | 90 | 1158 | |
| 11 | 59 | 31 | 10 | 51 | 9 | 71 | 37 | 91 | 27 | |
| 12 | 68 | 32 | 70 | 52 | 93 | 72 | 88 | 92 | 66 | |
| 13 | 166 | 33 | 159 | 53 | 103 | 73 | 102 | 93 | 147 | |
| 14 | 302 | 34 | 276 | 54 | 264 | 74 | 157 | 94 | 203 | |
| 15 | 161 | 35 | 215 | 55 | 189 | 75 | 168 | 95 | 163 | |
| 16 | 1298 | 36 | 1450 | 56 | 1862 | 76 | 1602 | 96 | 1339 | |
| 17 | 329 | 37 | 387 | 57 | 449 | 77 | 381 | 97 | 429 | |
| 18 | 2103 | 38 | 1934 | 58 | 1799 | 78 | 1598 | 98 | 1760 | |
| 19 | 1061 | 39 | 1000 | 59 | 1089 | 79 | 1099 | 99 | 898 | |
| 20 | 1163 | 40 | 1072 | 60 | 1257 | 80 | 1300 | 100 | 1034 | |

- 7.4** You have been asked to select a 10 per cent sample of firms from the sampling frame used for self-check question 7.3.
- a Select a 10 per cent systematic random sample and mark those firms selected for the sample on the sampling frame.
 - b Calculate the average (mean) annual output in tens of thousands of pounds over the past year for your sample.
 - c Given that the true average annual output is £6,608,900, why does systematic random sampling provide such a poor estimate of the annual output in this case?
- 7.5** You decide to collect data using a web questionnaire from managing directors of small-to medium-sized organisations. From the data you collect you need to be able to generalise about the attitude of such managing directors to recent changes in government policy towards these firms. Your generalisations need to be accurate to within plus or minus 5 per cent.
- a How many managing directors will you need to collect data from?
 - b Assuming a response rate of 5 per cent, how many questionnaires will you need to distribute?

- 7.6** You have been asked to use face-to-face questionnaires to collect data from local residents about their opinions regarding the siting of a new supermarket in an inner-city suburb (estimated catchment population 111,376 at the last census). The age and gender distribution of the catchment population at the last census is listed below.

| Gender | Age group | | | | | | | |
|---------|-----------|------|-------|-------|-------|------------|-----------|------|
| | 0–4 | 5–15 | 16–19 | 20–29 | 30–44 | 45–59 /64* | 60/65#–74 | 75 + |
| Males | 3498 | 7106 | 4884 | 7656 | 9812 | 12892 | 4972 | 2684 |
| Females | 3461 | 6923 | 6952 | 9460 | 8152 | 9152 | 9284 | 4488 |

*59 females, 64 males; 60 females, 65 males.

- a** Devise a quota for a quota sample using these data.
b What other data would you like to include to overcome likely variations between groups in their availability for interview and replicate the target population more precisely? Give reasons for your answer.
c What problems might you encounter in using interviewers?
- 7.7** For each of the following research questions it has not been possible for you to obtain a sampling frame. Suggest the most suitable non-probability sampling technique to obtain the necessary data, giving reasons for your choice.
- a** What support do people sleeping rough believe they require from social services?
b Which television advertisements do people remember watching last weekend?
c How do employers' opinions vary regarding the presence of institutional racism in the organisations in which they work?
d How are manufacturing companies planning to respond to the introduction of road tolls?
e Would users of the squash club be prepared to pay a 10 per cent increase in subscriptions to help fund two extra courts (answer needed by tomorrow morning!)?

Review and discussion questions

- 7.8** With a friend or colleague choose one of the following research questions (or one of your own) in which you are interested.
- What attributes attract people to jobs?
 - How are exporting organisations adapting their processes following the United Kingdom leaving the European Union?
- Use the flow charts for both probability sampling (Figure 7.3) and non-probability sampling (Figure 7.4) to decide how you could use each type of sampling independently to answer the research question.
- 7.9** Agree with a colleague to watch a particular documentary or consumer rights programme on the television or a streaming service. If possible, choose a documentary with a business or management focus. During the documentary, pay special attention to the samples from which the data for the documentary are drawn. Where possible, note down details of the sample such as who were interviewed or who responded to questionnaires and the reasons why these people were chosen. Where this is not possible, make a note of the information you would have liked to have been given. Discuss your findings with your colleague and come to a conclusion regarding the nature of the sample used, its representativeness and the extent to which it was possible for the programme maker to generalise from that sample.

- 7.10** Access online or obtain a copy of a quality daily newspaper and, within the newspaper, find an article that discusses a 'survey' or 'poll'. Share the article with a friend. Make notes of the process used to select the sample for the 'survey' or 'poll'. As you make your notes, note down any areas where you feel there is insufficient information to fully understand the sampling process. Aspects for which information may be lacking include the target population, size of sample, how the sample was selected, representativeness and so on. Discuss your findings with your friend.



Progressing your research project

Using sampling as part of your research

- Consider your research question(s) and objectives. You need to decide whether you will be able to collect data on the entire population or will need to collect data from a sample.
- If you decide that you need to sample, establish whether your research question(s) and objectives require probability sampling. If they do, make sure that a suitable sampling frame is available or can be devised, and calculate the actual sample size required, taking into account likely response rates.
- If your research question(s) and objectives do not require probability sampling, or you are unable to

obtain a suitable sampling frame, you will need to use non-probability sampling. Where appropriate, estimate the sample size you are likely to need to develop a fruitful analytic argument.

- Select the most appropriate sampling procedure or procedures after considering the advantages and disadvantages of all suitable procedures and undertaking further reading as necessary.
- Select your sample or samples following the procedure or procedures as outlined in this chapter.
- Remember to note down your choices and the reasons for these as you make them, as you will need to justify your choices when you write about your research method (Boxes 7.5 and 7.11 may be helpful here).
- Use the questions in Box 1.4 to guide your reflective diary entry.

References

- American Association for Public Opinion Research (2016) *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys* (9th edn). Lenexa, KA: AAPOR. Available at http://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf [Accessed 11 May 2017].
- Baker, R., Brick, J.M., Bates, N.A., Battaglia, M., Couper, M.P., Dever, J.A., Gile, K.J. and Tourangeau, R. (2013) 'Summary of the AAPOR task force on non-probability sampling', *Journal of Survey Statistics and Computing*, Vol. 1, pp. 90–143.
- Baruch, Y. and Holtom, B.C. (2008) 'Survey response rate levels and trends in organizational research', *Human Relations*, Vol. 61, pp. 1139–60.
- Becker, H.S. (1998) *Tricks of the Trade: How to Think About Your Research While You're Doing It*. Chicago: Chicago University Press.
- Bradley, N. (1999) 'Sampling for Internet surveys: An examination of respondent selection for Internet research', *Journal of the Market Research Society*, Vol. 41, No. 4, pp. 387–95.
- Chidlow, A., Ghauri, P.N., Yeniyurt, S. and Cavusgil, S.T. (2015) 'Establishing rigour in mail-survey procedures in international business research' *Journal of World Business*, Vol. 50, pp. 26–35.

- Committee of Advertising Practice (2010) *The BCAP Code: The UK Code of broadcast Advertising (Edition 1)*. London: Committee of Advertising Practice. Available at <https://www.asa.org.uk/asset/846F25EB-F474-47C1-AB3FF571E3DB5910/> [Accessed 2 August 2021].
- Committee of Advertising Practice (2014) *The CAP Code: The UK Code of Non-broadcast Advertising and Direct & Promotional Marketing (Edition 12)*. London: Committee of Advertising Practice. Available at <https://www.asa.org.uk/asset/47EB51E7-028D-4509-AB3C0F4822C9A3C4/> [Accessed 2 August 2021].
- Committee of Advertising Practice (2021) 'A quick guide to advertising consumer surveys' CAP News. Available at <https://www.asa.org.uk/news/a-quick-guide-to-advertising-consumer-surveys.html> [Accessed 2 August 2021].
- De Vaus, D.A. (2014) *Surveys in Social Research* (6th edn). Abingdon: Routledge.
- Dillman, D.A., Smyth, J.D. and Christian, J.M. (2014) *Internet, Phone, Mail and Mixed Mode Surveys: The Tailored Design Method* (4th edn). Hoboken, NJ: Wiley.
- Edwards, T., Tregaskis, O., Edwards, P., Ferner, A., Marginson, A. with Arrowsmith, J., Adam, D., Meyer, M. and Budjanovcanin, A. (2007) 'Charting the contours of multinationals in Britain: Methodological challenges arising in survey-based research', *Warwick Papers in Industrial Relations*, No. 86. Available at http://www.cbs.dk/files/cbs.dk/charting_the_contours_of_multinationals_in_britain.pdf [Accessed 14 May 2017].
- Ellis, P.D. (2010) *The Essential Guide to Effect Sizes*. Cambridge: Cambridge University Press.
- Francis, J.J., Johnston, M., Robertson, C., Glidewell, L., Entwistle, V., Eccles, M. and Grimshaw, M. (2010) 'What is an adequate sample size? Operationalizing data saturation for theory-based studies', *Psychology and Health*, Vol. 25, pp. 1229–45.
- Goods, C., Veen, A. and Barratt, T. (2019). "Is your gig any good?" Analysing job quality in the Australian platform-based food-delivery sector. *Journal of Industrial Relations*, Vol 61, No. 4, pp. 502–527.
- Groves, R.M. and Peytcheva, E. (2008) 'The impact of nonresponse rates on nonresponse bias', *Public Opinion Quarterly*, Vol. 72, No. 2, pp. 167–89.
- Guest, G., Bunce, A. and Johnson, L. (2006) 'How many interviews are enough? An experiment with data saturation and validity', *Field Methods*, Vol. 18, No. 1, pp. 59–82.
- Hammersley, M. (2015) 'Sampling and thematic analysis: A response to Fugard and Potts', *International Journal of Social Research Methodology*, Vol. 18, No. 6, pp. 687–8.
- Henry, G.T. (1990) *Practical Sampling*. Newbury Park, CA: Sage.
- Latpate, R., Kshirsagar, J., Gupta, V.K. and Chandra, G. (2021) *Advanced Sampling Methods*. Singapore: Springer.
- Lee, B. and Saunders, M.N.K (2017) *Doing Case Study Research for Business and Management Students*. London: Sage.
- Lee, R.M. (2000) *Doing Research on Sensitive Topics*. London: Sage.
- Mellahi, K. and Harris, L.C. (2016) 'Response rates in business and management research: An overview of current practice and suggestions for future direction', *British Journal of Management*, Vol. 27, No. 2, pp. 426–37.
- Miles, M.B., Huberman, A.M. and Saldaña, J (2019) *Qualitative Data Analysis: A Methods Sourcebook* (4th edn). Thousand Oaks, CA: Sage.
- Morse, J.M. and Clark, L. (2019) 'The nuances of grounded theory sampling and the pivotal role of theoretical sampling', In A. Bryant and K. Charmaz (eds.) *The Sage Handbook of Current Developments in Grounded Theory*. London: Sage, pp. 145–66.
- Neuman, W.L. (2014) *Social Research Methods* (7th edn). Harlow: Pearson.
- O'Reilly, M. and Parker, N. (2013) 'Unsatisfactory saturation: A critical exploration of the notion of the notion of saturated sample sizes in qualitative research', *Qualitative Research*. 13, pp. 190–97.
- Parry, E., Farndale, E., Brewster, C. and Morley, C. (2021) 'Balancing rigour and relevance: The case of methodological pragmatism in conducting large scale, multi-country and comparative management studies' *British Journal of Management*, Vol. 32, No. 2, pp. 273–82.

- Patton, M.Q. (2015) *Qualitative Research and Evaluation Methods: Integrating Theory and Practice* (4th edn). Thousand Oaks, CA: Sage.
- Pielsticker, D.I. and Hiebl, M.R.W. (2020) 'Survey response rates in family business research' *European Management Review*, Vol. 17, pp. 327–46.
- Rinken, S., Domínguez, J.-A., Trujillo, M., Lafuente, R., Sotomayer, R. and Serrano-de-Rosal, R. (2020) 'Combined mobile-phone and social-media sampling for web survey on social effects of Covid-19 in Spain', *Social Research Methods*, Vol. 14, No. 2, pp. 165–70.
- Rogelberg, S.G. and Stanton, J.M. (2007) 'Introduction: Understanding and dealing with organizational survey non-response', *Organizational Research Methods*, Vol. 10, No. 2, pp. 195–209.
- Saunders, M.N.K. (2012) 'Choosing research participants', in G. Symons and C. Cassell (eds) *The Practice of Qualitative Organizational Research: Core Methods and Current Challenges*. London: Sage, pp. 37–55.
- Saunders, M.N.K., Gray, D.E. and Bristow, A. (2017) 'Beyond the single organization: Insider insights from gaining access for large multiorganization survey HRD research', *Human Resource Development Quarterly*, Vol. 28, No. 3, pp. 401–425.
- Saunders, M.N.K. and Townsend, K. (2016) 'Reporting and justifying the number of participants in organisation and workplace research', *British Journal of Management*. Vol. 27, pp. 837–52.
- Saunders M.N.K. and Townsend, K. (2018) 'Choosing participants', in C. Cassell, A. Cunliffe and G. Grandy (eds), *Sage Handbook of Qualitative Business and Management Research Methods*. London: Sage, pp. 480–94.
- Sim, J., Saunders, B., Waterfield, J. and Kingstone, T. (2018) 'Can sample size in qualitative research be determined a priori?' *International Journal of Social Research Methodology*, Vol. 21, No. 5, pp. 619–34.
- Stutely, R. (2014) *The Economist Numbers Guide: The Essentials of Business Numeracy* (6th edn). London: Profile Books.
- Tucker, C. and Lepkowski, J.M. (2008) 'Telephone survey methods: Adapting to change', in J.M. Lepkowski, C. Tucker, J.M. Brick, E.D. De Leeuw, L. Japec, P.J. Lavrakas, M.W. Link and R.L. Sangster (eds), *Advances in Telephone Survey Methodology*. Hoboken, NJ: Wiley, pp. 3–28.

Further reading

- Baruch, Y. and Holtom, B.C. (2008) 'Survey response rate levels and trends in organizational research', *Human Relations*, Vol. 61, pp. 1139–60. This examines 490 academic studies using surveys published in 2000 and 2005 covering 100,000 organisations and over 400,000 individual respondents. The paper suggests likely response rates for different types of study and offers useful advice for reporting response rates.
- Latpate, R., Kshirsagar, J., Gupta, V.K. and Chandra, G. (2021) *Advanced sampling methods*. Singapore: Springer. This provides a detailed discussion of probability sampling procedures as well as non-sampling errors.
- Patton, M.Q. (2015) *Qualitative Research and Evaluation Methods: Integrating Theory and Practice* (4th edn). Thousand Oaks, CA: Sage. Chapter 5, 'Qualitative designs and data collection', contains a useful discussion of non-probability sampling techniques, with examples.
- Saunders, M.N.K. and Townsend, K. (2016) 'Reporting and justifying the number of participants in organisation and workplace research', *British Journal of Management*, Vol. 27, pp. 837–52. In addition to summarising the literature on sample size for interviewing, this examines sample selection practice and reporting for 248 academic studies using interviews published in 2003 and 2013. The paper suggests likely sample sizes for different types of interview study and offers useful advice for justifying sample size.

Collecting Primary Data Using Questionnaires

Saunders et al. (2023; 506-563)

Chapter 11



Collecting primary data using questionnaires

Learning outcomes

By the end of this chapter, you should be able to:

- identify the advantages and disadvantages of using questionnaires to collect data;
- distinguish between a range of self-completed (online, SMS, postal, delivery and collection) and researcher-completed (telephone, face-to-face) questionnaires;
- identify the different types of data questions collect;
- outline ways of establishing validity and testing reliability;
- select, justify and operationalise the use of appropriate questionnaire design and distribution methods to answer research questions and to meet objectives;
- justify pilot testing a questionnaire;
- take appropriate action to enhance response rates and to ensure the validity and reliability of the data collected;
- progress your research project by collecting data using a questionnaire.

11.1 Introduction

Within business and management research, the greatest use of questionnaires is made within the survey strategy (Section 5.5). However, both experiment and case study research strategies can make use of these methods. Although you probably have your own understanding of the term ‘questionnaire’, it is worth noting that there are a variety of definitions. Some people reserve it exclusively for questionnaires where the person answering the questions records their own answers, when it is **self-completed**. Others use it as a more general term to include interviews in which precisely the same set of questions are asked, and the respondent’s answers are recorded by the researcher.

In this book we use **questionnaire** as a general term to include all procedures for data collection in which each person is asked to respond to the same set of questions in a



predetermined order (De Vaus 2014). An alternative term, which is also widely used, is **instrument** (Ekinci 2015). The term therefore includes both questionnaires where questions are answered without a researcher being present (self-completed) and those where the researcher completes the questionnaire for the respondent. An overview of questionnaires that fall within this broad heading are outlined in the next section (11.2), along with their attributes and relative advantages and disadvantages for different completion modes and questionnaire mediums.

Please rate your experience . . .

Questionnaires are a part of our everyday lives. For modules in your course, your university has probably asked you and your fellow students to complete online module evaluation questionnaires, thereby collecting data on students' views. Similarly, when we visit a tourist attraction, have a meal in a restaurant or travel by air, there is often the opportunity to complete a visitor feedback form, comment card or passenger survey. Airlines are no exception, wanting to collect data from

their passengers so they can enhance their customers' experiences. While on a flight, and normally as the plane is nearing the destination, each passenger is asked via the aircraft's inflight entertainment system if they would be willing to answer a few questions about their experiences. If a passenger is willing, they then click on the 'passenger survey' icon displayed on their seat back screen and the first of the questions appears. Subsequently, they can rate their experiences by answering a series of closed (choice) questions using the plane's inflight entertainment system. The



Source: Matej Kastelic/Shutterstock



questionnaire starts with a brief introduction emphasising the importance of passengers' opinions in helping the Airline to improve:

Here at [Airline Name] we are dedicated to the continual improvement of our services and to the airline itself. To assist us in achieving this and to be in with a chance of winning 10,000 air miles we would be grateful if you could tell us what you thought of your experience flying with us today – thank you.

This is followed by closed questions such as those given below. Other topics about which questions are often asked include the service given by the cabin crew, the quality of the inflight entertainment system and the overall value for money of the airline. Personal details are also usually collected from each passenger, including their name, gender, age, country of origin and email address; passengers are informed that this will enable the airline to contact them, if they win the prize.

How did you check-in for your flight?

Online

Check-in counter

Kiosk

Please rate your check-in experience for each of the following:

| | Excellent | Very good | Good | OK |
|---|--------------------------|--------------------------|--------------------------|--------------------------|
| Ease of finding the check-in area | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Waiting time in queue | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Politeness of check-in staff | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Knowledge and helpfulness of check-in staff | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |

The use of questionnaires is discussed in many research methods texts. These range from those that devote a few pages to it, to those that specify precisely how you should construct and use them, such as Dillman et al.'s (2014) **tailored design method**. Because each person (respondent) is asked to respond to the same set of questions, the questionnaire provides an efficient way of collecting responses from a large sample prior to quantitative analysis (Chapter 12). However, before you decide to use a questionnaire, we should like to include a note of caution. Many authors (for example, Bell and Waters 2018) argue that it is far harder to produce a good questionnaire than you might think. You need to ensure that it will collect the precise data that you require to answer your research question(s) and achieve your objectives. This is of paramount importance because, like when the airline in the vignette is asking passengers to complete a questionnaire, you are unlikely to have more than one opportunity to collect the data. In particular, you will be unable to go back to those individuals who choose to remain anonymous and collect additional data using another questionnaire. These, and other issues, are discussed in Section 11.3.

The design of your questionnaire will affect the response rate and the validity and reliability of the data you collect (Section 11.4). These, along with response rates, can be maximised by:

- careful design of individual questions;
- clear and pleasing visual presentation;
- lucid explanation of the purpose;
- pilot testing;
- carefully and appropriately planned and executed delivery and return of completed questionnaires.

Our discussion of these aspects forms Sections 11.5 through to 11.8. In Section 11.5 we discuss designing individual questions, translating them into other languages and question coding. Constructing the questionnaire is discussed in Section 11.6 and pilot testing it in Section 11.7. Operationalising the distribution of the questionnaire is considered in Section 11.8 along with actions to help ensure high response rates.

11.2 Questionnaires: an overview

When to use questionnaires

We have found that many people use a questionnaire to collect data without considering other methods such as examination of archive and secondary sources (Chapter 8), observation (Chapter 9) and semi-structured or unstructured interviews (Chapter 10). Our advice is to evaluate all possible data collection methods and to choose those most appropriate to your research question(s) and objectives. Questionnaires are usually not particularly good for exploratory or other research that requires large numbers of open-ended questions (Section 10.2). They work best with standardised questions that you can be confident will be interpreted the same way by all respondents (Robson and McCartan 2016).

Questionnaires tend to be used for descriptive or explanatory research. Descriptive research, such as that undertaken using attitude and opinion questionnaires and questionnaires of organisational practices, enable you to identify and describe the variability in different phenomena. In contrast, explanatory or analytical research allows you to examine and explain relationships between variables, in particular cause-and-effect relationships. Alternatively, research requiring respondents to complete a quantitative diary regularly may use a short questionnaire administered repeatedly. These purposes have different research design requirements, which we discuss later (Section 11.3).

Although questionnaires may be used as the only data collection method, it can be better to link them with other methods in a mixed or multiple method research design (Section 5.4). For example, a questionnaire to discover customers' attitudes can be complemented by in-depth interviews to explore and understand these attitudes (Section 10.2).

Questionnaire completion modes and mediums

The design of a questionnaire differs according to its completion mode – whether it is completed by the respondent or a researcher and the medium through which it is distributed and returned or collected (Figure 11.1). **Self-completed questionnaires** are usually completed by the respondents and are often referred to as surveys. Such questionnaires are usually distributed to respondents online (**online questionnaire**), respondents either accessing the questionnaire through a web browser using a hyperlink (**Web questionnaire**)

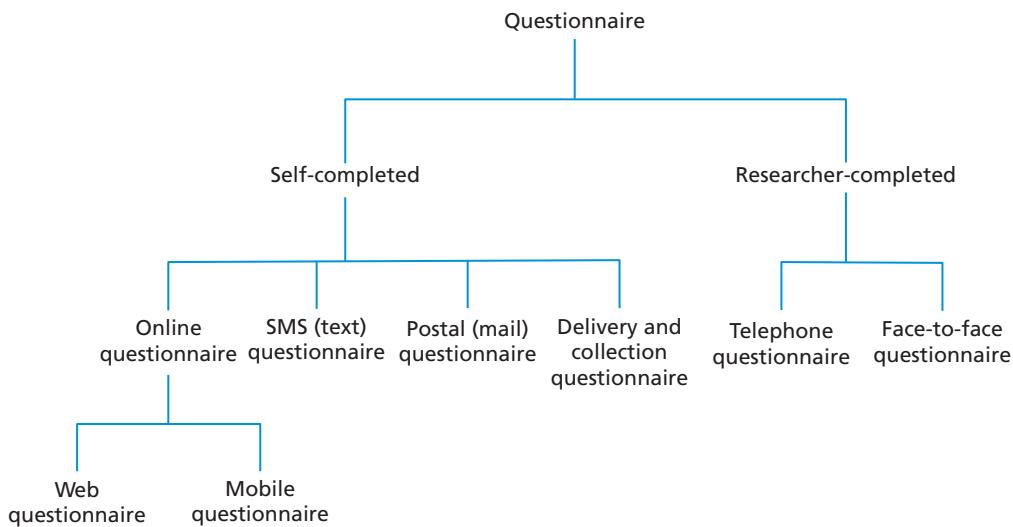


Figure 11.1 Questionnaire completion modes and mediums

on their computer, tablet or smartphone; or directly such as via a QR (quick response) code scanned into their smartphone or camera equipped device (**mobile questionnaire**). However, it is worth noting that such devices are increasingly blurring into each other (Kozinets 2020).

Questionnaires can also be delivered to each respondent's smartphone as a series of SMS (short message service) texts (**SMS questionnaires**), posted to respondents who return them by post after completion (**postal or mail questionnaires**) or delivered by hand to each respondent and collected later (**delivery and collection questionnaires**). Responses to **researcher-completed questionnaires** (also known as interviewer-completed questionnaires) are recorded by the researcher or a research assistant on the basis of each respondent's answers. Researcher-completed questionnaires undertaken using the telephone are known as **telephone questionnaires**. The final of these, **face-to-face questionnaires**, refers to those questionnaires where the researcher or a research assistant physically meet respondents and ask the questions face-to-face. These are also known as **structured interviews** but differ from semi-structured and unstructured (in-depth) interviews (Section 10.2), as there is a defined schedule of questions from which the researcher or research assistant should not deviate.

The choice of questionnaire

Your choice of questionnaire will be influenced by a variety of factors related to your research question(s) and objectives (Table 11.1), and in particular the:

- characteristics of the respondents from whom you wish to collect data;
- importance of reaching a particular person as respondent;
- importance of respondents' answers not being contaminated or distorted;
- size of sample you require for your analysis, taking into account the likely response rate;
- types of question you need to ask to collect your data;
- number of questions you need to ask to collect your data.

These factors will not apply equally to your choice of questionnaire, and for some research questions or objectives may not apply at all. The completion mode and medium you choose will dictate how confident you can be that the respondent is the person whom you wish to answer the questions and thus the reliability of responses (Table 11.1). Even if you address a postal questionnaire to a company manager by name, you have no way of ensuring that the manager will be the respondent. The manager's assistant or someone else could complete it! Online questionnaires, delivered by an emailed hyperlink, offer greater control because most people read and (hopefully) respond to their own emails. Similarly, SMS questionnaires, although only suitable for short questionnaires, are likely to be answered by the actual respondent as most people read and reply to text messages sent to them. With delivery and collection questionnaires, you can sometimes check who has answered the questions at collection. By contrast, researcher-completed questionnaires enable you to ensure that the respondent is whom you want. This improves the reliability of your data. In addition, you can record some details about non-respondents, allowing an assessment of the impact of bias caused by refusals.

Any contamination or distortion of respondents' answers will reduce your data's reliability (Table 11.1). Sometimes, if they have insufficient knowledge or experience, they may deliberately guess at the answer, a tendency known as **uninformed response**. This is more likely when the questionnaire has been incentivised (Section 11.8). Respondents to self-completed questionnaires are relatively unlikely to answer to please you or because they believe certain responses are more **socially desirable** (Dillman et al. 2014). They may, however, discuss their answers with others, thereby contaminating their response. Respondents to telephone and face-to-face questionnaires are more likely to answer to please due to their contact with you, although the impact of this can be minimised by good interviewing technique (Sections 10.7 to 10.9). Responses can also be contaminated or distorted when recorded. In extreme instances, research assistants may invent responses. For this reason, random checks of research assistants are often made by survey organisations. When writing your project report you will be expected to state your response rate (Section 7.5). When doing this you need to be careful not to make unsubstantiated claims if comparing with other questionnaires' response rates. While such comparisons place your response rate in context, a higher than normal response rate does not prove that your findings are unbiased (Rogelberg and Stanton 2007). Similarly, a lower than normal response rate does not necessarily mean that responses are biased.

The questionnaire you choose will affect the number of people who respond (Section 7.5). Researcher-completed questionnaires will usually have a higher response rate than self-completed questionnaires (Table 11.1). The size of your sample and the way in which it is selected will have implications for the confidence you can have in your data and the extent to which you can generalise (Section 7.5).

Longer questionnaires are more likely to result in careless responding than shorter ones, sometimes being reduced when their completion is supervised (Bowling et al. 2021). The presence of a researcher (or the use of cloud-based survey design, data collection and analysis software such as Qualtrics Research core™ and SurveyMonkey™) means that it is also easier to route different subgroups of respondents to answer different questions using a filter question (Section 11.4). The suitability of different types of question also differs between completion modes and questionnaire mediums.

Your choice of questionnaire completion mode and medium will also be affected by the resources you have available (Table 11.1), and in particular the:

- time available to complete the data collection;
- financial implications of data collection and entry;
- availability of research assistants and field workers to assist;
- cloud-based survey design, data collection and analysis software.

Table 11.1 Main attributes of questionnaires

| Attribute | Online | SMS | Postal | Delivery and collection | Telephone | Face-to-face |
|---|---|---|--|---|--|--------------|
| Population's characteristics for which suitable | IT literate individuals with access to the Internet, often contacted by email | Individuals with a mobile telephone | Literate individuals who can be contacted by post; selected by name, household, organisation, etc. | Individuals who can be telephoned; selected by name, household, organisation, etc. | Any; selected by name, household, organisation, in the street, etc. | |
| Confidence that right person has responded | High with email | High as have mobile phone number | Low | Low but can be checked at collection | High | |
| Likelihood of contamination or distortion of respondent's answer | Low, except where questions relate to use of Internet or associated technologies | Low | May be contaminated by consultation with others | Occasionally distorted or invented by researcher/research assistant | Occasionally contaminated by consultation or distorted/invented by researcher/research assistant | |
| Size of sample | Large, can be geographically dispersed | | | Dependent on number of field workers | Dependent on number of researchers/research assistants | |
| Likely response rate^a | Variable to low, 30–50% reasonable for Web within organisations, otherwise 10% or even lower | Low, often 10% or even lower | Variable, 30–50% reasonable | High, 50–70% reasonable | | |
| Feasible length of questionnaire | Equivalent of 6–8 A4 pages, minimise scrolling down | Short, as few questions as possible, preferably no more than 3 | 6–8 A4 pages | Up to half an hour | Variable depending on location | |
| Suitable types of question | Closed questions but not too complex; complicated sequencing fine if uses cloud-based software; must be of interest to respondent | Closed questions but not too complex; Questions need to be kept as succinct as possible | Closed questions but not too complex; simple sequencing only; must be of interest to respondent | Open and closed questions, including complicated questions; complicated sequencing feasible | | |

| Attribute | Medium | | | | |
|---|---|--|---|--|--|
| | Online | SMS | Postal | Delivery and collection | Telephone |
| Time taken to complete collection | 2–6 weeks from distribution (dependent on number of follow-ups) | Almost immediate | 4–8 weeks from posting (dependent on number of follow-ups) | Dependent on sample size, number of research assistants, etc. | Dependent on sample size, number of researchers/research assistants, etc., but slower than self-completed for same sample size |
| Main financial resource implications | Cloudbased software subscription, purchase of list of respondents' email addresses or data panel participants | Cloud-based software subscription, purchase of list of mobile phone numbers or data panel participants | Outward and return postage, photocopying, clerical support, data entry | Research assistants, telephone calls, clerical support; photocopying and data entry if not using CATI ^b ; survey tool if using CATI | Research assistants, telephone calls, clerical support; photocopying and data entry if not using CAPI ^c ; survey tool if using CAPI |
| Role of researcher/research assistants in data collection | None | | Delivery and collection of questionnaires; enhancing respondent participation | Enhancing respondent participation; guiding the respondent through the questionnaire and recording responses; answering respondents' questions | |
| Data input ^d | Automated through cloud-based software | Closed questions can be designed so that responses may be entered using optical mark readers after questionnaire has been returned | Response to all questions entered at time of collection using cloud-based software or CATI ^b | Response to all questions can be entered at time of collection using cloud-based software or CAPI ^d | |

^aDiscussed in Section 7.5. ^bComputer-aided telephone interviewing. ^cComputer-aided personal interviewing. ^dDiscussed in Section 12.4.
Sources: Authors' experience; Baruch and Holtom (2008); De Vaus (2014); Dillman et al. (2014); Saunders (2012); van de Heijden (2017).

The time needed for data collection increases markedly for delivery and collection and researcher-completed questionnaires where the samples are geographically dispersed (Table 11.1). One way you can overcome this constraint is to select your sample using cluster sampling (Section 7.6). For online questionnaires, **computer-aided personal interviewing (CAPI)** or **computer-aided telephone interviewing (CATI)**, you will need to consider the availability (and often the cost) of obtaining lists of email addresses or telephone numbers. For postal questionnaires you will need to consider the costs of reproducing the questionnaire, postage and entering the data for computer analysis. For telephone questionnaires you will also need to consider the cost of telephone calls rather than post. If you are working for an organisation, postage costs may be reduced by using *Freepost* for questionnaire return. This means that you pay only postage and a small handling charge for those questionnaires that are returned by post. However, the use of freepost rather than a stamp may adversely affect your response rates (see Table 11.5).

Data collected by questionnaires will, almost without exception, be analysed by computer. Virtually all cloud-based survey design, data collection and analysis software such as Qualtrics XM™ and SurveyMonkey™ allow you to design your questionnaire, capture, automatically code and save the data, and subsequently analyse the data within the software or export it as a data file for external analysis (Box 11.1). Data capture is most straightforward for closed questions where respondents select their answer from a prescribed list. Cloud-based survey software usually offers you the choice of exporting your data as either the actual text, or numeric codes (Box 11.1). For other questionnaires, responses will need subsequently to be coded, entered (typed) and saved in your analysis software (Section 12.4). As a rough rule, you should analyse questionnaire data by computer if they have been collected from 30 or more respondents.

11.3 Deciding what data need to be collected

Research design requirements

Unlike in-depth and semi-structured interviews (Chapter 10), the questions you ask in questionnaires need to be defined precisely prior to data collection. Whereas you can prompt and explore issues further with in-depth and semi-structured interviews, this will not be possible using questionnaires. In addition, the questionnaire offers only one chance to collect the data as it is often impossible to identify respondents or to return to collect additional information. This means that the time you spend planning precisely what data you need to collect, how you intend to analyse them (Chapter 12) and designing your questionnaire to meet these requirements is crucial if you are to answer your research question(s) and meet your objectives.

For most business and management research, the data you collect using questionnaires will be used for either descriptive or explanatory purposes. For questions where the main purpose is to describe a population's characteristics either at a fixed time or at a series of points over time to enable comparisons, you will normally need to deliver your questionnaire to a sample. The sample needs to be as representative and accurate as possible where it will be used to generalise about a population (Sections 7.4–7.6). You will also probably need to relate your findings to earlier research. It is therefore important that you select the appropriate characteristics to answer your research question(s) and to address your objectives. You will need to have:

- reviewed the literature carefully;
- discussed your ideas with colleagues, your project tutor and other interested parties.



Box 11.1 Focus on student research

Using cloud-based software to design a questionnaire

Ben's research project involved emailing a hyperlink to a Web questionnaire to small and medium-sized

enterprise owners to discover how they defined small business success. He designed his questionnaire using the cloud-based survey software Qualtrics as this would either allow him to analyse his data within the software or download his data as a matrix (table) of numeric codes (values) and use analysis software such as an Excel spreadsheet, IBM SPSS Statistics or a database.

Source: Copyright © 2021 Qualtrics LLC. Used With Permission.

For research involving organisations, we have found it essential to understand the organisational context in which we are undertaking the research. Similarly, for international or cross-cultural research it is important to have an understanding of the countries and cultures in which you are undertaking the research. Without this it is easy to make mistakes, such as using the wrong terminology or language, and to collect useless data. For many research projects an understanding of relevant organisations can be achieved through browsing company websites (Section 8.2), observation (Sections 9.3 and 9.4) and in-depth and semi-structured interviews (Section 10.4).

Explanatory research is usually deductive, using data to test a theory or theories. This means that, in addition to those issues raised for descriptive research, you need to define the theories you wish to test as relationships between variables prior to designing your questionnaire. You will need to have reviewed the literature carefully, discussed your ideas

widely and conceptualised your own research clearly prior to designing your questionnaire (Ghauri et al. 2020). In particular, you need to be clear about which relationships you think are likely to exist between variables:

- a dependent variable that may change in response to changes or manipulations in other independent variables;
- an independent variable that, when changed or manipulated, may cause changes in dependent variables;
- a mediating variable that transmits the effect of an independent variable to a dependent variable;
- a moderating variable that affects the relationship between an independent variable and a dependent variable;
- a control variable that needs to be kept constant to avoid it influencing the effect of the independent variable on the dependent variable.

As these relationships (Table 5.4) are likely to be tested through statistical analysis (Sections 12.8 to 12.13) of the data collected by your questionnaire, you need to be clear about the detail in which they will be measured at the design stage. Where possible, you should ensure that questions are compatible with those used in other relevant research so that comparisons can be made (Box 11.2).



Box 11.2 Focus on management research

Developing compatible questions to enable comparisons

In their article in the *Industrial Relations Journal*, Felstead and colleagues (2019) conceive and develop a short measure of job quality that is compatible with existing national skills and employment surveys. In developing their measure, they argue that three principles apply. First, that job quality needs to comprise a set of work features that can enhance or diminish worker wellbeing. Second, it needs to focus on attributes of the worker's actual job and not their personal circumstances or background. Third, there are a variety of attributes, such as pay that have the capability of enhancing or reducing worker wellbeing.

As Felstead and colleagues wished to design a short, easy to use online quiz, they could only use a

relatively small number of questions. They also wanted to allow quiz takers to benchmark their own answers with the findings from the representative national Skills and Employment Survey. This meant that they had to replicate as far as possible questions used in this survey.

Their resultant quiz comprised 21 questions. Of these 20 represented job demands and job resources using rating (scale) questions. Questions relating to job demands included 'How often does your job involve you working at very high speed?' For this question respondents were asked to choose one of seven responses ranging from 'never' to 'all of the time' (Felstead et al. 2019, p. 10). Questions relating to job resources, included 'How much influence do you personally have on deciding what tasks you are to do?' with the four possible responses ranging from 'none' to 'a great deal' (Felstead et al. 2019, p. 10). The final question, which was open ended, asked respondent to, in their own words, rate their job and explain their rating.

The quiz can be taken at <https://howgoodismyjob.co.uk/>

Types of data that can be collected

Dillman et al. (2014) distinguishes between three types of data that can be collected using questionnaires:

- factual and demographic;
- attitudes and opinions;
- behaviours and events.

These distinctions are important as they relate to the ease of obtaining accurate data and influence the way your questions are worded (Box 11.3). **Factual and demographic variables** contain data that are readily available to the respondent and are likely to, assuming the respondent is willing to disclose, be accurate. These variables include characteristics such as age, gender, marital status, education, occupation and income. They are used to explore how attitudes and opinions, and behaviours and events, differ, as well as to check that the data collected are representative of the total population (Section 7.5). **Attitude and opinion variables** contain data that respondents may have needed to think about before answering. They are likely to be influenced by the context in which the question was asked; recording how respondents feel about something or what they think or believe is true or false. **Behaviour and event variables** are also likely to be influenced by context. They contain data about what people did (behaviours) or what happened (events) in the past, is happening now, or will happen in the future.

Ensuring essential data are collected

A problem experienced by many students and organisations we work with is ensuring that data collected will enable the research question(s) to be answered and the objectives achieved. Although no method is infallible, one way is to create a **data requirements table** (Table 11.2). This summarises the outcome of a six-step process:

- 1** Decide whether the main outcome of your research is descriptive or explanatory.
- 2** Use your aim and objectives or overarching research question(s) to develop more specific investigative questions about which you need to gather data, noting how each relates to theory and key concepts in the literature.
- 3** Repeat the second stage if you feel that your investigative questions are not sufficiently precise.
- 4** Keeping in mind relevant theory and key concepts in the literature, identify the variables about which you must collect data to answer each investigative question.
- 5** Establish the level of detail required from the data for each variable.
- 6** Use (with acknowledgement) or adapt existing measurement questions or develop new questions to capture the data at the level required for each variable.

Investigative questions state with precision and depth those questions you need to answer in order to address satisfactorily overarching research question to meet your aim and objectives (Sections 2.5 and 2.6). They need to be generated with regard to your research question(s) and objectives. Some investigative questions may need subdividing into more detailed investigative questions. For each you need to be clear whether you are interested in facts/demographics, attitudes/opinions or behaviours/events (discussed earlier), as what appears to be a need to collect one sort of variable frequently turns out to be a need for another. We have found theory and key concepts from the literature, discussions with interested parties and pilot studies to be of help here.



Box 11.3 **Focus on student research**

Opinion, behaviour and attribute questions

Emily was asked by her employer to undertake an anonymous survey of financial advisors' ethical

values. In particular, her employer was interested in the advice given to clients. After some deliberation she came up with three questions that addressed the issue of putting clients' interests before their own:

Emily's choice of question or questions to include in her questionnaire was dependent on whether she needed to collect data on financial advisors' attitudes, opinions, or behaviours. Question 2 would collect

- 2** How do you feel about the following statement? 'Financial advisors should place their clients' interest before their own.'

| | | |
|-----------------------------------|----------------------------|--------------------------|
| (please tick the appropriate box) | strongly agree | <input type="checkbox"/> |
| | mildly agree | <input type="checkbox"/> |
| | neither agree nor disagree | <input type="checkbox"/> |
| | mildly disagree | <input type="checkbox"/> |
| | strongly disagree | <input type="checkbox"/> |

- 3** In general, do financial advisors place their clients' interests before their own?

| | | |
|-----------------------------------|---------------|--------------------------|
| (please tick the appropriate box) | always yes | <input type="checkbox"/> |
| | usually yes | <input type="checkbox"/> |
| | sometimes yes | <input type="checkbox"/> |
| | seldom yes | <input type="checkbox"/> |
| | never yes | <input type="checkbox"/> |

- 4** How often do you place your clients' interests before your own?

| | | |
|-----------------------------------|--------------------|--------------------------|
| (please tick the appropriate box) | 81–100% of my time | <input type="checkbox"/> |
| | 61–80% of my time | <input type="checkbox"/> |
| | 41–60% of my time | <input type="checkbox"/> |
| | 21–40% of my time | <input type="checkbox"/> |
| | 0–20% of my time | <input type="checkbox"/> |

data on respondents' opinions about financial advisors placing their clients' interest before their own. This question asks respondents how they feel. In contrast, question 3 asks respondents whether financial advisors in general place their clients' interests before their own. It is therefore concerned with their individual opinions regarding how financial advisors act. Question 4 focuses on how often the respondents placed their clients' interests before their own. Unlike the previous questions, it

is concerned with their actual behaviour rather than their opinion.

To answer her research questions and to meet her objectives Emily also needed to collect data to explore how ethical values differed between subgroupings of financial advisors. One theory she had was that ethical values were related to age. To test this, she needed to collect demographic data on respondents' ages. After some deliberation she came up with question 5:

- 5** How old are you?

| | | |
|-----------------------------------|--------------------------|--------------------------|
| (please tick the appropriate box) | Less than 30 years | <input type="checkbox"/> |
| | 30 to less than 40 years | <input type="checkbox"/> |
| | 40 to less than 50 years | <input type="checkbox"/> |
| | 50 to less than 60 years | <input type="checkbox"/> |
| | 60 years or over | <input type="checkbox"/> |

Table 11.2 Data requirements table

| Research aim/objectives/question(s): | | | | |
|--------------------------------------|----------------------|-------------------------------|---|--|
| Type of research: | | | | |
| Investigative questions | Variable(s) required | Detail in which data measured | Relation to theory and key concepts in the literature | Measurement question included in questionnaire ✓ |

| Investigative questions | Variable(s) required | Detail in which data measured | Relation to theory and key concepts in the literature | Measurement question included in questionnaire ✓ |
|-------------------------|----------------------|-------------------------------|---|--|
| | | | | |

You should then identify the variables about which you need to collect data to answer each investigative question and decide the level of detail at which these will be measured. Again, your literature review and associated research can suggest possibilities. However, if you are unsure about the detail needed you should measure at a more precise level. Although this is more time-consuming, it will give you flexibility in your analyses. In these you will be able to use analysis software to group or combine data (Section 12.2).

Once your data requirements table is complete (Box 11.4), it must be checked to make sure that all data necessary to answer your investigative questions are included. When



Box 11.4 Focus on student research

Data requirements table

As part of his work placement Greg was asked to discover employees' attitudes to the outside smoking area at his organisation's restaurants and bars. Discussion with

senior management and colleagues at the restaurant where he worked and reading relevant literature helped him to firm up his objective and investigative questions and the level of detail in which the data were measured. In addition, he wanted to be able to compare his findings with earlier research by Jackson and Taylor (2015) in the journal *Tourism and Hospitality Research* and Louka et al. (2006) in the *Journal of Health Psychology*.

One of his objectives is included in the extract from his table of data requirements:

| <ul style="list-style-type: none"> Research objective: To establish employees' attitudes to the outside smoking area at restaurants and bars. | | | | |
|---|---|---|---|-----------------------------------|
| <ul style="list-style-type: none"> Type of research: Predominantly descriptive, although wish to examine differences between restaurants and bars, and between different groups of employees. | | | | |
| Investigative questions | Variable(s) required | Detail in which data measured | Relation to theory and key concepts in literature | Check included in questionnaire ✓ |
| <ul style="list-style-type: none"> Do employees feel that restaurants and bars should provide an outside smoking area for smokers? (opinion) | <ul style="list-style-type: none"> Opinion of employee to the provision of an outside smoking area for smokers | <ul style="list-style-type: none"> Feel... very strongly that it should, quite strongly that it should, no strong opinions, quite strongly that it should not, very strongly that it should not [NB will need separate questions for restaurants and for bars] | | |





Box 11.4 Focus on student research (continued)

| Investigative questions | Variable(s) required | Detail in which data measured | Relation to theory and key concepts in literature | Check included in questionnaire ✓ |
|---|---|--|--|-----------------------------------|
| • Do employees' opinions differ depending on . . . | • (Opinion of employee – outlined above) | • (Included above) | | |
| • . . . whether or not a smoker? (behaviour) | • Smoker | • Smoker, former smoker or non-smoker | • use these 3 groups from Jackson and Taylor (2015) | |
| • . . . nationality (factual) | | • Country of origin | • Louka et al. (2006) highlights differences between nationalities | |
| • How representative are the responses of employee? (demographic) | • Gender of employee • Job • Number of hours worked | • Male, female • Will need to obtain a list of jobs from the organisation • Actual hours worked on week of questionnaire | • Note: UK government defines full-time work as at least 35 hours a week | |

checking, you need to ensure that only data which are essential to answering your research question(s) and meeting your objectives are included. The final column is to remind you to check that your questionnaire includes a (measurement) question that collects the precise data required!

11.4 Questionnaire validity and reliability

The validity and reliability of the data you collect (and the response rate you achieve) depend, to a large extent, on the design of your questions, the structure of your questionnaire and the rigour of your pilot testing (Section 11.7). A valid questionnaire will enable accurate data that actually measure the concepts you are interested in to be collected, while one that is reliable will mean that these data are collected consistently. Hardy and Ford (2014) argue that even if everyone understands a questionnaire, they may interpret it in different ways due to three forms of miscomprehension:

- instructional, where instructions such as ‘please rank the following in order of importance, ranking the most important 1, the next 2 and so on’ are not followed; the respondent doing something else such as ranking all as 1;
- sentinel, where the respondent enriches or depletes the syntax of a question; for example, a respondent answers a question about ‘management’ as her or his ‘line manager’;
- lexical, where the respondent deploys a different meaning to a word to that intended by the researcher; for example, where the word ‘satisfied’ in a question is intended to refer to obligations being fulfilled, but is interpreted as gratification.

Building on these ideas it is therefore crucial that the instructions given, and questions asked, are acted on or understood by the respondent in the way intended by the researcher. Similarly, the answers given by the respondent need to be understood by the researcher in the way intended by the respondent. This means the design stage is likely to involve substantial rewriting in order to ensure that the respondent follows instruction and decodes your questions in the way you intended (Section 11.5).

Establishing validity

Internal and external validity

When discussing validity researchers often talk about internal and external validity. **Internal validity** in relation to questionnaires refers to the ability of your questionnaire to measure what you intend it to measure. It therefore focuses on the ability of your questionnaire to represent the reality of what you are measuring and, as a consequence, your conclusions are warranted, and alternative explanations eliminated. In contrast, **external validity** is about the extent to which the findings from your questionnaire are generalisable to other real world relevant contexts. This presents you with a problem as, if you actually knew the reality of what you were measuring, there would be no point in designing your questionnaire and using it to collect data! Researchers get around this problem by looking for other relevant evidence that supports the answers found using the questionnaire, relevance being determined by the nature of their research question and their own judgement.

Self-generated validity and ecological validity

Researchers design questions on the premise that what they are measuring is of interest, relevant and understood by those responding. However, where this is not the case, it becomes possible for you to develop a question or series of questions to measure a construct that did not exist in the respondents’ minds before the received and began to complete the questionnaire. **Self-generated validity** occurs when, in response to your questions, the respondent generates answers about a construct or topic for which they had no prior knowledge or about which they had no views. In generating these answers they, in effect, produce the thought processes predicted by theory (Forbes and Avis 2020). This means their attitudes or opinions are constructed during the research process by answering the questions, their responses to earlier questions being used as a basis for subsequent responses. Such research is argued to be ecologically invalid as, although respondents offer an opinion, they had no knowledge or interest prior to reading the questions. Their opinion therefore has no basis other than the research process as it is the process that has produced the findings. This is important as findings from research with low ecological validity have little utility in real life settings. In relation to questionnaires, **ecological validity** therefore

refers to whether the questions actually measure an attitude or behaviour that can be generalised to and reflect realistic real-life situations.

Content validity

Often, when discussing the validity of a questionnaire, researchers refer to content validity, criterion-related validity and construct validity. **Content validity** refers to the extent to which the measurement device, in our case the questions in the questionnaire, provides adequate coverage of the investigative questions. Judgement of what is ‘adequate coverage’ can be made in a number of ways. One involves careful definition of the research through the literature reviewed and, where appropriate, prior discussion with others. Another is to use a panel of individuals to assess whether each question in the questionnaire is ‘essential’, ‘useful but not essential’ or ‘not necessary’.

Criterion validity

Criterion-related validity, sometimes known as **predictive validity**, is concerned with the ability of the questions (measures) to make accurate predictions. This means that if you are using the data collected by questions within your questionnaire to predict customers’ future buying behaviours then a test of these questions’ criterion-related validity will be the extent to which the responses predict these customers’ actual buying behaviours. In assessing criterion-related validity, you will be comparing the data from your questionnaire with that specified in the criterion in some way. Often this is undertaken using statistical analysis such as correlation (Section 12.6).

Construct, convergent and discriminant validity

Construct validity refers to the extent to which a set of questions (known individually as scale items and discussed later in this section) actually measures the presence of the construct you intended them to measure. It is therefore dependent upon lexical and sentinel miscomprehension for each scale item being minimised. The term is normally used when referring to constructs such as attitude scales (Box 11.5), customer loyalty and the like. It can be thought of as answering the question: ‘How well can I generalise from this set of questions to the construct I am trying to measure?’ Because validation of such constructs against existing data is difficult, other methods are used. Where different scales are used to measure the same construct, the overlap (or correlation) between these scales is known as **convergent validity**. In contrast, where different scales are used to measure theoretically distinct constructs, an absence of overlap (or correlation) between the scales means they are distinctive and have **discriminant validity**. These are discussed in more detail in a range of texts, including Bloomberg et al. (2014).

Testing for reliability

As we noted earlier, reliability refers to consistency. Although for a questionnaire to be valid it must be reliable, this is not sufficient on its own. Respondents may consistently interpret a question in your questionnaire in one way, when you mean something else! This might be because of lexical or sentinel miscomprehension for a specific question. Consequently, although the question is reliable, this does not really matter as it has no internal validity and so will not enable your research question to be answered. Reliability is therefore concerned with the robustness of your questionnaire and, in particular, whether or not it will produce consistent findings at different times and in different contexts, such as with different samples or, in the case of a researcher-completed questionnaire, with different research assistants or field workers. Alternatively, respondents may answer inconsistently due to instructional miscomprehension. Between 5 and 9 per cent of respondents

do not read instructions that accompany a questionnaire, this being due to familiarity with the task of completing questionnaires (Hardy and Ford 2014).

Mitchell (1996) outlines three common approaches to assessing reliability, in addition to comparing the data collected with other data from a variety of sources. Although the analysis for each of these is undertaken after data collection, they need to be considered at the questionnaire design stage. They are:

- internal consistency;
- alternative form;
- test re-test.

Internal consistency

Internal consistency involves correlating the responses to questions in the questionnaire with each other. However, it is nearly always only used to measure the consistency of responses across a subgroup of the questions. There are a variety of methods for calculating internal consistency, of which one of the most frequently used is **Cronbach's alpha**. This statistic is usually used to measure the consistency of responses to a sub-set of questions (scale items) that are combined as a scale (discussed in Section 11.5) to measure a particular concept. It consists of an alpha coefficient with a value between 0 and 1. Values of 0.7 or above indicate that the questions combined in the scale internally consistent in their measurement.

Alternative form

Alternative form offers some sense of the reliability within your questionnaire through comparing responses to alternative forms of the same question or groups of questions. Where questions are included for this purpose, usually in longer questionnaires, they are often called 'check questions'. However, it is often difficult to ensure that these questions are substantially equivalent. Respondents may suffer from fatigue owing to the need to increase the length of the questionnaire, and they may spot the similar question and just refer back to their previous answer! It is therefore advisable to use check questions sparingly.

Test re-test

The final approach is test re-test. This estimate of reliability is obtained by correlating data collected with those from the same questionnaire collected under as near equivalent conditions as possible. The questionnaire therefore needs to be delivered and completed twice by respondents. This may create problems, as it is often difficult to persuade respondents to answer the same questionnaire twice. In addition, the longer the time interval between the two questionnaires, the lower the likelihood that respondents will answer the same way. We therefore recommend that you use this method only as a supplement to other methods. Further details of this and other approaches can be found in Mitchell (1996) and in books discussing more advanced statistics and analysis software such as Field (2018).

11.5 Designing individual questions

The design of each question should be determined by the data you need to collect (Section 11.3). When designing individual questions researchers do one of three things (Bourque and Clark 1994):

- adopt questions used in other questionnaires;
- adapt questions used in other questionnaires;
- develop their own questions.

Adopting or adapting questions is far less time-consuming than the lengthy process of developing your own questions, providing you can still collect the data you need to answer your research question(s) and to meet your objectives. It allows you to replicate, or to compare your findings with, another study so that reliability to be assessed. Some cloud-based survey software includes questions that you may use. Alternatively, you may find questions and coding schemes that you feel will meet your needs in existing questionnaires, journal articles or in Internet-based question banks, such as the UK Data Service's Variable and Question Bank (2021). This provides searchable access to over 250,000 questions drawn from a range of UK and cross-national surveys since the mid-1990s.

Using existing questions is often sensible as it allows you to compare your findings with other research, although you need to be careful and should still pilot test your questionnaire (Section 11.7). Questions designed by researchers have been designed with a specific purpose in mind, which may not meet your research aim and objectives. Inevitably they cannot take account of any changes that have occurred since the questions were devised and tested. Unfortunately, there are a vast number of poorly worded or biased questions in circulation, so always assess each question carefully. In addition, you will need to check whether you require permission to use these questions because of copyright. Questions are usually subject to copyright unless there is an express indication that these may be used. Even where no formal copyright has been asserted you should, where possible, contact the author and obtain permission. In your project report you should always state where you obtained the questions and give credit to their author.

Types of question

Initially, you need only consider the type, wording and length of individual questions rather than the order in which they will appear on the form. Clear wording of questions using terms that will be familiar to, and understood by, respondents can improve the validity of the questionnaire. Shorter questions are easier to understand than longer ones and questions should, ideally, be no longer than 20 words, excluding possible answers (Sekaran and Bougie 2019). Most questionnaires include a combination of open and closed questions. **Open questions**, sometimes referred to as open-ended questions, allow respondents to give answers in their own way (Fink 2016). **Closed questions**, sometimes referred to as closed-ended questions (Fink 2016) or **forced-choice questions** (De Vaus 2014), provide two or more alternative answers from which the respondent is instructed to choose. Closed questions are usually quicker and easier to answer, as they require minimal writing. Responses are also easier to compare as they have been predetermined. However, if these predetermined responses are misunderstood by respondents, then they will not be valid (Hardy and Ford 2014). Within this section we highlight six types of closed question that we discuss later:

- list, where the respondent is offered a list of items, any of which may be selected;
- category, where only one response can be selected from a given set of categories;
- ranking, where the respondent is asked to place something in order;
- rating scale, in which a rating device is used to record responses;
- quantity, to which the response is a number giving the amount;
- matrix, where responses to two or more questions can be recorded using the same grid;

as well as:

- creating scales to measure constructs by combining individual rating scale questions.

We also consider issues associated with translating questions into other languages and pre-coding responses.

Open questions

Open questions are used widely in in-depth and semi-structured interviews (Section 10.7). In questionnaires they are useful if you are unsure of the response, such as in exploratory research, when you require a detailed answer, when you want to find out what is uppermost in the respondent's mind or do not wish to list all possible answers (Box 11.5). With such questions, the precise wording of the question and the amount of space determine partially the length and fullness of response. However, if you leave too much space the question becomes off-putting. Respondents tend to write more when answering open questions on online questionnaires than the paper based equivalent, although they are mainly just more verbose rather than offering more insights (Saunders 2012). An example of an open question (from a self-completed questionnaire) is:

- 6 Please list up to three things you like about your current employment:

- 1
2
3

This question collects data about each respondent's opinion of what they like about their current employment. Thus, if salary had been the reason uppermost in their mind this would probably have been recorded first. When questionnaires are returned by large numbers of respondents, responses to open questions are extremely time consuming to code (Section 12.2). For this reason, it is usually advisable to keep their use to a minimum.

List questions

List questions offer the respondent a list of responses from which she or he can choose either one or more responses. Such questions are useful when you need to be sure that the respondent has considered all possible responses. However, the list of responses must be defined clearly and be meaningful to the respondent. The response categories you can use vary widely and include 'yes/no', 'agree/disagree' and 'applies/does not apply' along with 'don't know' or 'not sure'. If you intend to use what you hope is a complete list, you may wish to add a catch-all category of 'other'. This has been included in question 7, which collects data on respondents' religion.

7 What is your religion?

(Please tick ✓ the appropriate box)

No religion

Christian (including Church of England, Catholic, Protestant and
all other Christian denominations)

Buddhist

Hindu

Jewish

Muslim

Sikh

Any other religion

Please describe

Question 7 collects demographic data on religion, the respondent ticking (checking) the response that applies. In this list question, the common practice of not asking respondents to both check those that do apply and those which do not has been adopted. Consequently, respondents are not asked to indicate those religions to which they do not belong. If you choose to do this, beware: non-response could also indicate uncertainty, or for some questions that an item does not apply! It is also likely that respondents will not read the list from which they have to select appropriate responses so carefully (Dillman et al. 2014). For researcher-completed questionnaires, it is often helpful to present the respondent with a prompt card listing all responses.



Box 11.5 Focus on research in the news

Bank risk officers put climate change at the top of the agenda

A survey that got our attention

By Billy Nauman

On any given day, our inboxes here at Moral Money are flooded with approximately a million (give or take) pitches about ESG surveys or reports. And truth be told, it is rare we even open them.

Apologies to all the PR people reading this and shaking their fists in rage, but there are so many of you, so few of us and not enough hours in the day. The big problem is most of these pitches follow a similar formula: [Group X] polled [Group Y] and found soaring interest in [insert ESG topic here]. Rarely do they provide any compelling information we haven't already seen.

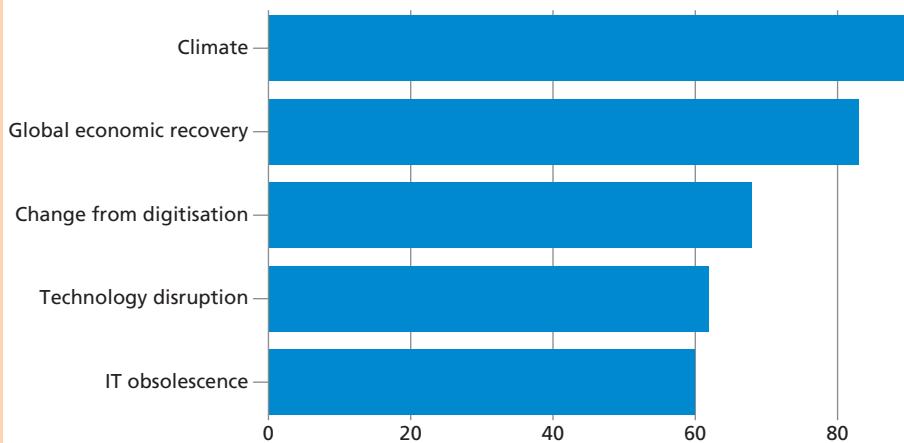
Sometimes, though, one will jump out and grab our attention, such as this new survey of bank risk managers from EY and the Institute of International Finance (IIF). In their poll, they found that interest in climate risk is (yes) soaring among banks. But what stood out about this survey is that it was not designed specifically to gauge opinions on ESG. The pollsters simply asked banks' risk officers what they thought was the biggest threat to their business. And climate organically popped up at the top of the heap.

Over a one-year timeframe, climate ranks third behind credit risk and cyber security. And over the next five years, risk officers see climate as the top problem banks will have to deal with.

This is the 11th year EY and the IIF have run this poll, and climate risk really 'came out of no-where', said Mark Watson, financial services managing director of EY Americas. One 'blindingly obvious' reason for this is the increase of extreme weather events such as wildfires and hurricanes that can interrupt day-to-day operations, he said. Banks

Most important risks for banks over the next five years

Chief risk officers name their top concerns



Source: EY, Institute of International Finance
© FT

are also concerned about upcoming climate stress tests from central banks, stranded assets and the transition risks that will accompany governments' moves to crack down on emissions. However, the heightened awareness is not all negative. Watson said banks were looking at climate more closely now because they saw a big opportunity. "I think people finally realise in financial services, we do not get to a zero-carbon economy without finance."



Source: Abridged from 'Bank risk officers put climate change at the top of the agenda', Billy Nauman, *Financial Times*, 25 June 2021. Copyright © 2021 The Financial Times Ltd

Category questions

In contrast, **category questions** are designed so that each respondent's answer can fit only one category. Such questions are particularly useful if you need to collect data about behaviour or attributes. The number of categories that you can include without affecting the accuracy of responses is dependent on the completion mode and questionnaire medium. Self-completed questionnaires and telephone questionnaires should usually have no more than five response categories (Fink 2016). Researcher-completed questionnaires can have more categories provided that a prompt card is used (Box 11.6) or, as in question 8, the researcher categorises the responses.



Box 11.6 Focus on student research

Use of a prompt card as part of a face-to-face questionnaire

As part of her face-to-face questionnaire, Jemma asked the following question:

Which of the following tourist sites did you visit while staying in Cusco?

[Show respondent cards 1 and 2 with the pictures of tourist sites. Read out names of the tourist sites one at a time. Record their response with a ✓ in the appropriate box].

| | Visited | Not visited | Not sure |
|--------------------------|--------------------------|--------------------------|--------------------------|
| Maras | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Moray | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Misminay Andean Village | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Sacsaywaman | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Priory of Santa Domingo | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Ollantaytambo | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Inca Pachacuteq Monument | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Qorikancha | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Pukapora (Red Fort) | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| The Sacred Valley | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |

Jemma gave card 1 (below) and subsequently card 2, both of which were A4 size, to each respondent, reading out the name of each tourist site and pointing to the photograph. She collected both cards after the question had been completed.



1. Maras



2. Moray



3. Misminay Andean village



4. Sacsaywaman



5. Priory of Santa Domingo



6. Ollantaytambo

Source: Copyright © 2018 Mark NK Saunders.

8 How often do you visit this retail park?

[Researcher: listen to the respondent's answer and tick ✓ as appropriate.]

- | | | |
|--|--------------------------------------|--------------------------|
| <input type="checkbox"/> First visit | 2 or more times a week | <input type="checkbox"/> |
| <input type="checkbox"/> Once a week | Less than once a week to fortnightly | <input type="checkbox"/> |
| <input type="checkbox"/> Less than fortnightly to once a month | Less often | <input type="checkbox"/> |

You should arrange responses in a logical order so that it is easy to locate the response category that corresponds to each respondent's answer. Your categories should be mutually exclusive (not overlapping) and should cover all possible responses. The layout of your questionnaire should make it clear which boxes refer to which response category by placing them close to the appropriate text.

Ranking questions

A **ranking question** asks the respondent to place things in rank order. This means that you can discover their relative importance to the respondent. In question 9, taken from an online questionnaire created in Qualtrics, the respondents are asked their opinions about the relative importance of a series of features when choosing a new car. The catch-all feature of 'other' is included to allow respondents to add one other feature, a subsequent question asking them to describe this.

9. Drag and drop the factors listed below so they are in order of importance to you in your choice of a new car.

Place the most important item at the top [1], the next second [2] and so on.

| | |
|---|---------------------------|
| 1 | Safety |
| 2 | Fuel efficiency |
| 3 | Eco-friendly |
| 4 | Suitability for daily use |
| 5 | Low price |
| 6 | High quality |
| 7 | Warranty and service |
| 8 | Other |

Source: Copyright © 2021 Qualtrics LLC. Used with permission.

With such questions, you need to ensure that the instructions are clear and will be understood by the respondent. In general, respondents find that ranking more than seven items takes too much effort, reducing their motivation to complete the questionnaire, so you should keep your list to this length or shorter (Bloomberg et al. 2014). Respondents can rank accurately only when they can see or remember all items. While this is straightforward for self-completed questions, providing all items are visible on the same screen or page, for researcher completed questionnaires it can be more difficult. Prompt cards (Box 11.6) on which you list all of the features to be ranked are one solution, but for telephone questionnaires you need to ask respondents to rank fewer items, as they will need to rely on their memory.

Rating scale questions

Rating scale questions are often used to collect opinion data. They most frequently use the **Likert-style rating** in which the respondent is asked how strongly they agree or disagree

with a statement or series of statements, usually on a four-, five-, six- or seven-point rating scale (Box 11.7). Coherent sets of rating scale questions are often combined to create a scale to measure a concept or construct, each question being termed a scale item (discussed later in this section). Statements should also include both positive and negative statements so as to ensure that the respondent reads each one carefully and thinks about which response box to tick. Possible responses to rating questions should be presented in a straight line (such as in question 10) rather than in multiple lines or columns, as this is how respondents are most likely to process the data (Dillman et al. 2014). If you intend to use a series of statements, you should keep the same order of response categories to avoid confusing respondents (Dillman et al. 2014). You should also, unlike in the opening vignette where whichever response category is selected the experience is at least ‘OK’, ensure response categories reflected the full spectrum of possible answers.

Question 10 (created using the cloud-based survey development software SurveyMonkey™) has been taken from an online questionnaire to an organisation’s employees and is designed to collect opinion data. In this rating question, an even number of

10. For the following statement, please select the response that matches your view most closely

I feel employees' views have influenced the decisions taken by management

Strongly agree Agree Disagree Strongly disagree

Source: Question created by SurveyMonkey™, Momentive Inc. (2021) San Mateo, California, USA



Box 11.7 Focus on management research

Eyes wide shut? Understanding and managing consumers' visual processing of country-of-origin clues

In a recent article in the *British Journal of Management*, Hakias and colleagues (2021) report on three eye-tracking experiments they undertook to investigate whether consumers naturally detected country-of-origin labels, whether such detection influences subsequent behavioural intentions and whether visual attention to these labels can be externally motivated.

In one of their eye-tracking experiments, after being exposed to same category products with country-of-origin labels ‘made in Romania’ and ‘made in Switzerland’, participants were asked to indicate their purchase intentions. These were recorded their answers using a rating of 1 through 7, where 1 = ‘not at all likely’, and 7 = ‘very likely’.

Those involved in the experiment also answered four questions comprising an existing four-item scale devised by Roth and Romeo (1992) to measure country image. Scale items assessed perceived innovativeness, design, prestige and workmanship of the country on a scale on 1 through 7, where 1 = ‘very low’, and 7 = ‘very high’; and example question being ‘How do you perceive the innovativeness of products that originate from Switzerland?’.

Hakias and colleagues found customers did notice country-of-origin cues, these predicting purchase intentions where at least 356 milliseconds had been spent looking at them.

points (four) has been used to force the respondent to express their feelings towards the statement by clicking on the ‘radio button’ under the response that matches their view most closely. By contrast, question 11, also from an online questionnaire created using SurveyMonkey™, contains an odd number of points (five). This inclusion of a neutral point allows the respondent to ‘sit on the fence’ by selecting the middle ‘neither agree nor disagree’ category when considering an implicitly negative statement. An alternative, the phrase ‘not sure’ is often used as it is less threatening to the respondent than admitting they do not know. This rating question is designed to collect data on employees’ opinions of the current situation.

11. For the following statement please select the answer that matches your view most closely

I believe there are 'them and us' barriers to communication in the company now

| | | | | |
|-----------------------|-----------------------|---------------------------|-----------------------|-----------------------|
| Strongly agree | Agree | Neither agree or disagree | Disagree | Strongly disagree |
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

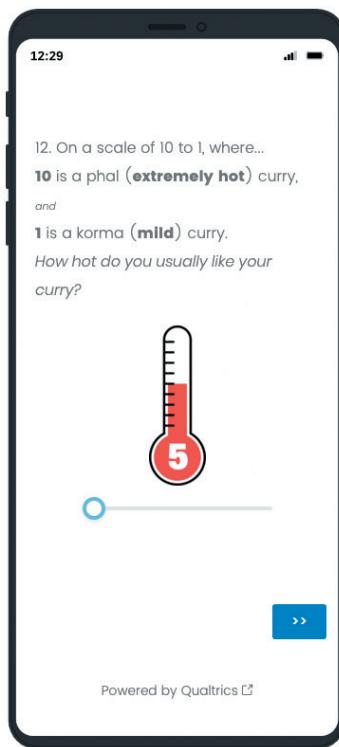
Source: Question created by SurveyMonkey™, Momentive Inc. (2021) San Mateo, California, USA

Both questions 10 and 11 are balanced rating scale questions as the possible answers are reflected around either an implicit (question 10) or an explicit (question 11) neutral point. The alternative is an unbalanced rating scale, such as question 12, which does not have a neutral point.

You can expand this form of rating question further to record finer shades of opinion, a variety of which are outlined in Table 11.3. However, respondents to telephone questionnaires find it difficult to distinguish between values when rating more than five points plus ‘don’t know’. In addition, there is little point in collecting data for seven or nine response categories, where these are subsequently combined in your analysis (Chapter 12). Colleagues and students often ask us how many points they should have on their rating scale. This is related to the likely measurement error. If you know that your respondents can only respond accurately to a three-point rating, then it is pointless to have a finer rating scale with more points!

In question 12 (created in Qualtrics and optimised in the software for completion on a mobile phone) a respondent’s opinion – how they usually like their curry – is captured on a 10-point numeric rating scale. In such rating questions it is important that the numbers reflect the answer of the respondent. Thus, 1 reflects a mild curry (korma) and 10 an extremely hot curry (phal), the number increasing as the temperature increases. Only these end categories (and sometimes the middle) are labelled, and these are known as self-anchoring rating scales. As in this question, a graphic that alters as the slider is moved can be used to reflect the rating scale visually and aid the respondent’s interpretation. The use of a slider has been shown to have no impact on responses when compared to more traditional radio-button formats (Roster et al. 2015) as in question 11. An additional category of ‘not sure’ or ‘don’t know’ can be added and should be separated slightly from the rating scale.

Another variation is the **semantic differential rating question**. These are often used in consumer research to determine underlying attitudes. The respondent is asked to rate a single object or idea on a series of bipolar rating scales. Each bipolar scale is described



Source: Copyright © 2021 Qualtrics LLC.
Used With Permission.

by a pair of opposite adjectives (question 13), designed to anchor respondents' attitudes. For these rating scales, you should vary the position of positive and negative adjectives from left to right to reduce the tendency to read only the adjective on the left (Bloomberg et al. 2014).

13 On each of the rows below, please click on the radio button to show how you feel about the service you received at our restaurant

| | | |
|-----------------|--|------------|
| Fast | <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input checked="" type="radio"/> <input type="radio"/> <input type="radio"/> | Slow |
| Unfriendly | <input type="radio"/> <input checked="" type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> | Friendly |
| Value for money | <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input checked="" type="radio"/> <input type="radio"/> <input type="radio"/> | Overpriced |

Source: Copyright © 2021 Qualtrics LLC. Used with permission.

Table 11.3 Response anchor categories for different types of rating scale questions

| Type of rating | Five categories | Seven categories |
|----------------|--|--|
| Agreement | strongly agree agree neither agree nor disagree/not sure/uncertain/unsure/undecided/no opinion* disagree strongly disagree | strongly agree agree/moderately agree/mostly agree* slightly agree/somewhat agree neither agree nor disagree/not sure/uncertain/unsure/undecided/no opinion* slightly disagree/somewhat disagree disagree/moderately disagree/mostly disagree* strongly disagree |
| Amount | far too much/nearly all/very large* too much/more than half/large* about right/about half/some* too little/less than half/small* far too little/almost none/not at all* | far too much/nearly all/very large* too much/more than half/large* slightly too much/quite large* about right/about half/some* slightly too little/quite small* too little/less than half/small* far too little/almost none/not at all* |
| Frequency | all the time/always* frequently/very often/most of the time/almost every time* sometimes/about as often as not/about half the time/occasionally* rarely/seldom/less than half the time/almost never/* never/practically never* | all the time/always/every time* almost all the time/almost always/usually* frequently/very often/most of the time* sometimes/about as often as not/about half the time* seldom/occasionally almost never/practically never/rarely* never/not at all* |
| Likelihood | very/almost always* good/often* reasonable/sometimes* slight/bit/seldom* none/not at all/rare* | extremely/almost always* very/usually* moderately/often* quite/reasonable/occasionally* somewhat/rarely * slight/bit/usually not* none/not at all/almost never* |

*Response dependent on question.

Source: Developed from Tharenou et al. (2007), Vagias (2006), and authors' experience

Quantity questions

The response to a **quantity question** is a number, which gives a factual amount of a characteristic. For this reason, such questions tend to be used to collect behaviour or attribute data. A common quantity question, which collects attribute data, is:

14 What is your year of birth?

| | | | |
|--|--|--|--|
| | | | |
|--|--|--|--|

(for example, for 1997 write:)

| | | | |
|---|---|---|---|
| 1 | 9 | 9 | 7 |
|---|---|---|---|

Because the response to this question data is coded by the respondent, the question can also be termed a **self-coded** question.

Matrix questions

A **matrix** or grid of questions enables you to record the responses to two or more similar questions at the same time. As can be seen from question 15, created in SurveyMonkey™, questions are listed down the left-hand side of the page, and responses listed across the top. The appropriate response to each question is then recorded in the cell where the row and column meet. Although using a matrix saves space, Dillman et al. (2014) suggests that respondents may have difficulties comprehending these designs and that they are a barrier to response.

15. The following items refer to your treatment by managers in general who are responsible for making decisions in Anytown Manufacturing Company that affect your work: *To what extent:*

| | to a large extent | to quite a large extent | to some extent | to quite a small extent | to a small extent | not at all |
|--|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| do they treat you with dignity? | <input type="checkbox"/> |
| do they treat you with respect? | <input type="checkbox"/> |
| are they at least as honest with bad news as good news in their communications with you? | <input type="checkbox"/> |

Source: Question created by SurveyMonkey™, Momentive Inc. (2021) San Mateo, California, USA

Using scales

Rating scale questions have been combined into scales to measure a wide variety of concepts such as customer loyalty, service quality, job satisfaction and Country image (Box 11.7). Referred to as **constructs**, these are attributes that can be inferred and assessed using a number of indicators but are not directly observable. Researchers infer the existence

of a construct using a series of measures (rating questions), these being combined into a scale that measures the construct. For each construct the resultant **scale** is represented by a composite scale score created by combining the scores for each of the rating questions. Each rating scale question is often referred to as a **scale item**. In the case of a simple Likert-type scale, for example, the scale (or composite) score for each case would be calculated by summing the scores of each of the rating questions (items) selected (De Vaus 2014).

Such aggregations are based on assumption that the scores attributed to each response (for example: ‘not at all’ = 1, ‘slight’ = 2, ‘reasonable’ = 3, ‘good’ = 4, ‘very’ = 5) are interval level data (Figure 12.1). This assumption has been made since Murphy and Likert developed their method of summative ratings in 1937! When aggregating scale item scores, it is important to ensure that scores for any items worded negatively are reverse coded. Using **reverse coding**, also known as **reverse scoring**, means high values will indicate the same type of response on every item. A detailed discussion of creating scales, including those by Likert and Guttman, can be found in DeVellis (2017). However, developing and validating your own scales is complex and time consuming. It therefore usually makes sense to use or adapt existing scales wherever possible (Schrauf and Navarro 2005). Since scaling techniques were first used in the 1930s, literally thousands of scales have been developed to measure attitudes and personality dimensions and to assess skills and abilities. Details of an individual scale can often be found by following up references in an article reporting research that uses that scale. In addition, there are a wide variety of handbooks that list these scales (e.g. American Psychological Association 2021; Bruner 2021). These scales can be used in your own research providing they:

- are valid for your research;
- can be operationalised by your intended respondents.

Box 11.8 provides a checklist to evaluate the suitability of existing scales for your own research.



Box 11.8 Checklist

Evaluating the suitability of existing scales

- ✓ Is the scale valid for your research?
 - Does the scale measure an appropriate construct for your research question?
 - Is the literature used to support the development of the scale robust?
 - Has the scale been empirically tested and validated?
 - Are the literature sources referenced?
 - Are the scale items justified fully by clear arguments?

- ✓ Can the scale be operationalised by the intended respondents?
 - Does the scale rely on putting the respondents into a scenario? In other words, do the respondents need to be thinking about something particular before responding to the scale items? (This may be appropriate in some cases such as when researcher is trying to understand attitude formation.)
 - Are each of the scale items and the predetermined, forced choice responses likely to be easy for the intended respondents to answer based upon their own experiences and daily lives?

Source: Developed by Sarah Forbes and Mark N.K. Saunders

It is worth remembering that you should only make amendments to the scale where absolutely necessary as significant changes could impact upon both the validity of the scale and, subsequently, your results! You also need to be aware that existing scales may be subject to copyright constraints. Even where there is no formal copyright, you should, where possible, contact the author and ask for permission. In your project report you should note where you obtained the scale and give credit to the author.

Question wording

The wording of each question will need careful consideration to ensure that the responses are valid – that is, measure what you think they do. Your questions will need to be checked within the context for which they were written rather than in abstract to ensure they are not misread and that they do not privilege a particular answer (Box 11.9). Given this, the checklist in Box 11.10 should help you to avoid the most obvious problems associated with wording that threatens the validity of responses.



Box 11.9 Focus on research in the news

The tale of the Brexit referendum question

By David Allen Green

The referendum question was: ‘Should the United Kingdom remain a member of the European Union or leave the European Union?’ The question was originally planned to be: ‘Should the United Kingdom remain a member of the European Union?’ The Electoral Commission assessed the original question and decided: ‘We have previously recommended the possibility of either a yes/no question for use at a referendum on European Union membership. However, in this assessment we have heard clearer views, particularly from potential campaigners to leave the European Union, about their concerns regarding the proposed yes/no question. Our assessment suggests that it is possible to ask a question which would not cause concerns about neutrality, while also being easily understood.’ The commission thereby recommended the wording used, and this was accepted by government and parliament.

Research had indicated there could be a difference. ‘It seemed to reveal there was 4 per cent in what the question was, whether it was a “yes/no” question or a “remain/leave” question.’

The referendum produced a 51.89 per cent vote for Leave. On a narrow and strict reading of the question, it meant there was a small but clear majority for the whole of the UK to leave the EU. In other words, there was a mandate for the ultimate objective. However, the same question, but in another form, might have had a different result.



Source: Abridged from ‘The tale of the Brexit referendum question’, David Allen Green, *Financial Times*, 3 Aug 2017. Copyright © 2017 The Financial Times



Box 11.10 Checklist

Evaluating question wording

- ✓ Does your question collect data at the right level of detail to answer your investigative question as specified in your data requirements table?
- ✓ Will respondents have the necessary knowledge to answer your question? A question on the implications of a piece of legislation would yield meaningless answers from those who were unaware of that legislation.
- ✓ Does your question appear to talk down to respondents? It should not!
- ✓ Does your question challenge respondents' mental or technical abilities? Questions that do this are less likely to be answered.
- ✓ Are the words used in your question familiar to all respondents, and will all respondents comprehend them in the same way? In particular, you should use simple words and avoid jargon, abbreviations and colloquialisms.
- ✓ Are there any words that sound similar and might be confused with those used in your question? This is a particular problem with researcher-completed questionnaires.
- ✓ Are there any words that look similar and might be confused if your question is read quickly? This is particularly important for self-completed questionnaires.
- ✓ Are there any words in your question that might cause offence? These might result in biased responses or a lower response rate.
- ✓ Can your question be shortened? Long questions are often difficult to understand, especially in researcher-completed questionnaires, as the respondent needs to remember the whole question. Consequently, they often result in no response at all.
- ✓ Are you asking more than one question at the same time? The question 'How often do you visit your mother and father?' contains two separate questions, one about each parent, so responses would probably be impossible to interpret.
- ✓ Does your question include a negative or double negative? Questions that include the word 'not' are sometimes difficult to understand. The question

'Would you rather not use a non-medicated shampoo?' is far easier to understand when rephrased as: 'Would you rather use a medicated shampoo?'

- ✓ Is your question unambiguous? This can arise from poor sentence structure, using words with different lexical meanings or having an unclear investigative question. If you ask 'When did you leave school?' some respondents might state the year, others might give their age, while those still in education might give the time of day! Ambiguity can also occur in category questions. If you ask employers how many employees they have on their payroll and categorise their answers into three groups (up to 100, 100–250, 250 plus), they will not be clear which group to choose if they have 100 or 250 employees.
- ✓ Does your question imply that a certain answer is correct? If it does, the question is biased and will need to be reworded, such as with the question 'Many people believe that too little money is spent on our public Health Service. Do you believe this to be the case?' For this question, respondents are more likely to answer 'yes' to agree with and please the researcher.
- ✓ Does your question prevent certain answers from being given? If it does, the question is biased and will need to be reworded. The question 'Is this the first time you have pretended to be sick?' implies that the respondent has pretended to be sick whether they answer yes or no!
- ✓ Is your question likely to embarrass the respondent? If it is, then you need either to reword it or to place it towards the end of the survey when you will, it is to be hoped, have gained the respondent's confidence. Questions on income can be asked as either precise amounts (more embarrassing), using a quantity question, or income bands (less embarrassing), using a category question. Questions on self-perceived shortcomings are unlikely to be answered.
- ✓ Have you incorporated advice appropriate for your completion mode and questionnaire medium (such as the maximum number of categories) outlined in the earlier discussion of question types?
- ✓ Are answers to closed questions written so that at least one will apply to every respondent and so that each of the responses listed is mutually exclusive?
- ✓ Are the instructions on how to record each answer clear?

Translating questions into other languages

Translating questions and associated instructions into another language requires care if your translated or target questionnaire is to be decoded and answered by respondents in the way you intended. For international research this is extremely important if the questions are to have the same meaning to all respondents. For this reason, Usunier et al. (2017) suggest that when translating the source questionnaire attention should be paid to:

- lexical meaning – the precise meaning of individual words (e.g. the French word *chaud* can be translated into two concepts in English and German, ‘warm’ and ‘hot’);
- idiomatic meaning – the meanings of a group of words that are natural to a native speaker and not deducible from those of the individual words (e.g. the English expression for informal communication, ‘grapevine’, has a similar idiomatic meaning as the German expression *Mundpropaganda*, meaning literally ‘mouth propaganda’);
- experiential meaning – the equivalence of meanings of words and sentences for people in their everyday experiences (e.g. terms that are familiar in the source questionnaire’s context such as ‘hybrid working’ or ‘dual career household’ may be unfamiliar in the target questionnaire’s context);
- grammar and syntax – the correct use of language, including the ordering of words and phrases to create well-formed sentences (e.g. in Japanese the ordering is quite different from English or Dutch, as verbs are at the end of sentences).

Usunier et al. (2017) outline a number of techniques for translating your source questionnaire. These, along with their advantages and disadvantages, are summarised in Table 11.4. In this table, the **source questionnaire** is the questionnaire that is to be translated, and the **target questionnaire** is the translated questionnaire. When writing your final project report, remember to include a copy of both the source and the target questionnaire as appendices. This will allow readers familiar with both languages to confirm that equivalent questions in both questionnaires have the same meaning.

Table 11.4 Translation techniques for questionnaires

| | Direct translation | Back-translation | Parallel translation |
|---------------|---|--|---|
| Approach | Source questionnaire to target questionnaire | Source questionnaire to target questionnaire; target questionnaire to source questionnaire; comparison of two new source questionnaires; creation of final version | Source questionnaire to target questionnaire by two or more independent translators; comparison of two target questionnaires; creation of final version |
| Advantages | Easy to implement, relatively inexpensive | Likely to discover most problems; easy to implement with translators at source country | Leads to good wording of target questionnaire |
| Disadvantages | Can lead to many errors (including those relating to meaning) between source and target questionnaire | Requires two translators, one a native speaker of the source language, the other a native speaker of the target language | Cannot ensure that lexical, idiomatic and experiential meanings are kept in target questionnaire |

Source: Developed from Usunier et al. (2017) ‘Translation techniques for questionnaires’ in International and Cross-Cultural Business Research. Copyright © 2017 Sage Publications.

Coding question responses

As you will be analysing your data by computer, question responses will need to be coded prior to entry. Online questionnaires designed using a cloud-based survey tool do this automatically, the software allowing the selected response to each closed question to either be given a numeric code or the actual answer recorded. Responses will be automatically saved and can subsequently be exported as a data file in a variety of formats such as Excel™, IBM SPSS Statistics compatible, or a CSV (comma-separated values) file (Box 11.1).

These codes allocated to response categories will affect your analyses. Where, as in questions 16 and 17, numeric codes are allocated to adjacent responses this makes it far easier to aggregate responses during analysis to ‘satisfactory’ (codes 5, 4 and 3) and ‘unsatisfactory’ (codes 2 and 1). For open questions the text entered by the respondent will be recorded verbatim and will need coding by you. Such responses are likely to require more complex coding using either the multiple-response or the multiple-dichotomy method. Coding is discussed in Section 12.2, and we recommend that you read this prior to designing your questions.

| | | | | | | |
|-----------|---|---|------------------------------------|--|------------------------------------|-------------------------------------|
| 16 | Is the service you receive? (Please circle O the number) | Excellent 5 | Good 4 | Reasonable 3 | Poor 2 | Awful 1 |
| 17 | Is the service you receive? (Please tick ✓ the box) | Excellent <input type="checkbox"/> 5 | Good <input type="checkbox"/> 4 | Reasonable <input type="checkbox"/> 3 | Poor <input type="checkbox"/> 2 | Awful <input type="checkbox"/> 1 |

For paper-based questionnaires you will need to allocate the codes yourself. For numerical responses, actual numbers can be used as codes. For other responses, you will need to design a coding scheme. As with online questionnaires, whenever possible, you should establish the coding scheme prior to collecting data and incorporate it into your questionnaire. As noted in Section 11.3, this should take account of the precise detail required to allow comparison with other relevant datasets and be compatible with their coding schemes (Section 12.2). Coding schemes can be printed on the paper questionnaire, thereby **pre-coding** the question and removing the need to code after data collection. Two ways of doing this are illustrated by questions 16 and 17, which collect data on the respondents’ opinions.

For open questions you will need to reserve space on your data collection form to code responses after data collection. Question 18 has been designed to collect attribute data in a sample survey of 5,000 people. Theoretically there could be hundreds of possible responses, and so sufficient spaces are left in the ‘For office use only’ box.

18 What is your full job title?
.....

For Office use only

11.6 Designing the questionnaire

Order and flow of questions

When designing your questionnaire, it is a good idea to spend time considering the order and flow of your questions. These should be logical to the respondent (and researcher) rather than follow the order in your data requirements table (Table 11.2). They should take account of possible bias caused by the ordering of the questions. For example, a question

asking a respondent to list the possible benefits of a new shopping centre could, if preceding a question about whether the respondent supports the proposed new shopping centre, bias respondents' answers in favour of the proposal.

To assist the flow of the questions it may be necessary to include **filter questions**. These identify those respondents for whom the following question or questions are not applicable, so they can skip those questions. Complex filter questions can be programmed using cloud-based software (and CAPI and CATI software) so that skipped questions are never displayed on the screen and as a consequence never asked (Dillman et al. 2014). In such situations the respondent is unlikely to be aware of the questions that have been skipped. However, you should beware of using more than two or three filter questions in paper-based self-completed questionnaires, as respondents tend to find having to skip questions annoying. The following example uses the answer to question 19 to determine whether questions 20 to 24 will be answered. (Questions 19 and 20 both collect factual data.)

- 19** Are you currently registered as unemployed? Yes ₁
If 'no' go to question 25 No ₂
- 20** How long have you been registered as unemployed? ₀₀ years ₀₀ months
(for example, for no years and six months write:) ₀ years ₆ months

Where you need to introduce new topics, phrases such as 'the following questions refer to . . .' or 'I am now going to ask you about . . .' are useful, although respondents may ignore or miscomprehend instructions (Section 11.4). For researcher-completed questionnaires, you will have to include instructions for the researcher or research assistant (Box 11.11).



Box 11.11

Focus on student research

Introducing a series of rating questions in a telephone questionnaire

As part of a telephone questionnaire, Stefan needed to collect data on respondents' opinions about motorway service stations. To do this he asked respondents to

rate a series of statements using a Likert-type rating scale. These were recorded as a matrix. Because his survey was conducted by telephone, and he wanted respondents to express an opinion, the rating scale was restricted to four categories: strongly agree, agree, disagree, strongly disagree.

In order to make the questionnaire easy to follow, Stefan used italic script to highlight the instructions and the words that the research assistant needed to read in bold. An extract is given below:

Now I'm going to read you several statements. Please tell me whether you strongly agree, agree, disagree or strongly disagree with each.

Read out statements 21 to 30 one at a time and after each ask . . .

Do you strongly agree, agree, disagree or strongly disagree?

Record respondent's response with a tick ✓

| | strongly agree | agree | disagree | strongly disagree |
|---|---------------------------------------|---------------------------------------|---------------------------------------|---------------------------------------|
| 21 I think there should be a greater number of service stations on motorways | <input type="checkbox"/> ₄ | <input type="checkbox"/> ₃ | <input type="checkbox"/> ₂ | <input type="checkbox"/> ₁ |



Box 11.12 Checklist

Assessing the question order

- ✓ Are questions at the beginning of your questionnaire more straightforward and ones the respondent will enjoy answering? Questions about attributes and behaviours are usually more straightforward to answer than those collecting data on opinions.
- ✓ Are questions at the beginning of your questionnaire obviously relevant to the stated purpose of your research? For example, questions requesting contextual information may appear irrelevant.
- ✓ Are questions and topics that are more complex placed towards the middle of your questionnaire? By this stage most respondents should be

undertaking the survey with confidence but should not yet be bored or tired.

- ✓ Are personal and sensitive questions towards the end of your questionnaire, and is their purpose explained clearly? On being asked these a respondent may refuse to answer; however, if they are at the end of a researcher-completed questionnaire you will still have the rest of the data!
- ✓ Are filter questions and routing instructions easy to follow so that there is a clear route through the questionnaire?
- ✓ (For researcher-completed questionnaires) Are instructions to the researcher easy to follow?
- ✓ Are questions grouped into obvious sections that will make sense to the respondent?
- ✓ Have you re-examined the wording of each question and ensured it is consistent with its position in the questionnaire as well as with the data you require?

Although the checklist in Box 11.12 should help you to avoid the most obvious problems associated with question order and flow in all questionnaires, the advice contained may sometimes appear contradictory. Where this is the case, you need to decide what is most important for your particular population.

Visual presentation

Visual presentation is important for all questionnaires. All questionnaires should be designed to make reading questions and filling in responses easy. The visual presentation of online and other self-completed questionnaires should, in addition, be attractive to encourage the respondent to fill it in and to return it, while not appearing too long. A two-column layout for a paper-based questionnaire can look attractive without decreasing legibility (Ekinci 2015). For online questionnaires a single column is preferable while, due to the screen size, only one question per page is often preferable for mobile questionnaires (Section 11.5, question 12) (Dillman et al. 2014). However, where the choice is between an extra screen (or page) and a cramped questionnaire the former is likely to be more acceptable to respondents (Dillman et al. 2014). Cloud-based survey software contain a series of style templates for typefaces, colours and page layout, as well as optimisation routines for screen, tablet and mobile phone. These are all helpful in producing a professional-looking questionnaire more quickly. However, it is worth noting that, whatever the completion mode and questionnaire medium, the best way of obtaining the maximum number of valid responses to questions is to keep both the visual presentation of the questionnaire and the wording of each question simple (Dillman 2014).

Research findings on the extent to which the length of your questionnaire will affect your response rate are mixed (De Vaus 2014). There is a widespread view that longer

questionnaires will reduce response rates relative to shorter questionnaires (Edwards et al. 2002). However, a very short questionnaire may suggest that your research is insignificant and hence not worth bothering with. Conversely, a questionnaire that takes over an hour to complete might just be thrown away by the intended respondent. In general, we have found that a length of between four and eight A4 pages (or equivalent) has been acceptable for both online and paper-based within-organisation self-completed questionnaires. In contrast, SMS questionnaires need to have far fewer questions, preferably five or less. Telephone questionnaires of up to half an hour have caused few problems, although this is dependent upon a respondent's location and time of day. Similarly, the acceptable length for face-to-face questionnaires can vary from only a few minutes in the street to over two hours in a more comfortable environment (Section 10.6). Based on these experiences and, noting that respondents generally respond more carelessly when answering longer questionnaires (Bowling et al. 2021), we recommend you do not make the questionnaire longer than is really necessary to meet your research questions and objectives.

Box 11.13 summarises the most important layout issues as a checklist of common mistakes to avoid.



Box 11.13 Checklist

Avoiding common mistakes in questionnaire layout

- ✓ (For self-completed questionnaires) Do questions appear well spaced on the screen or page? A cramped design will put the respondent off reading it and reduce the response rate. Unfortunately, a long questionnaire is equally off-putting!
- ✓ (For paper-based self-completed questionnaires) Is the questionnaire going to be printed on good-quality paper? Poor-quality paper implies that the survey is not important.
- ✓ (For self-completed questionnaires) Is the questionnaire going to be displayed or printed on a warm pastel colour? Warm pastel shades, such as yellow and pink, generate slightly more responses than white (Edwards et al. 2002) or cool colours, such as green or blue. White is a good neutral colour, but bright or fluorescent colours should be avoided.
- ✓ (For researcher-completed questionnaires) Will the questions and instructions be printed on one side of the paper only? A researcher will find it difficult to read the questions on the back of pages if you are using a questionnaire attached to a clipboard!
- ✓ Is your questionnaire easy to read? Questionnaires should be displayed in 12 point or 10 point using a plain font. Excessively long and unduly short lines reduce legibility. Similarly, respondents find CAPITALS, *italics* and shaded backgrounds more difficult to read. However, if used consistently, they can make completing the questionnaire easier.
- ✓ Have you ensured that the use of shading, colour, font sizes, spacing and the formatting of questions is consistent throughout the questionnaire?
- ✓ Is your questionnaire laid out in a format that respondents are accustomed to reading? Research has shown that many people skim-read questionnaires (Dillman et al. 2014). Instructions that can be read one line at a time from left to right moving down the page are, therefore, more likely to be followed correctly.
- ✓ Is your questionnaire optimised for the questionnaire medium you intend to use?

Explaining the purpose of the questionnaire

Online questionnaire's email invitation and welcome screen

Most online self-completed questionnaires are accompanied by an email invitation to participate that explains the purpose of the research and include a welcome screen which summarises the email and offers instructions on how to complete the questionnaire. While for web and mobile questionnaire's these should fit on one screen if possible, for SMS questionnaires the messages will need to be far more succinct.

For some research projects you may also send an email or text message prior to sending your questionnaire. This will be used by the respondent to decide whether to grant you access. This is often the only opportunity you have to convince the respondent to participate in your research and ways to help ensure this are discussed in Section 6.4.

The welcome screen is the first part of the online questionnaire that a respondent will see. Unfortunately, between 4 per cent and 9 per cent of your sample will not read instructions (Hardy and Ford 2014), while others will use it to decide whether to answer the accompanying questionnaire.

Dillman et al. (2014) and others note the messages contained in an online questionnaire's email invitation to participate (covering email) will affect the response rate. The results of Dillman et al.'s research, along with requirement of most ethics committees to stress that participation is voluntary, are summarised in the annotated email (Figure 11.2). As you will see, this includes a hyperlink to the questionnaire. Although not included in this example email, some cloud-based software allows you to also include an access code for participants. Where this is the case, you need to remember to include it in your email! At the start of your online questionnaire, your welcome screen needs to summarise this information and explain concisely why you wish the respondent to complete the survey (Box 11.14). This should summarise the main messages in the covering email. It should be followed by a question that asks explicitly for the respondent's consent (Box 11.16).

Postal and delivery and collection questionnaire's invitation letter and opening remarks

Postal questionnaires and delivery and collection questionnaires are usually accompanied by an invitation letter (covering letter) requesting participation (Figure 11.3). At the start of your postal questionnaire, or delivery and collection questionnaire, you need to summarise the information in this invitation letter explain clearly and concisely why you want the respondent to complete the survey. Dillman et al. (2014) argue that, to achieve as high a response rate as possible, this should be included as the first page of the questionnaire in addition to the covering letter. He suggests that in addition to a summary of the main messages in the invitation to participate letter (Figure 11.3) you include:

- a clear unbiased banner or title, which conveys the topic of the questionnaire and makes it sound interesting;
- a subtitle, which conveys the research nature of the topic (optional);
- a neutral graphic illustration or logo to add interest and to set the questionnaire apart.

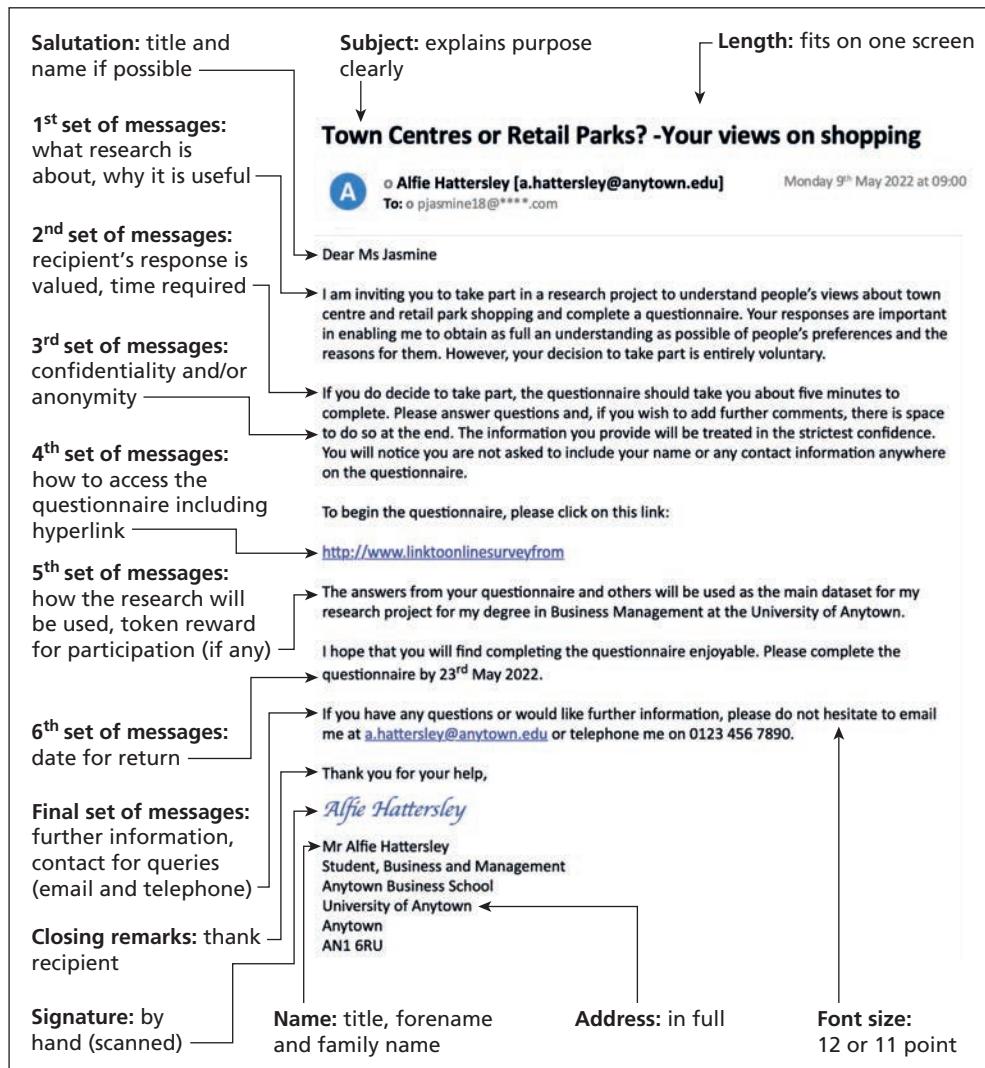


Figure 11.2 Annotated email invitation to participate (covering email)

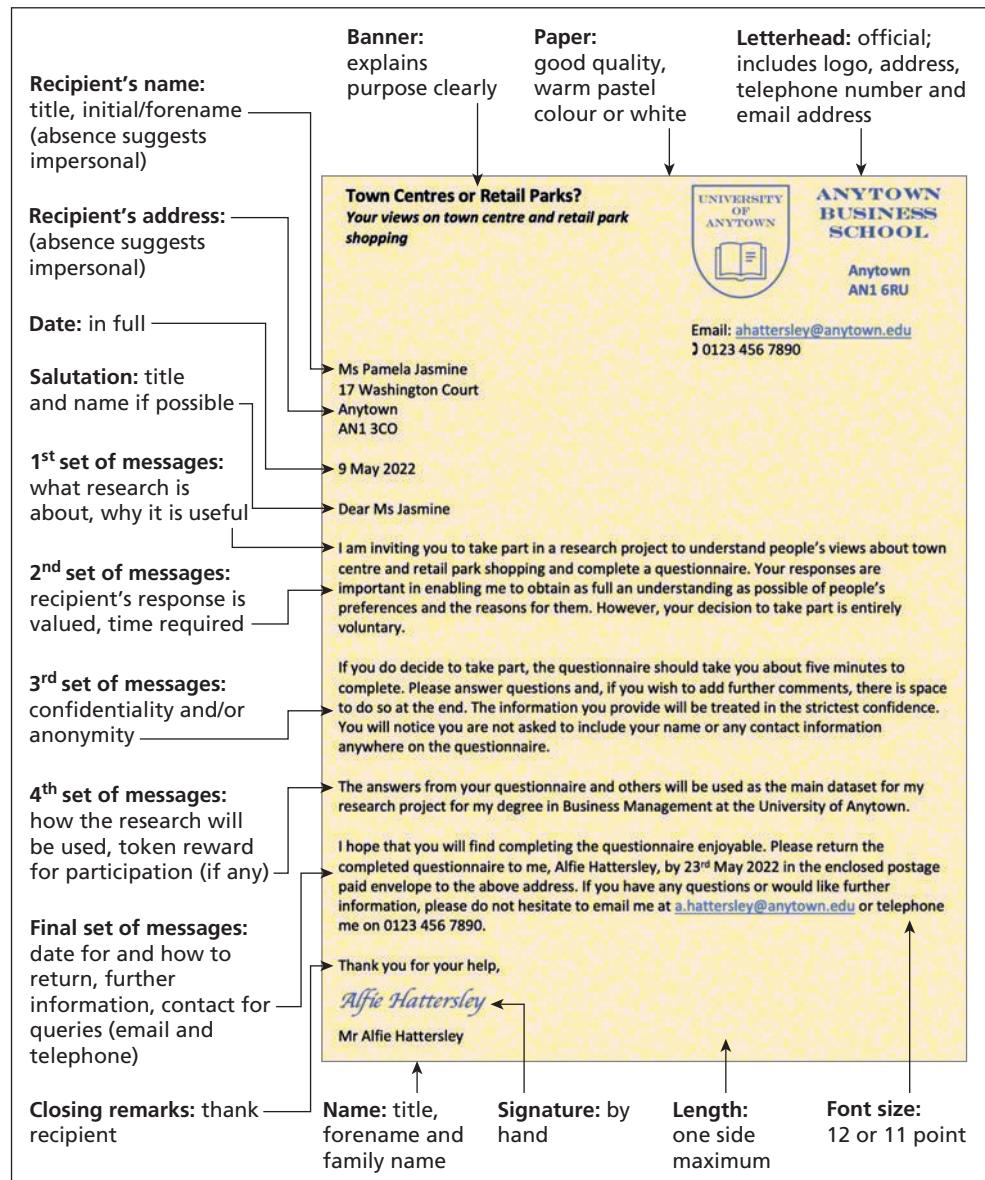


Figure 11.3 Structure of an invitation letter (covering letter)



Box 11.14 **Focus on student research**

Introducing an online questionnaire

Lily asked her project tutor to comment on what she hoped was the final draft of her online questionnaire. This included the following introduction:

ANYTOWN PRIVATE HOSPITAL STAFF SURVEY

Dear Sir or Madam

I am undertaking research on behalf of Anytown Private Hospital and we are inviting some people to take part. THE RESEARCH WILL HELP US DEVELOP THE FUTURE OF THE HOSPITAL. If you would like to take part in this research, please answer the questionnaire.

Thank you for your time.

Not surprisingly, her project tutor suggested that she re-draft her introduction. Her revised introduction follows:

Anytown Private Hospital

Staff Survey 2022

Your voice matters

Dear Colleague

This survey is being carried out to find out how you feel about the Hospital's policies to support colleagues like you in your work. Please answer the questions freely. You cannot be identified from the information you provide and no information about individuals will be given to the Hospital.



All the information you provide will be treated in the strictest confidence. Your decision to participate in this research is entirely voluntary.

If you do not wish to take part, just close this web page. If you do decide to take part, the questionnaire should take you about five minutes to complete. Please answer the questions in the space provided. Try to complete the questions at a time when you are unlikely to be disturbed. Also, do not spend too long on any one question. Your first thoughts are usually your best! Even if you feel the questions asked may not apply directly to your working life, please do not ignore them. Your answers are essential in building an accurate picture of the issues that are important to improving our support for people working for this Hospital.

There are no costs associated with completing the questionnaire other than your time.

When you have completed the questionnaire, please click on the < submit > button to send it to me.

I hope you will be willing to complete and submit the questionnaire and thank you for your time. A summary of the findings will be published on the Hospital intranet. If you have any queries or would like further information about this project, please telephone me on 01234–5678910 or email me on l.woollons@anytownhealthcare.com.

Thank you for your help.

Lily Woollons

Lily Woollons

Human Resources Department

Anytown Private Hospital

Anytown AN99 9HS

Researcher-completed questionnaire's invitation and opening remarks

Researcher-completed questionnaires will require the invitation and opening information to be phrased as a short introduction, given in the researcher's own words to each respondent. A template for a telephone questionnaire (developed from De Vaus 2014), which you, as researcher, would paraphrase, is given in the next paragraph.

Good morning/afternoon/evening. My name is [your name] from [your organisation]. I am undertaking a research project to find out [brief description of purpose of the research]. Your telephone number was drawn from a random sample of [brief description of the total population]. The questions I should like to ask will take about [number] minutes. If you have any queries, I shall be happy to answer them. [Pause] Before I continue, please can you confirm that this is [read out the telephone number] and that I am talking to [read out name/occupation/position in organisation to check that you have the right person]. Please can I confirm that you consent to answering the questions and ask you them now?

Obviously, you will need to amend this template if using face-to-face questionnaires and, if selecting respondents using a quota sample, will need to ensure your opening questions enable your sample quotas to be identified. You will also need to have prepared answers to the more obvious questions that the respondent might ask you. These include the purpose of the research, how you obtained the respondent's telephone number, who is conducting or sponsoring the research, and why someone else should not answer the questions instead of the respondent.

Closing the questionnaire

At the end of your questionnaire, you need to explain clearly what you want the respondent to do with their completed questionnaire. It is usual to start this section by thanking the respondent for completing the questionnaire, and restating the contact's name, email address and telephone number for any queries they may have from the covering email (Figure 11.2) or letter (Figure 11.3). Sometimes, as in Box 11.14, you may wish to make a summary of your research findings available to respondents. If you do make this offer, don't forget to provide the summary!

For postal, and delivery and collection questionnaires, you should restate the date by which you would like it returned and how and where to return it. A template for postal questionnaires, which can be adapted for other questionnaire mediums is given in the next paragraph:

Thank you for taking the time to complete this questionnaire. If you have any queries, please do not hesitate to contact [your name] by telephoning [contact work/university telephone number with answer machine/voice mail] or emailing [work/university email address].

Please return the completed questionnaire by [date] in the envelope provided to:

[your name]

[your address]

11.7 Pilot testing

Prior to using your questionnaire to collect data it should be pilot tested with respondents who are similar to those who will actually complete it. The purpose of the **pilot test** is to refine the questionnaire so that respondents will have no problems in answering the questions and there will be no problems in recording the data. In addition, it will enable you to obtain some assessment of the questions' validity and the likely reliability of the data that will be collected both for individual questions and, where appropriate, scales comprising a number of rating scale questions. Preliminary analysis using the pilot test data can be undertaken to ensure that the data collected will enable your investigative questions to be answered.

Initially, you should ask an expert or group of experts to comment on the suitability of your questions. As well as allowing suggestions to be made on the structure of your questionnaire, this will help establish content validity and enable you to make necessary amendments prior to pilot testing using a group as similar as possible to the final population in your sample and the same distribution method. For any research project there is a temptation to skip the pilot testing. However, this is a false economy and, even if you have limited time, it is important to try out your questionnaire (Bell and Waters 2018). Without a trial run, you have no way of knowing whether your questionnaire will succeed.

The number of people with whom you pilot your questionnaire and the number of pilot tests you conduct will be dependent on your research question(s), your objectives, the size of your research project, the time and money resources you have available, and how well you have initially designed your questionnaire. Where surveys are particularly important, such as referenda and national censuses, there will be numerous field trials, starting with individual questions (Box 11.9) and working up to larger and more rigorous pilots of later drafts.

For smaller-scale surveys you are unlikely to have sufficient financial or time resources for large-scale field trials. However, it is still important that you pilot test your questionnaire. The number of people you choose should be sufficient to include any major variations in your population that you feel are likely to affect responses. For most student questionnaires this means that the minimum number for a pilot is 10 (Fink 2016), although for large surveys between 100 and 200 responses is usual (Dillman et al. 2014). Occasionally you may be extremely pushed for time. In such instances it is better to pilot test the questionnaire using friends or family than not at all! This will provide you with at least some idea of your questionnaire's **face validity**: that is, whether the questionnaire appears to make sense.

As part of your pilot, you should check each completed pilot questionnaire to ensure that respondents have had no problems understanding or answering questions and have followed all instructions correctly (Fink 2016). Their responses will provide you with an idea of the reliability and suitability of the questions (Box 11.15). For self-completed questionnaires, additional information about problems can be obtained by giving respondents a further short questionnaire. Bell and Waters (2018) suggest you should use this to find out:

- how long the questionnaire took to complete;
- the clarity of instructions;
- which, if any, questions were unclear or ambiguous;
- which, if any, questions the respondents felt uneasy about answering;
- whether in their opinion there were any major topic omissions;
- whether the layout was clear and attractive;
- any other comments.

Researcher-completed questionnaires need to be tested with the respondents for all these points other than layout. One way of doing this is to form an assessment as each questionnaire progresses. Another is to ask any research assistants you are employing.



Box 11.15 Focus on student research

Pilot testing a questionnaire

Zaineb pilot tested her questionnaire with 10 people who had similar characteristics to her potential

22. How would you describe your current relationship status?

- | | |
|---|--|
| single, never married married or domestic partnership widowed divorced | <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> |
|---|--|

None of these, I'm separated!

respondents. When looking at the completed questionnaires she noticed that two of her respondents had amended question 22 on marital status.

On this basis, Zaineb added another possible response 'separated' to question 22.

However, you can also check by asking the respondent additional questions at the end of their questionnaire. In addition, you will need to pilot test the questionnaire with the research assistants to discover whether:

- there are any questions for which visual aids should have been provided;
- they have difficulty in finding their way through the questionnaire;
- they are recording answers correctly.

Once you have completed pilot testing you should email or write to these respondents thanking them for their help.

11.8 Distributing the questionnaire

Having designed, pilot tested and amended your questionnaire, and selected your sample, you are ready to distribute it and collect data. Within business and management research reports, it is often not clear whether respondents felt compelled to respond to the questionnaire (Baruch and Holtom 2008). Respondents' feelings of compulsion are usually signified by stating the questionnaire was 'administered', whereas non-compulsion is signified by phrases such as 'invited to fill out a questionnaire voluntarily' or 'voluntary response'. In collecting data using your questionnaire it is important that you abide by your university's or professional body's code of ethics (Sections 6.5 and 6.6). Although, when a respondent answers questions and returns their questionnaire, they are giving their implied consent, they have rights just like all research participants.

Inevitably you will need to gain access to your sample (Sections 6.2 to 6.4) and attempt to maximise the response rate. We have already discussed likely response rates for different of questionnaire medium in relation to probability sampling and sample sizes (Section 7.5). In addition, a large number of studies have been conducted to assess the impact of different strategies for increasing the response to postal questionnaires; many of which also can be applied to inline questionnaires. Fortunately, the findings of these studies have been analysed and synthesised by Edwards et al. (2002), Anseel et al. (2010) and Mellahi and Harris (2016). As you can see from Table 11.5, response rates can be

Chapter 11 Collecting primary data using questionnaires

Table 11.5 Relative impact of strategies for raising postal questionnaire response rates

| Strategy | Relative impact |
|--|-----------------------|
| Incentives | |
| <i>Monetary incentive v. no incentive</i> | Very high |
| <i>Incentive included with questionnaire v. incentive on questionnaire return</i> | High |
| <i>Non-monetary incentive (such as free report) v. no incentive</i> | Low |
| Length | |
| <i>Shorter questionnaire v. longer questionnaire</i> | Very high |
| Appearance | |
| <i>Brown envelope v. white envelope</i> | High but variable |
| <i>Coloured ink v. standard</i> | Medium |
| <i>Folder or booklet v. stapled pages</i> | Low |
| <i>More personalised (name, hand signature etc.) v. less personalised</i> | Low |
| <i>Coloured questionnaire v. white questionnaire</i> | Very low |
| <i>Identifying feature on the return v. none</i> | Very low but variable |
| Distribution | |
| <i>Recorded delivery v. standard delivery</i> | Very high |
| <i>Stamped return envelope v. business reply or franked</i> | Medium |
| <i>First class post outwards v. other class</i> | Low |
| <i>Sent to work address v. sent to home address</i> | Low but variable |
| <i>Pre-paid return v. not pre-paid</i> | Low but variable |
| <i>Stamped outward envelope v. franked</i> | Negligible |
| <i>email v. paper (within organisations and providing all use email regularly)</i> | Medium |
| Contact | |
| <i>Pre-contact (advanced notice) v. no pre-contact</i> | Medium |
| <i>Follow-up v. no follow-up</i> | Medium |
| <i>Postal follow-up including questionnaire v. postal follow-up excluding questionnaire</i> | Medium |
| <i>Pre-contact by telephone v. pre-contact by post</i> | Low |
| <i>Mention of follow-up contact v. none</i> | Negligible |
| Content | |
| <i>More interesting/relevant v. less interesting/relevant topic</i> | Very high |
| <i>User-friendly language v. standard</i> | Medium |
| <i>Demographic and behaviour questions only v. demographic, behaviour and attitude questions</i> | Medium |
| <i>More relevant questions first v. other questions first</i> | Low |
| <i>Most general question first v. last</i> | Low |
| <i>Sensitive questions included v. sensitive questions not included</i> | Very low |
| <i>Demographic questions first v. other questions first</i> | Negligible |
| <i>'Don't know' boxes included v. not included</i> | Negligible |
| Origin | |
| <i>University sponsorship as a source v. other organisation</i> | Medium |
| <i>Sent by more senior or well-known person v. less senior or less well-known</i> | Low but variable |
| Communication | |
| <i>Explanation for not participating requested v. not requested</i> | Medium |
| <i>Confidentiality/anonymity stressed v. not mentioned</i> | Medium |
| <i>Choice to opt out from study offered v. not given</i> | Low |

| Strategy | Relative impact |
|--|------------------|
| Instructions given v. <i>not given</i> | Low but variable |
| <i>Benefits to respondent stressed</i> v. other benefits | Very low |
| Benefits to sponsor stressed v. other benefits | Negligible |
| Benefits to society stressed v. other benefits | Negligible |
| Response deadline given v. no deadline | Negligible |

Note: Strategies in italics increase response rates relative to those in normal font

Source: Developed from Anseel et al. 2010; Edwards et al. 2002; Mellahi and Harris 2016

improved by careful attention to a range of factors, including visual presentation, length, content, distribution methods and associated communication as well as being clearly worded. However, organisations and individuals are increasingly being bombarded with requests to respond to questionnaires and so may be unwilling to complete your questionnaire (Chidlow et al. 2015). The techniques you use to help to maximise responses will inevitably be dependent, at least in part, on the way in which your questionnaire is distributed. It is the processes associated with delivering each of the six questionnaire mediums, that we now consider.

Online questionnaires

For both Web and mobile questionnaires, it is important to have a clear timetable that identifies the tasks to be done and the resources needed. A good response is dependent on the recipient being motivated to answer the questionnaire and return it. Although the covering email and visual appearance will help to ensure a high level of response, it must be remembered that, unlike postal and delivery and collection questionnaires, you and your respondent may see different images displayed on their screens. It is therefore crucial that your cloud-based software can optimise the questionnaire for different displays and, ensure the questionnaire design is clear across all display media (Dillman et al. 2014).

Web and mobile questionnaires are usually delivered via a Web link. This normally uses email or a Web page to display the hyperlink to the questionnaire and is dependent on having a potential respondents' email addresses. Such electronic database lists can be purchased from specialist organisations (Section 7.4). Online panel companies can also be paid to deliver your questionnaire to a specified quota of respondents (Section 7.9). When using the Internet to distribute questionnaires, you should abide by generally acceptable uses (**netiquette**) and, minimise the likelihood of the email invitation being flagged as spam by email filters (Box 11.16).

For within-organisation research, questionnaires can be easily delivered as a hyperlink within an email to employees, provided all of the sample have access to it and use email. If you choose to use email with a direct hyperlink to the questionnaire, we suggest that you:

- 1 Contact recipients yourself by email (or use an online panel company) and advise them to expect a questionnaire – a pre-survey contact (Section 6.5).
- 2 Include the hyperlink to the questionnaire in the email invitation (Figure 11.2). You should make sure that this will arrive when recipients are likely to be receptive. For most organisations Fridays and days surrounding major public holidays have been shown to be a poor time.



Box 11.16 Checklist

Abiding by netiquette and minimising spam flagging

Netiquette

- ✓ Ensure emails and postings to user groups are relevant and that you do not send junk emails (spam).
- ✓ Remember that invitations to participate sent to over 20 user groups at once are deemed as unacceptable by many net vigilantes and so you should not exceed this threshold.
- ✓ Avoid sending your email to multiple mailing lists as this is likely to result in individuals receiving

multiple copies of your email (this is known as cross-posting).

- ✓ Avoid the use of email attachments as these can contain viruses.
- Minimising spam
- Send individual emails rather than bulk emailing options.
- Avoid using the 'Cc.' and 'Bcc.' recipient fields.
- Avoid words such as 'offer', 'free', 'cash', 'win', 'promotion', 'prize' and similar in your email invitation.
- Test messages with a spam test analyser.

Sources: Developed from Hewson et al. 2003; Dillman et al. 2014.

- 3 Summarise the purpose of the research and include an explicit request for the respondent's consent in the welcome screen at the start of the questionnaire (Box 11.17).
- 4 Email a first follow-up one week after the initial email invitation to all recipients. This should thank early respondents and remind non-respondents to answer (a copy of the hyperlink should be included again).
- 5 Email a second follow-up to those who have not responded after three weeks. This should include the information that was in the initial invitation including the hyperlink. The email invitation should be amended to further emphasise the importance of completing the questionnaire.
- 6 Also use a third follow-up if time allows or your response rate is low.
- 7 When the respondent completes the questionnaire, their responses will be saved automatically. However, you may need to select the online survey tool option that prevents multiple responses from one respondent.

Alternatively, the questionnaire can be advertised online or in printed media and potential respondents invited to access the questionnaire by clicking on a hyperlink or scanning a QR (quick response) code using their tablet or mobile phone. Adopting either approach observes netiquette (Box 11.16) and means that respondents can remain anonymous. The stages involved are:

- 1 Ensure that a website has been set up that explains the purpose of the research and has the hyperlink to the questionnaire (this takes the place of the covering email).
- 2 Advertise the research website widely using a range of media (for example, an email pre-survey contact or a banner advertisement on a page that is likely to be looked at by the target population) and highlight the closing date.
- 3 When respondents complete the questionnaire, their responses will be saved automatically. However, you will need to select the online survey tool option that prevents multiple responses from one respondent.

Response rates from web advertisements and QR codes are likely to be very low, and there are considerable problems of non-response bias as the respondent has to take extra



Box 11.17 Focus on student research

Request for respondent's consent in an online questionnaire

Ana had decided to collect her data using an online questionnaire. She emailed potential respondents explaining the purpose of her research and requesting

their help. At the end of her email, she included a hyperlink to the online questionnaire created in Qualtrics.

The first page of Ana's online questionnaire included a summary of the main messages in her email. This was followed by a formal request to the respondent for their consent, which stressed that the decision to participate was entirely voluntary and that they could withdraw at any time.

Thank you for your interest in my research. Before you start the questionnaire, I need to make sure you know what my research is about, what your involvement will be, and for you to confirm that you agree to take part.

By agreeing to take part in this research you are stating that you understand the following:

- I am participating in a research study;
- I have been given an explanation of the research / am about to participate in and I know what is involved in my participation;
- my participation in this research is voluntary and I am free to withdraw at any time without giving any reason;
- my identity cannot be linked to my data and all information I give remains anonymous;
- if I have any questions about the research I can contact Ana on anal23@anytown.edu.

Do you agree to take part?

Yes No

Source: Copyright © 2021 Qualtrics LLC. Used With Permission.

steps to locate and complete the questionnaire. Consequently, it is likely to be very difficult to obtain a sample from which you might generalise. This is not to say that this approach should not be used as it can, for example, enable you to contact difficult-to-access groups. It all depends, as you would expect us to say, on your research question and objectives!

SMS questionnaires

SMS (text) questionnaires are used typically to obtain feedback immediately after an event such as a purchase delivery, meal at a restaurant or similar. For these questionnaires the introduction is invariably shorter as a maximum of 918 characters can be sent by text message. SMS questionnaires are usually sent using cloud-based survey software being delivered directly to recipients' mobile phones comprising very few questions (preferably three or less). Questions are delivered one question at a time; subsequent questions only being delivered if a question is answered. If you choose to use an SMS questionnaire, we suggest that you:

- 1 Obtain and import a list of potential respondents' mobile phone numbers into the cloud-based software and schedule the distribution of the questionnaire at a time when you believe they will be able to take part.

- 2 For the first question, text recipients and ask if they would be willing to take part in the research.
- 3 Subsequent questions will be sent by text message immediately after the respondent answers the question.
- 4 On receipt of a response to the last question, ensure the software is set up to text the respondent and thank them for taking part.

Postal questionnaires

For postal questionnaires, it is important to have a concise and clear covering letter and good visual presentation to help to ensure a high level of response. As with online questionnaires, a clear timetable and well-executed administration process are important (Box 11.18).

Our advice for postal questionnaires (developed from De Vaus 2014) can be split into six stages:

- 1 Ensure that questionnaires and letters are printed, and envelopes addressed.
- 2 Contact recipients by post, telephone or email and advise them to expect a questionnaire – a pre-survey contact (Section 6.5). This stage is often omitted for cost reasons.
- 3 Post the survey with a covering letter and a return envelope. You should make sure that this will arrive when recipients are likely to be receptive. For most organisations Fridays and days surrounding major public holidays have been shown to be a poor time.
- 4 Post the first follow-up one week after posting out the survey to all recipients. This should take the form of a postcard designed to thank early respondents and to remind rather than to persuade non-respondents.
- 5 Post the second follow-up to people who have not responded after three weeks. This should contain another copy of the questionnaire, a new return envelope and a new covering letter. The covering letter should be reworded to emphasise further the importance of completing the questionnaire. For anonymous questionnaires a second follow-up will not be possible, as you should not be able to tell who has responded!
- 6 Also use a third follow-up if time allows or your response rate is low. For this it may be possible to use ‘signed for’ delivery (post), telephone calls or even call in person to emphasise the importance of responding.

Additionally, De Vaus (2014) advises placing a unique identification number on each questionnaire, which is recorded on your list of recipients. This makes it easy to check and follow up non-respondents and, according to Dillman et al. (2014) and Edwards et al. (2002), has little, if any, effect on response rates. However, identification numbers should not be used if you have assured respondents that their replies will be anonymous!

Delivery and collection questionnaires

For delivery and collection questionnaires either you or a research assistant will deliver and call to collect the questionnaire. It is therefore important that your covering letter states when the questionnaire is likely to be collected. As with postal questionnaires, follow-ups can be used, calling at a variety of times of day and on different days to try to catch the respondent.



Box 11.18 Focus on management research

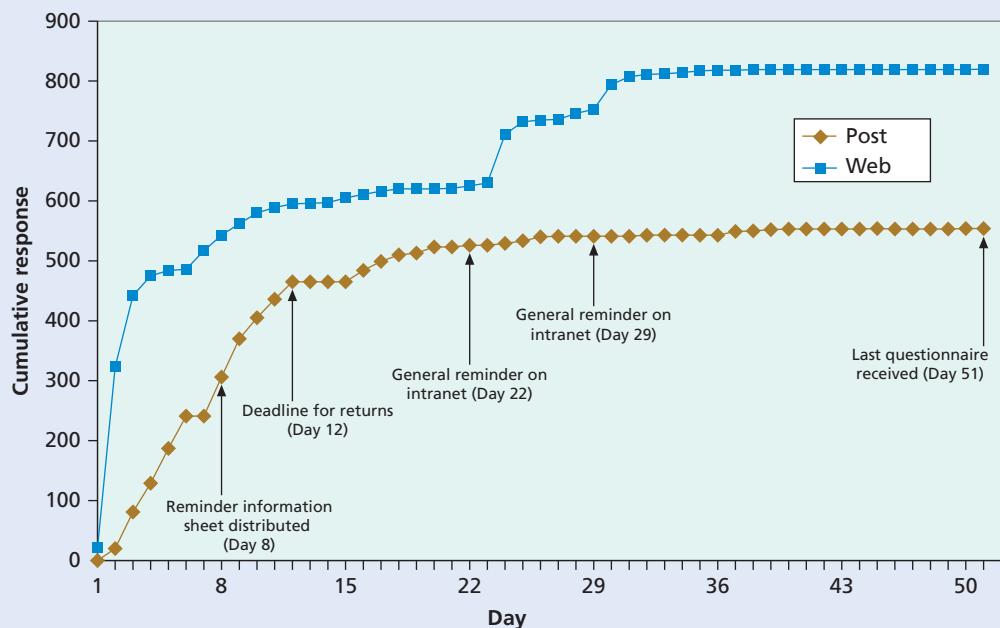
Questionnaire administration

Mark undertook an attitude survey of employees in a large organisation using a questionnaire. Within the organisation, 50 per cent of employees received an online questionnaire by a hyperlink in an email, the remaining 50 per cent receiving a postal questionnaire.

General information regarding the forthcoming survey was provided to employees using the staff intranet, the normal method for such communications. Subsequently, each employee received five personal contacts including the questionnaire:

- One week before the questionnaire was delivered a pre-survey notification email or letter, jointly from the organisation's Chief Executive and Mark, was delivered in the same manner as the potential respondent would receive their questionnaire.
- Covering email or letter and hyperlink or questionnaire to online questionnaire.
- Personal follow-up/reminder designed as an information sheet re-emphasising the deadline for returns at the end of that week.
- First general reminder (after the deadline for returns) posted on the staff intranet.
- Second general reminder (after the deadline for returns) posted on the staff intranet.

The following graph records the cumulative responses for both the online and postal questionnaire, emphasising both the impact of deadlines, follow-up/ reminders and the length of time required (over 7 weeks) to collect all the completed questionnaires.



Cumulative questionnaire returns for Web and post questionnaires

Source: Unpublished data

A variation of this process that we have used widely in organisations allows for delivery and collection of questionnaires the same day and eliminates the need for a follow-up. The stages are:

- 1 Ensure that all questionnaires and covering letters are printed and a collection box is ready.
- 2 Contact respondents by email, internal post, telephone or text/SMS advising them to attend a meeting or one of a series of meetings to be held (preferably) in the organisation's time (Section 6.5).
- 3 At the meeting or meetings, distribute the questionnaire with a covering letter to each respondent.
- 4 Introduce the questionnaire, stress its anonymous or confidential nature and that participation is voluntary.
- 5 Ensure that respondents place their questionnaires in a collection box before they leave the meeting.

Although this adds to costs, as employees are completing the questionnaire in work time, response rates as high as 98 per cent are achievable!

Telephone questionnaires

The quality of data collected using telephone questionnaires will be affected by the researcher's competence to conduct interviews. This is discussed in Sections 10.5 and 10.7. Once your sample has been selected, you need to:

- 1 Ensure that all questionnaires are printed or, for CATI, that the survey tool has been programmed and tested.
- 2 Where possible and resources allow, contact respondents by email, post or telephone advising them to expect a telephone call (Section 6.5).
- 3 Telephone each respondent, recording the date and time of call and whether or not the questionnaire was completed. You should note any specific times that have been arranged for call-backs. For calls that were not successful you should note the reason, such as no reply or telephone disconnected.
- 4 For unsuccessful calls where there was no reply, try three more times, each at a different time and on a different day, and note the same information.
- 5 Make call-back calls at the time arranged.

Face-to-face questionnaires

Conducting face-to-face questionnaires uses many of the skills required for in-depth and semi-structured interviews (Section 10.7). Issues such as researcher appearance and preparedness are important and will affect the response rate (Section 10.5). However, once your sample has been selected you need to:

- 1 Ensure that all questionnaires are printed or, for CAPI, that the survey tool has been programmed and tested.
- 2 Contact respondents by email, post or telephone advising them to expect a researcher to call within the next week. This stage is often omitted for cost reasons.
- 3 (For large-scale surveys) Divide the sample into assignments that are of a manageable size (50–100) for one research assistant.

- 4 Contact each respondent or potential respondent in person, recording the date and time of contact and whether or not the questionnaire was completed. You should note down any specific times that have been arranged for return visits. For contacts that were not successful, you should note down the reason.
- 5 Try unsuccessful contacts at least twice more, each at a different time and on a different day and note down the same information.
- 6 Visit respondents at the times arranged for return visits.

11.9 Summary

- Questionnaires collect data by asking people to respond to exactly the same set of questions. They are often used as part of a survey strategy to collect descriptive and explanatory data about facts/demographics, attitudes/opinions and behaviours/events. Data collected are normally analysed quantitatively.
- Your choice of questionnaire will be influenced by your research question(s) and objectives and the resources that you have available. There are four common self-completed mediums: online, SMS, postal, delivery and collection; and two common researcher-completed mediums: telephone and face-to-face.
- Prior to designing a questionnaire, you must establish precisely what data you need to collect to answer your research question(s) and to meet your objectives. One way of helping to ensure that you collect these data is to use a data requirements table.
- The validity and reliability of the data you collect and the response rate you achieve depend largely on the design of your questions, the structure of your questionnaire and the rigour of your pilot testing.
- When designing your questionnaire, you should consider the wording of individual questions prior to the order in which they appear. Questions can be divided into open and closed. The six types of closed questions are list, category, ranking, rating scale, quantity and matrix.
- Responses for closed questions in online and SMS questionnaires are coded automatically within the cloud-based survey software. For other questionnaire distribution modes closed questions should, wherever possible, be pre-coded on your questionnaire to facilitate data input and subsequent analyses.
- The order and flow of questions in the questionnaire should be logical to the respondent. This can be assisted by filter questions and linking phrases.
- The visual appearance of the questionnaire should be attractive, easy to read and the responses easy to fill in.
- Questionnaires must be introduced carefully to the respondent to ensure a high response rate. For self-completed questionnaires this should take the form of an invitation email or letter and for online questionnaires also summarised in the welcome screen; for researcher-completed questions it will be done by the researcher or a research assistant.
- All questionnaires should be pilot tested prior to their distribution, to assess the validity and likely reliability of the questions.
- Distribution of questionnaires is dependent on the completion mode and questionnaire medium.

Self-check questions

Help with these questions is available at the end of the chapter.

- 11.1** In what circumstances would you choose to use a delivery and collection questionnaire rather than an online questionnaire? Give reasons for your answer.
- 11.2** The following questions have been taken from a questionnaire about flexibility of labour.

i Do you agree or disagree with the use of zero hours contracts by employers? (Please tick appropriate box)

Strongly agree 4

Agree 3

Disagree 2

Strongly disagree 1

ii Have you ever been employed on a zero hours contract? (Please tick appropriate box)

Yes 1

No 2

Not sure 3

iii What is your marital status? (Please tick appropriate box)

Single 1

Married or living in long-term relationship 2

Widowed 3

Divorced 4

Other 5

(..... Please describe)

iv Please describe what you think would be the main impact on employees of a zero hours contract

For each question identify:

- a** the type of data variable for which data are being collected;
b the type of question.

You should give reasons for your answers.

- 11.3** You are undertaking research on the use of children's book clubs by householders within mainland Europe. As part of this, you have already undertaken in-depth interviews with households who belong, and do not belong, to children's book clubs. This, along with a literature review, has suggested a number of investigative questions from which you start to construct a table of data requirements.

- a** For each investigative question listed, decide whether you will need to collect factual/demographic, attitude/opinion or behaviour/event data.
- b** Complete the 'variable(s) required' and 'detail in which data measured' in the table of data requirements for each of the investigative questions already listed. (You may embellish the scenario to help in your choice of variables required and the detail in which the data will be measured as you feel necessary, but you do not have to explore the relation to theory and key concepts in the literature.)

| Research objective: To establish mainland Europe's householders' opinions about children's book clubs. | | |
|--|----------------------|-------------------------------|
| Type of research: Predominantly descriptive, although wish to explain differences between householders. | | |
| Investigative questions | Variable(s) required | Detail in which data measured |
| A Do householders think that children's book clubs are a good or a bad idea? | | |
| B What things do householders like most about children's book clubs? | | |
| C Would householders be interested in an all-ages book club? | | |
| D How much per year do households spend on children's books? | | |
| E Do households' responses differ depending on (i) number of children? (ii) whether already members of a children's book club? | | |

- 11.4** Design pre-coded or self-coded questions to collect data for each of the investigative questions in Question 11.3. Note that you will need to answer self-check question 11.3 first.
- 11.5** What issues will you need to consider when translating the questions you designed in answer to question 11.4?
- 11.6** You work for a major consumer research bureau that has been commissioned by 11 major UK companies to design, deliver and analyse the data collected using a telephone questionnaire. The purpose of this questionnaire is to describe and explain relationships between adult consumers' lifestyles, opinions and purchasing intentions. Write the introduction to this telephone questionnaire, to be read by a research assistant to each respondent. You may embellish the scenario and include any other relevant information you wish.
- 11.7** You have been asked by a well-known national charity 'Work for All' to carry out research into the effects of long-term unemployment throughout the UK. The charity intends to use the findings of this research as part of a major campaign to highlight public awareness about the effects of long-term unemployment. The charity has drawn up a list of names and postal addresses of people who are or were long-term unemployed with whom they have had contact over the past six months. Write a covering letter to accompany the postal questionnaire. You may embellish the scenario and include any other relevant information you wish.
- 11.8** You have been asked to give a presentation to a group of managers at a ground and air source heating company to gain access to undertake your research. As part of the presentation, you outline your methodology, which includes pilot testing the questionnaire. In the ensuing question and answer session, one of the managers asks you to justify the need for a pilot study, arguing that 'given the time constraints the pilot can be left out'. List the arguments that you would use to convince him that pilot testing is essential to your methodology.

Review and discussion questions

- 11.9** Obtain a copy of a 'customer questionnaire' from a department store or restaurant. For each question on the questionnaire establish whether it is collecting factual/demographic, attitude/opinion or behaviour/event data. Do you consider any of the questions are potentially misleading? If yes, how do you think the question could be improved? Discuss the answer to these questions in relation to your questionnaire with a friend.
- 11.10** Visit the website of a cloud-based survey design, data collection and analysis software provider. A selection of possible providers can be found by typing 'online questionnaire provider' or 'online survey provider' into the Google search engine. Use the online survey tool to design a simple questionnaire. To what extent does the questionnaire you have designed meet the requirements of the checklists in Boxes 11.10, 11.12 and 11.13?
- 11.11** Visit your university library or use the Internet to view a copy of a report for a recent national government survey in which you are interested. If you are using the Internet, the national government websites listed in Table 8.1 are a good place to start. Check the appendices in the report to see if a copy of the questionnaire used to collect the data is included. Of the types of question – open, list, category, ranking, rating, quantity and grid – which is most used, and which is least frequently used? Note down any that may be of use to you in your research project.



Progressing your research project

Using questionnaires in your research

- Return to your research question(s) and objectives. Decide on how appropriate it would be to use questionnaires as part of your research strategy. If you do decide that this is appropriate, note down the reasons why you think it will be sensible to collect at least some of your data in this way. If you decide that using a questionnaire is not appropriate, justify your decision.
- If you decide that using a questionnaire is appropriate, re-read Chapter 7 on sampling and, in conjunction with this chapter (Table 11.1 is a good place to start), decide which of the six main questionnaire mediums will be most appropriate. Note down your choice of questionnaire and the reasons for this choice.
- Construct a data requirements table and work out precisely what data you need to answer your investigative questions. Remember that you will need to relate your investigative questions and data requirements to both theory and key concepts in the literature you have reviewed

and any preliminary research you have already undertaken.

- Design the separate questions to collect the data specified in your data requirements table. Wherever possible, try to use closed questions and to adhere to the suggestions in the question wording checklist (Box 11.10). Read Sections 12.2 and 12.3 and pre-code questions on the questionnaire to aid subsequent analysis whenever possible.
- Order your questions to make reading the questions and filling in the responses as logical as possible to the respondent. Wherever possible, try to adhere to the checklist for layout (Boxes 11.12 and 11.13). Remember that researcher-completed questionnaires will need instructions for the researcher or research assistant.
- Write the introduction to your questionnaire and, where appropriate, a covering letter.
- Pilot test your questionnaire with as similar a group as possible to the final group in your sample. Pay special attention to issues of validity and reliability.
- Distribute your questionnaire paying attention to netiquette for online questionnaires (Box 11.16) and remember to send out a follow-up survey to non-respondents whenever possible.
- Use the questions in Box 1.4 to guide your reflective diary entry.

References

- American Psychological Association (2021) *APA PsycTESTS*. Available at <https://www.apa.org/pubs/databases/psychtests/index> [Accessed 16 Nov. 2021].
- Anseel, F., Lievens, F., Schollaert, E. and Choragwicka, B. (2010) 'Response rates in organizational science, 1995–2008: A meta-analytic review and guidelines for survey researchers', *Journal of Business Psychology*, Vol. 25, pp. 335–49.
- Baruch, Y. and Holtom, B.C. (2008) 'Survey response rate levels and trends in organizational research', *Human Relations*, Vol. 61, pp. 1139–60.
- Bell, J. and Waters, S. (2018) *Doing Your Research Project* (7th edn). Maidenhead: Open University Press.
- Bloomberg, B., Cooper, D.R. and Schindler, P.S. (2014) *Business Research Methods* (4th edn). Boston, MA and Burr Ridge, IL: McGraw-Hill.
- Bourque, L.B. and Clark, V.A. (1994) 'Processing data: The survey example', in M.S. Lewis-Beck (ed.) *Research Practice*. London: Sage, pp. 1–88.
- Bowling, N.A., Gibson, A.M., Houpt, J.W. and Brower, C.K. (2021) 'Will the questions ever end? Person-level increases in careless responding during questionnaire completion', *Organizational Research Methods*, Vol. 24, No. 4, pp. 718–38.
- Bruner II, G.C. (2021) *Marketing Scales Handbook: Multi Item Measures for Consumer Insight Research Volume 11*. Fort Worth, TX: GBII Productions.
- Chidlow, A., Ghauri, P.N., Yeniyurt, S. and Cavusgil, S.T. (2015) 'Establishing rigour in mail-survey procedures in international business research', *Journal of World Business*, Vol. 50, pp. 26–35.
- De Vaus, D.A. (2014) *Surveys in Social Research* (6th edn). Abingdon: Routledge.
- DeVellis, R.F. (2017) *Scale Development: Theory and Applications* (4th edn). Thousand Oaks, CA: Sage.
- Dillman, D.A., Smyth, J.D. and Christian, J.M. (2014) *Internet, Phone, Mail and Mixed Mode Surveys: The Tailored Design Method* (4th edn). Hoboken, NJ: Wiley.
- Edwards, P., Roberts, I., Clarke, M., Di Giuseppe, C., Pratap, S., Wentz, R. and Kwan, I. (2002) 'Increasing response rates to postal questionnaires: Systematic review', *British Medical Journal*, No. 324, May, pp. 1183–91.
- Ekinci, Y. (2015) *Designing Research Questionnaires for Business and Management Students*. London: Sage.
- Felstead, A., Gallie, D., Green, F. and Henseke, G. (2019) 'Conceiving, designing and trialling a short-form measure of job quality: A proof-of-concept study', *Industrial Relations Journal*, Vol. 50, No. 1, pp. 2–19.
- Field, A. (2018) *Discovering Statistics Using IBM SPSS Statistics* (5th edn). London: Sage.
- Fink, A. (2016) *How to Conduct Surveys* (6th edn). Thousand Oaks, CA: Sage.
- Forbes, S. and Avis, M. (2020) 'Construct creation from research questions', *European Journal of Marketing*, Vol. 54, No. 8, pp. 1817–38.
- Ghauri, P., Grønhaug, K. and Strange, R. (2020) *Research Methods in Business Studies* (5th edn). Cambridge, Cambridge University Press.
- Halkias, G., Florack, A., Diamantopoulos, A. and Palcu, J. (2021, early view) 'Eyes wide shut? Understanding and managing consumers' visual processing of country-of-origin cues', *British Journal of Management*, pp 1–15. DOI: 10.1111/1467-8551.12545.
- Hardy, B. and Ford, L.R. (2014) 'It's not me, it's you: Miscomprehension in surveys', *Organizational Research Methods*, Vol. 17, No. 2, pp. 138–62.

Chapter 11 Collecting primary data using questionnaires

- Hewson, C., Yule, P., Laurent, D. and Vogel, C. (2003) *Internet Research Methods: A Practical Guide for the Social and Behavioural Sciences*. London: Sage.
- Jackson, L.A. and Taylor, M. (2015) 'Revisiting smoking bans in restaurants: Canadian employees' perspectives', *Tourism and Hospitality Research*, Vol. 15, No. 2, pp. 91–104.
- Kozinets, R.V. (2020) *Netnography. The Essential Guide to Qualitative Social Media Research* (3rd edn). London: Sage.
- Louka, P., Maguire, M., Evans, P. and Worrell, M. (2006) "I think that it's a pain in the ass that I have to stand outside in the cold and have a cigarette": Representations of smoking and experiences of disapproval in UK and Greek Smokers', *Journal of Health Psychology*, Vol. 11, No. 3, pp. 441–51.
- Mellahi, K. and Harris, L.C. (2016) 'Response rates in business and management research: an overview of current practice and suggestions for future direction', *British Journal of Management*, Vol. 24, No. 2, pp. 426–37.
- Mitchell, V. (1996) 'Assessing the reliability and validity of questionnaires: An empirical example', *Journal of Applied Management Studies*, Vol. 5, No. 2, pp. 199–207.
- Qualtrics (2021). *Qualtrics Online Survey Software*. Available at <https://www.qualtrics.com/uk/core-xm/survey-software/> [Accessed 27 Nov. 2021].
- Robson, C. and McCartan, K. (2016) *Real World Research: A Resource for Users of Social Research Methods in Applied Settings* (4th edn). Chichester: John Wiley.
- Rogelberg, S.G. and Stanton, J.M. (2007) 'Introduction: Understanding and dealing with organizational survey non-response', *Organizational Research Methods*, Vol. 10, No. 2, pp. 195–209.
- Roster, C.A., Lucianetti, L. and Albaum, G. (2015) 'Exploring slider vs. categorical response formats in web-based surveys', *Journal of Research Practice*, Vol. 11, No. 1, Article D1, p. 15.
- Roth, M. S. and J. B. Romeo (1992). 'Matching product category and country image perceptions: A framework for managing country-of-origin effects', *Journal of International Business Studies*, Vol. 23, pp. 477–497.
- Saunders, M.N.K. (2012) 'Web versus mail: The influence of survey distribution mode on employees' response', *Field Methods*, Vol. 24, No. 1, pp. 56–73.
- Schrauf, R.W. and Navarro, E. (2005) 'Using existing tests and scales in the field', *Field Methods*, Vol. 17, No. 4, pp. 373–93.
- Sekaran, U. and Bougie, R. (2019) *Research Methods for Business* (8th edn). Chichester: Wiley.
- SurveyMonkey (2021) *SurveyMonkey*. Available at www.surveymonkey.com [Accessed 27 Nov. 2021].
- Tharenou, P., Donohue, R. and Cooper, B. (2007) *Management Research Methods*. Melbourne: Cambridge University Press.
- UK Data Service (2021c) *Variable and Question Bank*. Available at <https://discover.ukdataservice.ac.uk/variables> [Accessed 27 Nov. 2021].
- Usunier, J.-C., Van Herk, H. and Lee, J.A. (2017) *International and Cross-Cultural Management Research*. (2nd edn) London: Sage.
- van de Heijden, P. (2017) 'The practicalities of SMS research', *Journal of Marketing Research*, Vol. 59, No. 2, pp. 157–72.
- Vagias, W. M. (2006). *Likert-type scale response anchors*. Clemson International Institute for Tourism and Research Development, Department of Parks, Recreation and Tourism Management, Clemson University. Available at: https://mwcc.edu/wp-content/uploads/2020/09/Likert-Scale-Response-Options_MWCC.pdf [Accessed 19 November 2021].

Further reading

- De Vaus, D.A. (2014) *Surveys in Social Research* (6th edn). Abingdon: Routledge. Chapters 7 and 8 provide a detailed guide to constructing and delivering questionnaires, respectively.
- Dillman, D.A., Smyth, J.D. and Christian J.M. (2014) *Internet, Phone, Mail and Mixed Mode Surveys: The Tailored Design Method* (4th edn). Hoboken, NJ: Wiley. The fourth edition of this classic text contains an extremely detailed and well-researched discussion of how to design and deliver online, telephone and postal-based questionnaires to maximise response rates.
- Hall, J.F. (2021) *Journeys in Survey Research*. Available at <http://surveyresearch.weebly.com/> [Accessed 27 Nov. 2021]. This site contains a wealth of information about the use of questionnaires and has an informative section on survey research practice.

Analyzing Data Quantitatively

Saunders et al. (2023; 572-642)

Chapter 12



Analysing data quantitatively

Learning outcomes

By the end of this chapter, you should be able to:

- recognise different types of data variables and understand the implications for quantitative analyses;
- code data variables and create a data matrix using statistical analysis software;
- explore and present data variables using appropriate tables and graphs;
- describe individual data variables using appropriate statistics;
- examine associations and differences between data variables using appropriate statistics;
- assess the strength of relationships between data variables using appropriate statistics;
- make predictions from data variables using appropriate statistics;
- examine trends in data variables using appropriate statistics;
- progress your research project by analysing data quantitatively.

12.1 Introduction

Virtually any business and management research you undertake is likely to involve some numerical data, or contain data that has or could be quantified, to help you answer your research question(s) and to meet your objectives. Quantitative data refer to all such primary and secondary data and can range from simple counts such as the frequency of occurrences of an advertising slogan to more complex data such as test scores, prices or rental costs. However, to be useful these data need to be analysed and interpreted. Quantitative analysis techniques assist you in this process. They range from creating simple tables or graphs that show the frequency of occurrence and using statistics such as indices to enable comparisons, through establishing statistical relationships between variables, to complex statistical modelling.

Before we begin to analyse data quantitatively, we need to ensure that our data are already quantified or that they are quantifiable and can be transformed into **quantitative data**, that is

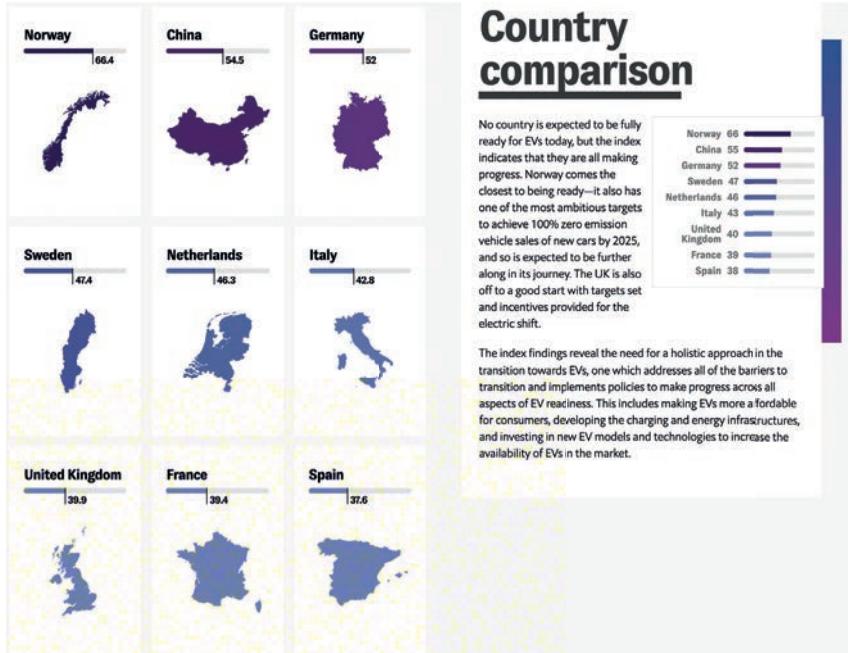


data that can be recorded as numbers and analysed quantitatively. This means that prior to undertaking our analysis, we may need to transform other forms of data (such as text, voice and visual) into sets or categories giving each category a numerical code.

The Economist's rEV Index

The Economist's (2021) rEV Index assesses the UK's Four nations and regions on their readiness for electric vehicles (EVs). It also provides comparative scores for leading EV markets including European countries and China. The index provides an idea of what it means to be ready for EVs from the perspectives of both vehicle manufacturers and users. Using data from reputable international, national and industry sources, the index aims to measure the extent to which different parts of the UK (and other countries) are equipped with policies, infrastructure and consumer attitudes to transfer to EVs.

The index is a scaled score of between 0 and 100, where zero represents the weakest environment for the adoption of EVs, and 100 the strongest. This allows for easy comparison between countries and between regions of the UK. Each of these scaled scores is calculated from 23 different indicators organised into eight pillars that are weighted depending upon their relative importance. For example, anxiety around the



The Economist's rEV Index for the UK

Source: © The Economist (2021)

availability of charging points is the topmost concern in the UK so this pillar weighted most highly. The eight pillars are:

- **rEV inputs**

- affordability (weighted at 15%);
- purchase incentives (weighted at 10%);
- consumer sentiment (weighted at 10%);
- charging infrastructure (weighted at 25%);
- energy infrastructure (weighted at 10%);
- regulations (weighted at 10%).



- **rEV outputs**

- uptake (weighted at 10%);
- availability (weighted at 10%).

The index scores reveal Norway as the most EV ready country with an index of 66, followed by China

(index = 55), and Germany (index = 52), although differences within these countries are not given. Within the UK (index = 40), London is the most EV ready region (index = 48), and the North West and East of England (both with an index of 33) the least ready.

Within quantitative analysis, calculations and diagram drawing are usually undertaken using analysis software ranging from spreadsheets such as Excel™ to more advanced data management and statistical analysis software such as IBM SPSS Statistics™, SAS™, or Stata™. You might also use more specialised survey design and analysis online software such as Qualtrics Research CORE™ and SurveyMonkey™, statistical shareware such as the R Project for Statistical Computing, or content analysis and text mining software such as WordStat™. However, while this means you do not have to be able to draw charts by hand, undertake calculations using a calculator or count frequencies of occurrences of words and phrases by hand, if your analyses are to be straightforward and of any value you need to:

- distinguish between different types of data and recognise the implications for quantitative analysis;
- have prepared your data with quantitative analyses in mind;
- be aware of and know when to use different tables, graphs and statistical analysis techniques.

This is not to say that there is only one possible technique for any analysis situation. As we will see, a range of factors need to be taken into account when selecting the most appropriate graphs, tables, graphs and statistics. Consequently, if you are unsure about which of these to use, you need to seek advice.

This chapter builds on the ideas outlined in earlier chapters about secondary data (Chapter 8) and primary data collection (Chapters 9 to 11), including issues of sample size (Section 7.5). It assumes that you will use a spreadsheet or more advanced statistical analysis software to undertake all but the simplest quantitative analyses. Although it does not focus on a particular analysis software, you will notice in the Focus on student research boxes that many of the analyses were undertaken using widely available software such as Excel and IBM SPSS Statistics. There are numerous statistics books already published that concentrate on specific software packages. These include Dancey and Reidy (2020), Field (2018) or Pallant (2020) on SPSS, and Scherbaum and Shockley (2015) and Winston (2022) on Excel. Likewise, this chapter does not attempt to provide an in-depth discussion of the wide range of graphical and statistical techniques available or cover more complex statistical modelling, as these are already covered elsewhere (Dawson 2017; Hair et al. 2018; Hays 1994). Rather it discusses issues that need to be considered at the planning and analysis stages of your research project and outlines analytical techniques that our students have found of most use for quantitative analysis and interpretation of data. In particular, the chapter is concerned with:

- data types and precision of measurement (Section 12.2);
- preparing data for quantitative analysis (Section 12.3);
- data entry and checking (Section 12.4);
- selecting and interpreting tables and graphs to explore and present data (Sections 12.5 to 12.7);

- selecting and interpreting statistics to describe data (Section 12.8);
- understanding statistical assumptions and hypothesis testing (Section 12.9);
- selecting and interpreting statistics to examine associations and differences (Section 12.10);
- selecting and interpreting statistics to examine relationships (Section 12.11);
- selecting and interpreting statistics to make predictions (Section 12.12);
- selecting and interpreting statistics to examine trends (Section 12.13).

Ideally, these should be considered before obtaining your data. This is equally important for both primary and secondary data analysis, although you obviously have far greater control over the type, format and coding of primary data.

12.2 Data types and precision of measurement

Many business statistics textbooks classify data for quantitative analysis into *data types* according to precision of measurement, often in ascending order of numerical precision (Berman Brown and Saunders 2008; Dancey and Reidy 2020). These different levels of numerical precision dictate the range of techniques available to you for exploration, presentation, description, and examination of your data. They are discussed in more detail in subsequent sections of this chapter.

Understanding the type of data is extremely important when analysing your data quantitatively, for two reasons. First, it is extremely easy with analysis software to generate statistics from your data that are inappropriate for the data type and are consequently of little value (Box 12.1). Second, as we will see in Sections 12.6 to 12.8, the more precise the measurement, the greater the range of analytical techniques available to you. Data that have been collected and coded using a precise numerical measurement can also be regrouped to a less precise level where they can also be analysed (Box 12.2). For example, a student's score in a test could be recorded as the actual mark (discrete data) or as the position in their class (ranked data). By contrast, less precise data cannot be made more precise. Therefore, if you are not sure about the precision of measurement you require, it is usually better to collect data at the highest level of precision possible and to regroup them if necessary.

Categorical data

Data for quantitative analysis can be divided into two distinct groups: categorical and numerical (Figure 12.1). **Categorical data** refer to data whose values cannot be measured numerically but can be either classified into sets (categories) according to the characteristics that identify or describe the variable or placed in rank order (Berman Brown and Saunders 2008). They can be further subdivided into descriptive and ranked. An auto manufacturer might categorise their vehicles' powertrains as petrol, diesel, hybrid and electric. You might classify aspects of an image in terms of the gender of the person depicted and whether or not she or he is smiling. The verbal responses to an open-ended interview question asking participants to describe their journey to work could, once transcribed into text, be used to generate data about their main mode of travel to work. These could be categorised as 'bicycle', 'bus', 'rail', 'car' or 'walk'. Alternatively, you may be looking at particular concepts in illustrations in annual reports such as whether the central figure in each is male or female.



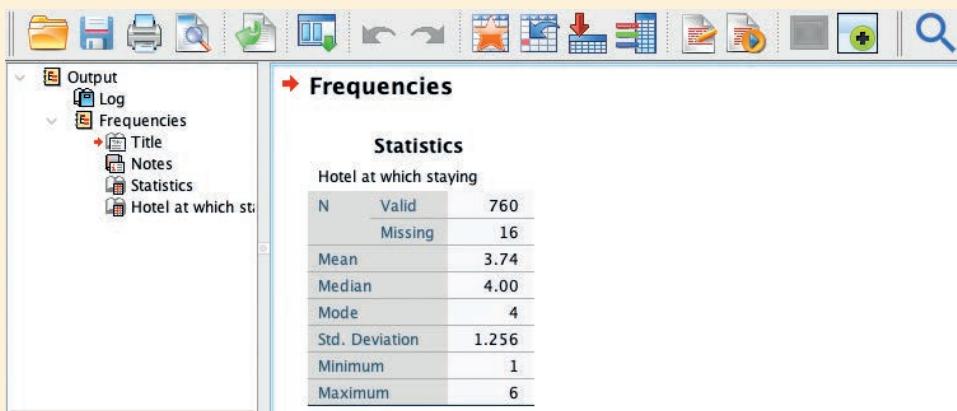
Box 12.1 Focus on student research

The implications of coding and data types for analysis

Pierre's research was concerned with customers' satisfaction for a small hotel group of six hotels. In collecting the data, he had asked 760 customers to indicate the hotel at which they were staying when they completed their Internet questionnaires. When he downloaded his data, the survey design software had

automatically allocated a numerical code to represent the hotel, named the variable and labelled each of the codes. The code labels for the six hotels were:

| Hotel at which staying | Code |
|------------------------|------|
| Amsterdam | 1 |
| Antwerp | 2 |
| Eindhoven | 3 |
| Nijmegen | 4 |
| Rotterdam | 5 |
| Tilburg | 6 |



In his initial analysis, Pierre used the analysis software to calculate descriptive statistics for every data variable, including the variable 'Hotel'. These included the mean, (3.74), median and mode (both 4 – the code for Nijmegen), standard deviation (1.256), minimum value (1 – the code for Amsterdam), the maximum value (6 – the

code for Tilburg). Looking at his computer screen, Pierre wondered which of the three averages (mean, median or mode) was the most appropriate. He decided that, as his data were descriptive (nominal), the mode would be the most useful appropriate. The median, mean and standard deviation statistics were inappropriate for this data type.



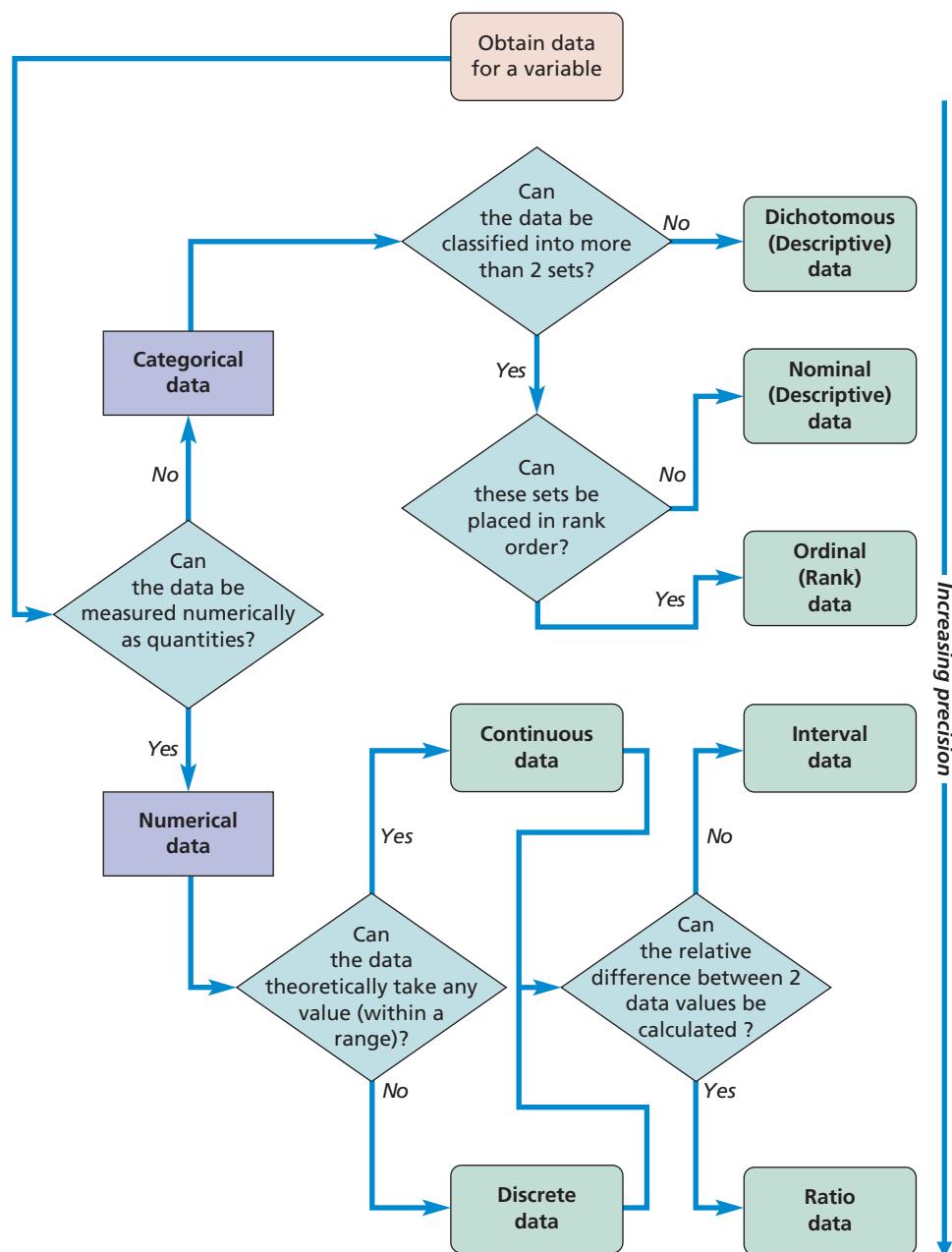
Box 12.2 Focus on student research

Precision of data measurement

As part of a marketing questionnaire, Rashid asked individual customers to rank up to five features of a new product in order of importance to them. Data collected

were, therefore, categorical and ranked (ordinal). Initial analyses made use of these ranked data. Unfortunately, a substantial minority of customers had ticked, rather than ranked, those features of importance to them.

All responses that had been ranked originally were therefore re-coded to 'of some importance'. This reduced the precision of measurement from ranked (ordinal) to descriptive (nominal) but enabled Rashid to use all responses in the subsequent analyses.

**Figure 12.1** Defining the data type

Nominal or descriptive data

Although the sources of these data differ, they are all known as **nominal data** or **descriptive data** as it is impossible to define such a category numerically or rank it. Rather, these data simply count the number of occurrences in each category of a variable. For virtually all analyses the categories should be unambiguous and discrete; in other words, having one particular feature, such as a vehicle being electric, excludes it from being in all other

powertrain categories. This prevents questions arising regarding which category an individual case belongs to. Although these data are purely descriptive, you can count them to establish which category has the most and whether cases are spread evenly between categories. Some statisticians (and statistics) also separate descriptive data where there are only two categories. These are known as **dichotomous data**, as the variable is divided into two categories, such as the variable ‘result’ being divided into ‘pass’ and ‘fail’.

Ordinal or rank data

Ordinal (or rank) data are a more precise form of categorical data. In such instances you know the relative position of each case within your data set, although the actual numerical measures (such as scores) on which the position is based are not recorded (Box 12.2). A researcher exploring an organisation’s online communication may rank each of that organisation’s tweets over a three-month period as positive, neutral or negative. You might rank individual festival goers’ photographs uploaded to the festival website in terms of the prominence given to music related aspects; categorising this as high, medium, low or absent. Similarly, a questionnaire asking a rating or scale question, such as how strongly a respondent agrees with a statement, also collects ranked (ordinal) data. Despite this, some researchers argue that, where such data are likely to have similar size gaps between data values, they can be analysed as if they were numerical interval data (Blumberg et al. 2014).

Numerical data

Numerical data are those whose values are measured or counted numerically as quantities (Berman Brown and Saunders 2008). This means that numerical data are more precise than categorical as you can assign each data value a position on a numerical scale. It also means that you can analyse these data using a far wider range of statistics. There are two possible ways of subdividing numerical data: into interval or ratio data and, alternatively, into continuous or discrete data (Figure 12.1).

Interval and ratio data

If you have **interval data** you can state the difference or ‘interval’ between any two data values for a particular variable, but you cannot state the relative difference. This means that values on an interval scale can meaningfully be added and subtracted, but not multiplied and divided. The Celsius temperature scale is a good example of an interval scale. Although the difference between, say, 20°C and 30°C is 10°C, it does not mean that 30°C is one and a half times as warm. This is because 0°C does not represent a true zero. When it is 0°C outside, there is still some warmth, rather than none at all! In contrast, for **ratio data**, you can also calculate the relative difference or ratio between any two data values for a variable. Consequently, if a multinational company makes a profit of \$1,000,000,000 in one year and \$2,000,000,000 the following year, we can say that profits have doubled. Similarly, if you are estimating the number of people attending events such as political rallies using aerial photographs you might estimate the number of people at one event is half as many as at another.

Continuous and discrete data

Continuous data are those whose values can theoretically take any value (sometimes within a restricted range) provided that you can measure them accurately enough (Dancey and Reidy 2020). Data such as furnace temperature, delivery distance and length of service

are therefore continuous data. Similarly, data such as the amount of time a product is displayed in a television advertisement is continuous data. **Discrete data** can, by contrast, be measured precisely. Each case takes one of a finite number of values from a scale that measures changes in discrete units. These data are often whole numbers (**integers**) such as the number of mobile phones manufactured, number of occurrences of a particular word or phrase in employer associations' communications, or number of illustrations containing BAME (Black, Asian and Minority Ethnic) people in each issue of a fashion magazine over the last 10 years. However, in some instances (e.g. UK shoe size) discrete data will include non-integer values.

Definitions of discrete and continuous data are, in reality, dependent on how your data values are measured. The number of customers served by a large organisation is strictly a discrete datum as you are unlikely to get a part customer! However, for a large organisation with many customers you might treat this as a continuous datum, as the discrete measuring units are exceedingly small compared with the total number being measured.

12.3 Preparing data for quantitative analysis

When preparing data for quantitative analysis you need to be clear about the:

- definition and selection of cases;
- numeric codes used to classify data to ensure they will enable your research questions to be answered.

We now consider these.

Definition and selection of cases

The definition, selection and number of cases required for quantitative analysis (sample size) have already been discussed in Section 7.5 where we defined a case as an individual unit for which data have been collected. A case might be a respondent who had completed a questionnaire, an individual organisation or country for which secondary data had already been compiled, a magazine advertisement, a television commercial or an organisation's tweets. The data set would comprise the data collected from the respondents, organisations or countries, magazine advertisements, television commercials or an organisation's tweets you intend to analyse. Principles of probability sampling, outlined in Sections 7.4 to 7.7, apply when selecting such cases. However, for some research questions your data set might comprise one or only a few cases. A single case might be defined as the published report of a national inquiry, whereas if your data comprised main political parties' most recent general election manifestos, this would generate only a few cases, one for each political party. These cases would be most likely to be selected using non-probability sampling (Sections 7.8 and 7.9). It is therefore crucial to ensure that the cases selected will be sufficient to enable you to analyse the data quantitatively, answer your research question and meet your objectives.

Using numeric codes

Data for quantitative analysis should, with few exceptions, be recorded using numeric codes for each variable. This makes subsequent analyses, in particular those that require re-coding of data to create new variables, more straightforward. Data collection software

such as Qualtrics™ can be used to export responses selected to questions as numeric codes saving you the effort of entering your data yourself. Where you enter data yourself, the use of predetermined numeric codes enables you to do this quickly and with fewer errors using the numeric keypad on your keyboard. Unfortunately, with numeric codes meaningless analyses, such as calculating a mean (average) gender from codes 1 and 2, or the mean hotel location (Box 12.1), are easier! A common exception to using a numeric code for categorical data is where a postcode or zip code is used as the code for a geographical reference. If you are using a spreadsheet, you will need to keep a list of codes for each variable. Statistical analysis software can store these so that each code is automatically labelled.

Coding categorical data

For many secondary data sources (such as government surveys), a suitable coding scheme will have already been devised when the data were first collected. However, for other secondary sources such as documents (text, voice and visual) and all primary data you will need to decide on a coding scheme. Prior to this, you need to establish the highest level of precision required by your analyses (Figure 12.1).

Existing coding schemes can be used for many variables. These include industrial classification (Prosser 2009), occupation (Office for National Statistics 2021), social class and socioeconomic classification (Office for National Statistics nd a) and ethnic group (Office for National Statistics nd b), social attitude variables (National Centre for Social Research 2021) as well as coding schemes devised and used by other researchers. Wherever possible, we recommend you use these as they:

- save time;
- are normally well tested;
- allow comparisons of your findings with other research findings.

Where possible these codes should be specified in your online survey tool or data collection form as **pre-set codes**, provided there are a limited number of categories (Section 11.5). For such coding at data collection, the person filling in the form selects their response category and the associated code, this being added automatically when using survey design software (Section 11.5). Even if you decide not to use an existing coding scheme, perhaps because of a lack of detail, you should ensure that your codes are still compatible. This means that you will be able to compare your data with those already collected.

Coding of variables after data collection is necessary when you are unclear regarding the likely categories or there are a large number of possible categories in the coding scheme. To ensure that the coding scheme captures the variety in the data (and that it will work!) it is better to wait until data from the first 50 to 100 cases are available and then develop the coding scheme. This is called the **codebook** and can be used for both data from open questions' responses in questionnaires (Box 12.3) as well as visual and text data including tweets (Box 12.4). As when designing your data collection method(s) (Chapters 8–11), it is essential to be clear about the intended analyses, in particular the:

- level of precision required;
- coding schemes used by other research with which comparisons are to be made.



Box 12.3 Focus on student research

Developing a codebook for open questions with multiple responses

As part of his research project, Amil used a questionnaire to collect data from the customers of a local themed restaurant. The questionnaire included an open list question, which asked 'List up to three things you like about this restaurant'. Respondents

could therefore provide more than one answer to the question, in other words multiple responses. Their answers included over 50 different 'things' that the 186 customers responding liked about the restaurant, the maximum number mentioned by any one customer being constrained to three by the phrasing of the question.

Once data had been collected, Amil devised a hierarchical coding scheme based on what the customers liked about the restaurant.

Extract from coding scheme used to classify responses:

| Categories | Sub-categories | Response | Code |
|-----------------------|-------------------------------|----------|-------|
| Physical surroundings | | | 1–9 |
| | Decoration | | 1 |
| | Use of colour | | 2 |
| | Comfort of seating | | 3 |
| Dining experience | | | 10–49 |
| | Menu | | 10–19 |
| | Choice | | 11 |
| | Regularly changed | | 12 |
| | Food | | 20–29 |
| | Freshly prepared | | 21 |
| | Organic | | 22 |
| | Served at correct temperature | | 23 |
| | Staff attitude | | 30–39 |
| | Knowledgeable | | 31 |
| | Greet by name | | 32 |
| | Know what diners prefer | | 33 |
| | Discreet | | 34 |
| | Do not hassle | | 35 |
| | Good service | | 36 |
| | Friendly | | 37 |
| | Have a sense of humour | | 38 |
| | Drinks | | 40–49 |
| | Value for money | | 41 |
| | Good selection of wines | | 42 |
| | Good selection of beers | | 43 |
| | Served at correct temperature | | 44 |

The hierarchical coding scheme meant that individual responses could subsequently be re-coded into categories and sub-categories to facilitate a range of different analyses. These were undertaken

using statistical analysis software. Codes were allocated for each of up to three 'things' a customer liked, each of the three 'things' being represented by a separate variable.



Box 12.4 Focus on management research

Developing a codebook for tweet text data

Mirbabai and Marx (2020) conducted research into patterns of sense-breaking in social media crisis communications and their impact on collective sense-making and sense-giving. Published in *Behaviour and Information Technology*, their research comprises a case study of the 2017 Manchester Arena bombing after the Ariana Grande concert and explores how new novel (sense-breaking) information was introduced through tweets. It includes a social network analysis of 708,147 Twitter postings and a content analysis of 2006 individual tweets.

To undertake their content analysis, they manually selected the 2006 unique tweets that had been retweeted 50 or more times in the first six hours of crisis communication. These were coded by both authors separately using a self-developed codebook that, among other things, classified each tweet in terms of the information frame. The two authors' inter-rater reliability score of 0.892, confirming their coding could be considered as reliable. The 18 unique information frames used to code each tweet are, along with those for each tweet's author listed below:

| Information frames | Author roles |
|----------------------------------|-----------------------|
| Accommodation or lift offerings | Influencer (politics) |
| Injured people | Journalist |
| Islamist terrorism | Influencer (news) |
| Explosion caused by bomb | Fan page (music) |
| Emotional sensemaking | |
| Mass panic and chaos | |
| Missing persons | |
| Reported fatalities | |
| Unsupervised children at hotels | |
| Politics and media critique | |
| Arrested suspect | |
| Second bomb | |
| Gunman at Oldham hospital | |
| Suicide bomber | |
| Controlled explosion of 2nd bomb | |
| Appeal for blood donations | |

Mirbabai and Marx found individual type roles (such as private persons, journalists, and social media influencers) were initiators of sense-breaking in the early stages of the crisis when there was most uncertainty, these messages revolving around the incident reporting, crisis support and rumours. Their analysis, they argue, places increased emphasis on the role of individuals sense-giving efforts on collective sensemaking.

| Information frames | Author roles |
|--------------------------------------|--------------------|
| Explosion | Private person |
| Police and emergency rescue presence | Media organisation |

Creating a codebook for categorical data

To create your codebook for each variable you:

- 1 Examine the data and establish broad categories.
- 2 Subdivide the broad categories into increasingly specific subcategories dependent on your intended analyses.
- 3 Allocate codes to all categories at the most precise level of detail required.
- 4 Note the actual responses that are allocated to each category and produce a codebook.

- 5** Ensure that those categories that may need to be aggregated are given adjacent codes to facilitate re-coding.

Subsequently codes are attached to specific segments (or units) of data (Rose et al. 2015). Segments may be individual words, based on identifying and counting particular words in the content of your sample, as in our example about attitudes towards an organisational policy. Alternatively, the segment may be larger than a word, being related to the occurrence of particular phrases or to sentences or paragraphs. In coding occurrences in the data, you are coding the **manifest content**, which is components that are clearly visible in the data and can be counted. Larger segments (sentences or paragraphs) are often used where it is important to contextualise content to be able to categorise the meanings behind the manifest content. The meaning behind the manifest content is termed the **latent content** and is often difficult to infer even from larger segments. However, in general manifest content is likely to be reflected in the use of the word or phrase and latent content is likely to be reflected in the use of larger segments. A segment may also focus on the characteristics of those involved, as in our example where gender, age, occupation and work department were recorded, or other characteristics that are relevant to record and analyse for your research. Segment in visual data varies from individual images to visual sequences.

Coding involves you working through your data to code segments of these data according to the categories you have devised. We suggest you start with a sample of your data and test your codes and, if necessary, modify your codebook before applying it across all of your data. An important way for you to assess whether your system of categories is transparent and capable of being applied consistently by others is for you and a friend to code a sample of the same data separately using this system of categories and then to compare your coding. This is known as **inter-rater reliability** and can be assessed by measuring the extent to which two or more coders (raters) agree. One way of doing this is to calculate the percentage agreement using the following formula:

$$PA = \frac{A}{n} \times 100$$

where:

PA = percentage agreement

A = number of agreements between the two coders

n = number of segments coded

Although there is no clear agreement regarding an acceptable percentage agreement, McHugh (2012) suggests that scores of 80 per cent or higher would normally be considered acceptable. A more sophisticated measure of inter-rater reliability for two raters is Cohen's Kappa (McHugh, 2012), which can usually be calculated using your analysis software.

Coding numerical data

The actual numbers recorded, such as a respondent's age in years or the number of tickets sold for a football match, are often used as codes for numerical data, even though this level of precision may not be required. Once these data have been entered in a data matrix (Section 12.4), you can use analysis software to group or combine data to form additional variables with less detailed categories. This process is referred to as **re-coding**. For example, a Republic of Ireland employee's salary could be coded to the nearest euro and entered into the matrix as 73543 (numerical discrete data). Later, re-coding could be used to place it in a group of similar salaries, from €70,000 to €79,999 (categorical ranked data).

Coding missing data

Where you have been able to obtain at least some data for a case, rather than none, you should ensure that each variable for each case in your data set has a code. Where data have not been collected for some variables, you therefore need a code to signify these data are missing. The choice of code to represent missing data is up to you, although some statistical analysis software has a code that is used by default. A missing data code can also be used to indicate why data are missing. Missing data are important as they may affect whether the data you have collected are representative of the population. If missing data follow some form of pattern, such as occurring for particular questions or for a subgroup of the population, then your results are unlikely to be representative of the population and so you should not ignore the fact they are missing. However, if data are missing at random, then it is unlikely that this will affect your results being representative of the population (Little and Rubin 2019). Reasons for missing data in relation to questionnaires include:

- the data were not required from the respondent, perhaps because of a skip generated by a filter question in a questionnaire;
- the respondent refused to answer the question (a **non-response**);
- the respondent did not understand the question;
- the respondent did not know the answer or did not have an opinion. Sometimes this is treated as implying an answer; on other occasions it is treated as missing data;
- the respondent may have missed a question by mistake, or the respondent's answer may be unclear;
- leaving part of a question in a survey blank implies an answer; in such cases the data are not classified as missing (Section 11.5).

Content analysis

Content analysis is a specific analytical technique of categorising and coding text, images and expressions as data that have been created to be seen, read interpreted and acted upon by people, data being analysed with such uses in mind (Krippendorff 2018). It therefore tends to focus upon textual matter, symbols, images, messages, mass media content and interactions; using a systematic replicable coding scheme to enable quantitative analysis. Although there are numerous definitions of content analysis, most draw on an early definition by Berelson (1952:18) as a 'technique for the objective, systematic and quantitative description of the manifest content of communication'.

The 'objective' nature of content analysis emphasises that different researchers should be able to replicate their analysis by using the explicit categories to code components and produce an identical outcome. What you choose to code and subsequently count is dependent upon your research question. You may, for example, ask what are the attitudes towards an organisational policy and who holds these views. Content analysis of interview recordings (voice), interview transcripts (text), or organisation communications could be used to code variables such as attitude towards the policy, these attitude data being categorised as positive, neutral or negative. You would identify terms denoting negative, neutral or positive attitudes, these typically being pre-determined before your analysis commences. You would categorise and code specific instances of these in the text and identify the characteristics of the holders of each of these attitudes defining these categories using variables such as gender, age, occupation, work department and so forth.

'Systematic' emphasises that content analysis should be conducted in a consistent, transparent and replicable way with clear rules for defining and applying codes being detailed in a code book or coding manual. This coding scheme can draw on existing

schemes developed by other researchers or developed inductively from the data (Box 12.4) using similar techniques to those outlined in Section 13.6. Holsti (1969) advocates five general principles for the systematic development of variables' categories in content analysis. These should:

- link obviously to the scope and purpose of the research topic, not least so that the relationship of these categories to the research question and objectives is evident (Section 2.6);
- be exhaustive so that every relevant component of data may be placed into an analytical category;
- be mutually exclusive so that each component of data may only be placed into one analytical category, rather than possibly fitting into more than one;
- be independent so that components of data exhibiting related but not the same characteristics cannot be coded into the same category; and
- be developed from a single classification to avoid conceptual confusion.

Subsequent quantitative analysis range from calculating the frequency of different categories for a variable (Section 12.6) to examining relationships between variables created (Section 12.7). Using our earlier example about attitudes towards an organisational policy and who holds these views, you could calculate the frequency for each category of the variable attitude towards the policy and establish the relative importance of negative, neutral or positive attitudes. It would also be possible for you to present these data graphically (Section 12.5) to, for example, show the relative amounts for each of the categories; and test statistically whether differences in attitudes were associated significantly with variables such as gender (Section 12.7). However, while the significance or otherwise of such relationships could be established, explaining in detail why they were significant just using content analysis would be difficult.

12.4 Data entry and checking

When entering your data into analysis software you need to ensure the:

- data layout and format meet that required by the analysis software;
- data, once entered, have been saved and a back-up copy made;
- data have been checked for errors and any found corrected;
- need to weight cases has been considered.

Data layout

Some primary data collection methods, such as online questionnaires, computer-aided personal interviewing (CAPI) and computer-aided telephone interviewing (CATI) automatically enter and save data electronically. These data can subsequently be exported in a range of formats compatible with different analysis software. Cloud-based survey design, data collection and analysis software such as Qualtrics Research CORE™ and SurveyMonkey™ go one stage further and integrate the analysis in the same software as questionnaire design and data capture (Qualtrics 2021; SurveyMonkey 2021). Alternatively, digital secondary data (Section 8.3) can be downloaded in a format compatible with your analysis software. However, where you have to prepare and enter data yourself for computer analysis, you will need to be clear about the precise data layout requirements of your analysis software.

Table 12.1 A simple data matrix

| | Id | Variable 1 | Variable 2 | Variable 3 | Variable 4 |
|---------------|-----------|-------------------|-------------------|-------------------|-------------------|
| Case 1 | 1 | 27 | 1 | 2 | 1 |
| Case 2 | 2 | 19 | 2 | 1 | 2 |
| Case 3 | 3 | 24 | 2 | 3 | 1 |

Virtually all analysis software will accept your data if they are entered in tabular format as a **data matrix** (Table 12.1). Once data have been entered into your analysis software, it is usually possible to save them in a format that is compatible with other software. Within a data matrix, each column usually represents a separate variable for which you have obtained data. Each matrix row contains the variables for an individual case, that is, an individual unit for which data have been obtained. If your data have been collected using a questionnaire, each row will contain the coded data from one questionnaire; if your data are pictures tweeted by people attending a heavy metal music concert then each row will contain the coded data relating to a picture tweeted. Secondary data that have already been stored in a data file are almost always held as a data matrix. For such data sets you usually select the subset of variables and cases you require and save these as a separate matrix. If you enter your own data, these are input directly into your chosen analysis software one case (row) at a time using codes to record the data (Box 12.5). Larger data sets with more data variables and cases result in larger data matrices. Although data matrices store data using one column for each variable, this may not be the same as one column for each question for data collected using surveys (Box 12.6).



Box 12.5 Focus on student research

A spreadsheet data matrix

Lucy was interested in what people videoed with their smartphones when they attended a trade show. Thirty trade show visitors who had used their smartphones consented to allow her to use the video clips they had taken. In all, she had 217 videos to analyse. Lucy decided to treat each video clip as a separate case. In her Excel spreadsheet, the first variable (*id*) was the video clip identifier. This meant that she could link data for each case (row) in her matrix to the video clip when checking for errors. The second variable (*age*) contained

numerical (ratio) data, the age of each person who had taken the video clip (at the time the video had been taken). Subsequent variables contained further data: the third (*gender*) recorded this dichotomous (categorical) data using code 1 for a male and 2 for a female person taking the video clip. The fourth variable (*length*) recorded the length of the video clip in seconds (numerical, ratio data). The fifth variable (*focus*) recorded the overall focus of the video clip. In developing her codebook for this nominal (categorical) variable Lucy had noted that the video clips focussed on three categories: products (code 1) services (code 2) and people (code 3). The codes used by Lucy, therefore, had different meanings for different variables. Subsequent variables related to different aspects of the content of the video clips, the codes being recorded in Lucy's codebook.

| | A | B | C | D | E | F | G |
|---|----|-----|--------|-------|--------|--------|-------------|
| 1 | id | age | gender | focus | length | people | interaction |
| 2 | 1 | 27 | 1 | 2 | 5.9 | 3 | |
| 3 | 2 | 35 | 2 | 2 | 7.8 | 1 | |
| 4 | 3 | 41 | 2 | 3 | 23.2 | 1 | |



Box 12.6 Focus on student research

Data coding for more advanced statistical analysis software

As part of a market research project, Zack needed to discover which of four products (tomato ketchup, brown sauce, soy sauce and mayonnaise) had been purchased within the last month by consumers. He therefore needed to collect four data items from each respondent:

- 1 Which of the following items have you purchased within the last month?

| Item | Purchased | Not purchased | Not sure |
|----------------|----------------------------|----------------------------|----------------------------|
| Tomato ketchup | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 3 |
| Brown sauce | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 3 |
| Soy sauce | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 3 |
| Mayonnaise | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 3 |

The data Zack collected from each respondent formed four separate nominal (categorical) variables in the data matrix using numerical codes (1 = purchased, 2 = not purchased, 3 = not sure). This is known as multiple-dichotomy coding.

| | tomato | brown | soy | mayonaise | like1 | like2 | like3 | like4 | like5 |
|---|--------|-------|-----|-----------|-------|-------|-------|-------|-------|
| 1 | 1 | 1 | 1 | 2 | 23 | 31 | 17.00 | 4.00 | 5.00 |
| 2 | 2 | 2 | 2 | 3 | 12 | 15 | 12.00 | 5.00 | . |
| 3 | 1 | 2 | 3 | 1 | 23 | 12 | 4.00 | . | . |
| 4 | | | | | | | | | |

Zack also included a question (question 2 below) that could theoretically have millions of possible responses for each of the 'things'. For such questions, the number of 'things' that each respondent mentions may also vary. Our experience suggests that virtually all respondents will select five or fewer. Zack therefore left space to code up to five responses after data had been collected in the nominal (categorical) variables 'like1', 'like2', 'like3', 'like4' and 'like5'. This is known as multiple-response coding. When there were fewer than five responses given, the code '.' was entered automatically by the software into empty cells for the remaining 'like' variables, signifying missing data.

- 2 List up to five things you like about tomato ketchup

- Tomato ketchup purchased within the last month? Yes/No
- Brown sauce purchased within the last month? Yes/No
- Soy sauce purchased within the last month? Yes/No
- Mayonnaise purchased within the last month? Yes/No

Each of these data items is a separate variable. However, the data were collected using one matrix question in an interviewer completed telephone questionnaire:

For office use only

| | | | | |
|-------|--------------------------|--------------------------|--------------------------|--------------------------|
| | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |

We strongly recommend that you save your data regularly as you are entering it, to minimise the chances of deleting by accident! In addition, you should save multiple backup or security copies on your smartphone or other mass storage device, and the cloud, ensuring your data are stored securely.

If you intend to enter data into a spreadsheet, the first variable is in Column A, the second in Column B and so on. Each cell in the first row (1) should contain a short variable name to enable you to identify each variable. Subsequent rows (2 onwards) will each contain the data for one case (Box 12.5). Statistical analysis software follows the same logic, although the variable names are usually displayed ‘above’ the first row (Box 12.6).

The **multiple-response method** of coding uses the same number of variables as the maximum number of different responses from any one case. For Question 2 these were named ‘like1’ through to ‘like5’ (Box 12.6). Each of these variables would use the same codes and could include any of the responses as a category. Statistical analysis software often contains multiple-response procedures to analyse such data. The alternative, the **multiple-dichotomy method** of coding, uses a separate variable for each different answer (Box 12.5). For Question 2 (Box 12.6) a separate variable could have been used for each ‘thing’ listed: for example, flavour, consistency, bottle shape, smell, price and so on. You subsequently would code each variable as ‘listed’ or ‘not listed’ for each case. However, although the multiple dichotomy method makes it easy to calculate the number of responses for each ‘thing’ (Field, 2018), it means where there are a large number of different responses, a large number of variables will be required. As entering data for a large number of variables is more time-consuming, it is important to ensure your statistical analysis software can calculate the number of responses for each ‘thing’ if you intend to use the multiple response method.

Entering and saving data

If you have downloaded secondary data as a file, or have used online survey tool, your data will already have been entered (input) and saved. However, you may need to enter and save the data as a file yourself. In either situation, it is essential that you ensure that your data have been recorded correctly and save the file regularly. When saving our data files, we have found it helpful to include the word DATA in the filename. When entering data, the well-known maxim ‘rubbish in, rubbish out’ certainly applies! More sophisticated analysis software allows you to attach individual labels to each variable and the codes associated with each of them. If this is feasible, we strongly recommend that you do this. By ensuring the labels replicate the exact words used in the data collection, you will reduce the number of opportunities for misinterpretation when analysing your data. Taking this advice for the variable ‘like1’ in Box 12.6 would result in the variable label ‘List up to three things you like about this restaurant’, each value being labelled with the actual response in the coding scheme.

Checking for errors

No matter how carefully you code and subsequently enter data there will always be some errors. The main methods to check data for errors are as follows:

- Look for illegitimate codes. In any coding scheme, only certain numbers are allocated. Other numbers are, therefore, errors. Common errors when entering the data yourself are the inclusion of letters O and o instead of zero, letters l or I instead of 1, and number 7 instead of 1.

- Look for illogical relationships. For example, if a person is coded to the ‘higher managerial occupations’ socioeconomic classification category and she describes her work as ‘manual’, it is likely an error has occurred.
- For questionnaire data, check that rules in filter questions are followed. Certain responses to filter questions (Section 11.6) mean that other variables should be coded as missing values. If this has not happened, there has been an error.

For each possible error, you need, if possible, to discover whether it occurred at coding or data entry and then correct it. By giving each case a unique numeric identifier (Box 12.5), it is possible to link the matrix to the original data. You must, however, remember to ensure the identifier is on the data collection form and entered along with the other data into the matrix.

Data checking is very time-consuming and so is often not undertaken. Beware: not doing it is very dangerous and can result in incorrect results from which false conclusions are drawn!

Weighting cases

Most data you use will be collected from a sample. For some forms of probability sampling, such as stratified random sampling (Section 7.6), you may have used a different sampling fraction for each stratum. Alternatively, you may have obtained a different response rate for each of the strata. To obtain an accurate overall picture you will need to take account of these differences in response rates between strata. A common method of achieving this is to use cases from those strata that have lower proportions of responses so each represents more than one case in your analysis (Box 12.7). Most statistical analysis software allows you to do this by **weighting** cases.

To weight the cases, you:

- 1 Calculate the percentage of the population responding for each stratum.
- 2 Establish which stratum had the highest percentage of the population responding.
- 3 Calculate the weight for each stratum using the following formula:



Box 12.7 Focus on student research

Weighting cases

Doris had used stratified random sampling to select her sample. The percentage of each stratum’s population that responded is given below:

- upper stratum: 90 per cent;
- lower stratum: 65 per cent.

To account for the differences in the response rates between strata she decided to weight the cases prior to analysis.

Each case in the upper stratum counted as 1 case in her analysis.

The weight for the upper stratum was : $\frac{90}{90} = 1$

Each case in the lower stratum counted for 1.38 cases in her analysis.

The weight for the lower stratum was : $\frac{90}{65} = 1.38$

Doris entered these weights as a separate variable in her data matrix and used the statistical analysis software to apply them to the data.

$$\text{Weight} = \frac{\text{highest proportion of population responding for any stratum}}{\text{proportion of population responding in stratum for which calculating weight}}$$

(Note: if your calculations are correct this will always result in the weight for the stratum with the highest proportion of the population responding being 1.)

4 Apply the appropriate weight to each case.

Beware: many authors (for example, Hays 1994) question the validity of using statistics to make inferences from your sample if you have weighted cases.

12.5 Exploring and presenting data: an overview

Once your data have been entered, checked and errors corrected, you are ready to start your analysis. We have found Tukey's (2020) **exploratory data analysis (EDA)** approach useful in these initial stages. This approach emphasises using graphs to explore and understand your data. Although within data analysis the term graph has a specific meaning: '... a visual display that illustrates one or more relationships among numbers' (Kosslyn 2006: 4), it is often used interchangeably with the term 'chart' both by authors and data analysis software. Even more confusingly, what are referred to as 'pie charts' are actually graphs! Tukey (2020) also emphasises the importance of using your data to guide your choice of analysis techniques. As you would expect, we believe that it is important to keep your research question(s) and objectives in mind when exploring your data. However, exploratory data analysis allows you flexibility to introduce previously unplanned analyses to respond to new findings. It therefore formalises the common practice of looking for other relationships in data which your research was not initially designed to test. This should not be discounted, as it may suggest other fruitful avenues for analysis.

Even at this stage it is important that you structure and label clearly each graph and table to avoid possible misinterpretation. Box 12.8 provides a checklist of points to remember when designing a graph or table.



Box 12.8 Checklist

Designing your graphs and tables

For both graphs and tables

- ✓ Does it have a brief but clear and descriptive title?
- ✓ Are the units of measurement used stated clearly?
- ✓ Are the sources of data used stated clearly?
- ✓ Are there notes to explain abbreviations and unusual terminology?
- ✓ Does it state the size of the sample on which the values in the graph/table are based (where needed)?

For graphs

- ✓ Does it have clear axis labels?
- ✓ Are bars and their components in the same logical sequence?
- ✓ Is more dense shading used for smaller areas?
- ✓ Have you avoided misrepresenting or distorting the data?
- ✓ Is a key or legend included (where necessary)?

For tables

- ✓ Does it have clear column and row headings?
- ✓ Are columns and rows in a logical sequence?
- ✓ Are numbers in columns right justified?

We have found it best to begin exploring data by looking at individual variables and their components. The key aspects you may need to consider will be guided by your research question(s) and objectives, and are likely to include (Kosslyn 2006) for single variables:

- specific amounts represented by individual data values;
- relative amounts such as:
 - highest and lowest data values;
 - trends in data values;
 - proportions and percentages for data values;
 - distributions of data values.

Once you have explored these, you can then begin to compare variables and interdependences between variables, by (Kosslyn 2006):

- comparing intersections between the data values for two or more variables;
- comparing cumulative totals for data values and variables;
- looking for relationships between cases for variables.

These are summarised in Table 12.2. Most analysis software can create tables and graphs. Your choice will depend on those aspects of the data to which you wish to direct your readers' attention and the precision at which the data were measured. This section is concerned only with tables and two-dimensional graphs, including pictograms, available with most spreadsheets (Table 12.2). Three-dimensional graphs are not discussed, as

Table 12.2 Data presentation by data type: A summary

| | Categorical | | Numerical | |
|--|---|-------------------------------|---|---|
| | Nominal (Descriptive) | Ordinal (Ranked) | Continuous | Discrete |
| To show one variable so that any <i>specific amount</i> can be read easily | Table/frequency distribution (data often grouped) | | | |
| To show the relative amount for categories or values for one variable so that <i>highest</i> and <i>lowest</i> are clear | Bar graph/chart, pictogram or data cloud (data may need grouping) | | Histogram or frequency polygon (data must be grouped) | Bar graph/chart or pictogram (data may need grouping) |
| To show the <i>trend</i> for a variable | | Line graph or bar graph/chart | Line graph or histogram | Line graph or bar graph/chart |
| To show the <i>proportion</i> or <i>percentage</i> of occurrences of categories or values for one variable | Pie chart or bar graph/chart (data may need grouping) | | Histogram or pie chart (data must be grouped) | Pie chart or bar graph/chart (data may need grouping) |
| To show the <i>distribution</i> of values for one variable | | | Frequency polygon, histogram (data must be grouped) or box plot | Frequency polygon, bar graph/chart (data may need grouping) or box plot |

(continued)

Chapter 12 Analysing data quantitatively

Table 12.2 Data presentation by data type: A summary (*Continued*)

| | Categorical | | Numerical | | | |
|--|--|---------------------|----------------------------|----------|--|--|
| | Nominal (Descriptive) | Ordinal (Ranked) | Continuous | Discrete | | |
| To show the <i>interrelationship</i> between two or more variables so that any <i>specific</i> amount can be read easily | Contingency table/cross-tabulation (data often grouped) | | | | | |
| To compare the relative amount for categories or values for two or more variables so that <i>highest</i> and <i>lowest</i> are clear | Multiple bar graph/chart (continuous data must be grouped; other data may need grouping) | | | | | |
| To compare the <i>proportions</i> or <i>percentages</i> of occurrences of categories or values for two or more variables | Comparative pie charts or percentage component bar graph/chart (continuous data must be grouped; other data may need grouping) | | | | | |
| To compare the <i>distribution</i> of values for two or more variables | | | Multiple box plot | | | |
| To compare the <i>trends</i> for two or more variables so that <i>intersections</i> are clear | Multiple line graph or multiple bar graph/chart | | | | | |
| To compare the frequency of occurrences of categories or values for two or more variables so that <i>cumulative totals</i> are clear | Stacked bar graph/chart (continuous data must be grouped; other data may need grouping) | | | | | |
| To compare the <i>proportions</i> and <i>cumulative totals</i> of occurrences of categories or values for two or more variables | Comparative proportional pie charts (continuous data must be grouped; other data may need grouping) | | | | | |
| To show the <i>interrelationship</i> between cases for two variables | | | Scatter graph/scatter plot | | | |

Source: © Mark Saunders, Philip Lewis and Adrian Thornhill 2022

these can often mislead or hinder interpretation (Kosslyn 2006). Those tables and graphs most pertinent to your research question(s) and objectives will eventually appear in your research report to support your arguments. You should therefore save a copy of all tables and graphs you create.

12.6 Exploring and presenting individual variables

To show specific amounts

Tables

The simplest way of summarising data for individual variables so that specific amounts can be read is to use a **table (frequency distribution)**. For categorical data, the table summarises the number of cases (frequency) in each category. For variables where there are likely to be a large number of categories (or values for numerical data), you will need to group the data into categories that reflect your research question(s) and objectives.

To show highest and lowest values

Bar graphs and bar charts

Tables attach no visual significance to highest or lowest data values unless emphasised by alternative fonts. Graphs can provide visual clues, although both categorical and numerical data may need grouping. For categorical and discrete data, bar graphs and pictograms are both suitable. Generally, bar graphs provide a more accurate representation and should be used for research reports. In a **bar graph**, also often known as a **bar chart**, the height or length of each bar represents the frequency of occurrence. Bars are separated by gaps, usually half the width of the bars. Bar graphs where the bars are vertical (as in Figure 12.2) are sometimes called bar or column charts. This bar graph emphasises that the European Union member state with the highest proportion of energy from renewable sources in 2019 was Iceland, while Malta, and then either Luxembourg or the Netherlands had the lowest proportions of energy from renewable sources. By presenting the bars in alphabetical order of country, it makes it easier to locate a specific country rather than compare relative amounts (Figure 12.4).

Pictograms

In a **pictogram** (also known as a pictograph), each bar is replaced by a picture or series of pictures chosen to represent the data and convey a general impression or gain an audience's attention. For this reason, they are often used in infographics. To illustrate the impact of doing this, we have used data of worldwide Harley-Davidson motorcycle shipments to generate both as a pictogram (Figure 12.3) and, later, a histogram (Figure 12.5). In the pictogram each picture represents 20,000 motorcycles. Pictures in pictograms can, like bars in bar graphs and histograms, be shown in columns or horizontally. The height of the column or length of the bar made up by the pictures represents the frequency of occurrence. In this case we felt it was more logical to group the pictures as a horizontal bar rather than vertically on top of each other. You will have probably also noticed that, in the pictogram, there are gaps between the 'bars'. While this normally signifies discrete categories of data, it is also acceptable to do this for continuous data (such as years) when drawing a pictogram, to aid clarity. Although analysis software allows you to convert a bar graph or histogram to a pictogram easily and accurately, it is more difficult to establish the actual data values from a pictogram. This is because the number of units part of a picture

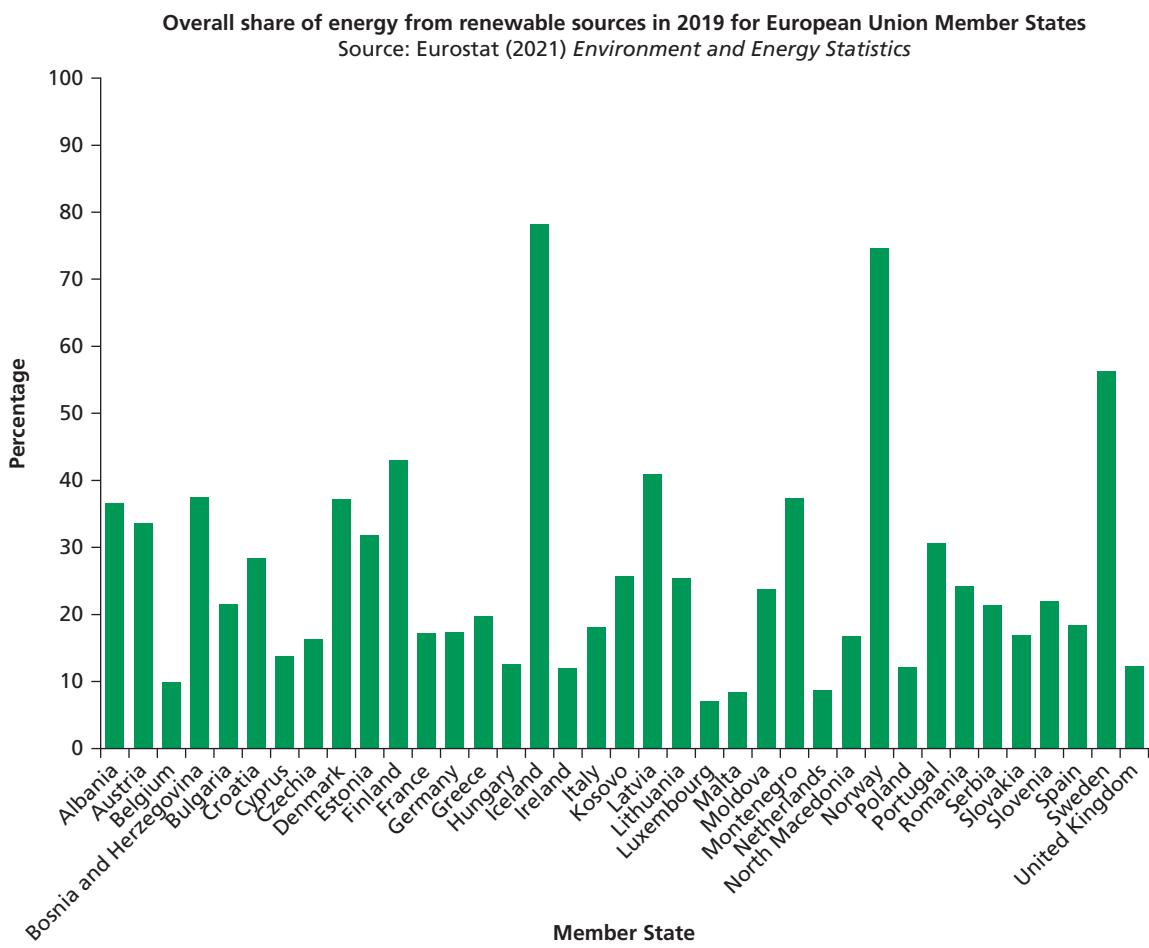


Figure 12.2 Bar graph

Source: Adapted from Eurostat (2021) © European Communities 2021

represents is not immediately clear. For example, in Figure 12.3, how many motorcycles shipped would a rear wheel represent?

Pictograms have a further drawback, namely that it is very easy to misrepresent the data. Both Figures 12.3 and 12.5 show shipments of Harley-Davidson motorcycles declined between 2006 and 2010 and again from 2014. Using our analysis software, the first of these declines could have been represented using a picture of a motorcycle in 2006 that was nearly one and a half times as long as the picture in 2010. However, to keep the proportions of the motorcycle accurate, the picture would have needed to be nearly one and a half times as tall. Consequently, the actual area of the picture for 2006 would have been over twice as great and would have been interpreted as motorcycle shipments being twice as large in 2006 than 2010! Because of this we would recommend that if you are using a pictogram, you decide on a standard value for each picture and do not alter its size. In Figure 12.3 we choose one image to represent 20,000 motorcycles and included a key to indicate the value each image represented.

Worldwide Harley-Davidson motorcycle shipments 1996-2020

Source: Harley-Davidson Inc. (2021)

**Figure 12.3** Pictogram

Source: Adapted from Harley-Davidson Inc. (2021)

*To show relative amounts**Bar graphs*

To emphasise the relative values represented by each of the bars in a bar graph, the bars may be reordered in either descending or ascending order of the frequency of occurrence represented by each bar (Figure 12.4). It is now clear from the order of the bars that Iceland had the highest percentage of energy from renewable sources in 2019, with Luxembourg and Malta having the lowest proportions.

Word clouds

For text data the relative proportions of key words and phrases can be shown using a **word cloud** (Box 12.9), there being numerous free word cloud generators such as Wordclouds.com™ available online. In a word cloud the frequency of occurrence of a particular word or phrase is represented by the font size of the word or occasionally the colour.

Histograms

Most researchers use a histogram to show highest and lowest values for continuous data. Prior to being drawn, data will often need to be grouped into class intervals. In a **histogram**, the area of each bar represents the frequency of occurrence, and the continuous nature of the data emphasised by the absence of gaps between the bars. For equal

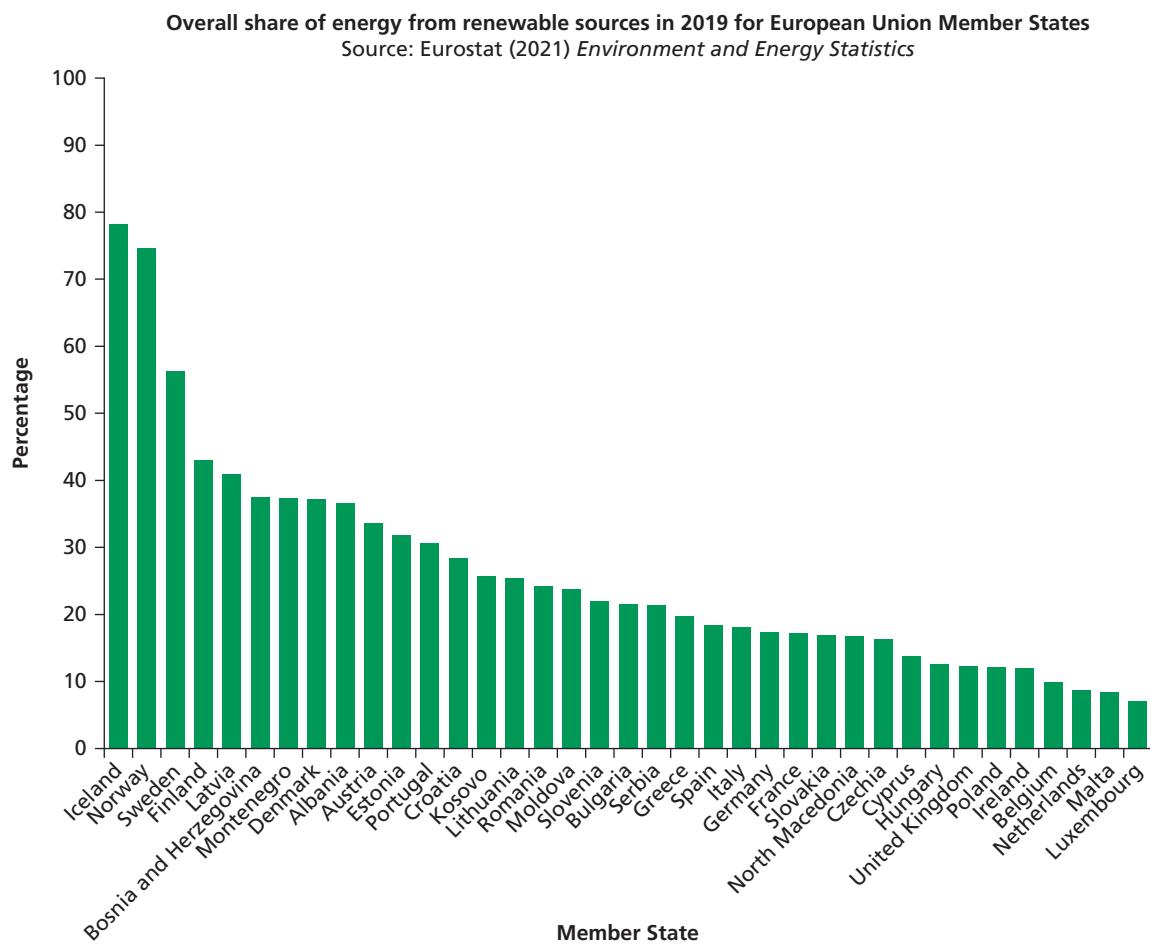


Figure 12.4 Bar graph (data reordered)

Source: Adapted from Eurostat (2021) © European Communities 2021.



Box 12.9 Focus on student research

Using a word cloud to display the frequency of key terms

Luca undertook a research project evaluating types of pay structure. This involved him conducting interviews in organisations that each used a different pay structure. Luca wanted to understand the reasons why each

had decided to adopt a particular structure and to evaluate perceptions about that structure's use in practice. To demonstrate the frequency of key terms used by his interview participants he thought it might be useful to produce a word cloud for each set of interviews exploring a particular pay structure. Since these word clouds would represent the actual terms used by his interview participants, they also helped Luca to demonstrate how he had derived his codes from his data. This word cloud represents the terms used by interview participants in an organisation that had implemented a Job Families pay structure.



Source: © Mark Saunders 2018

width class intervals, the height of your bar still represents the frequency of occurrences (Figure 12.5) and so the highest and lowest values are easy to distinguish. For histograms with unequal class interval widths, this is not the case (Box 12.12), and each bar's height represents the relative frequency of occurrences. In Figure 12.5 the histogram emphasises that the highest number of Harley-Davidson motorcycles shipped worldwide was in 2006, and the lowest number in 1986.

Analysis software treats histograms for data of equal width class intervals as a variation of a bar chart. Unfortunately, analysis software can rarely cope automatically with the

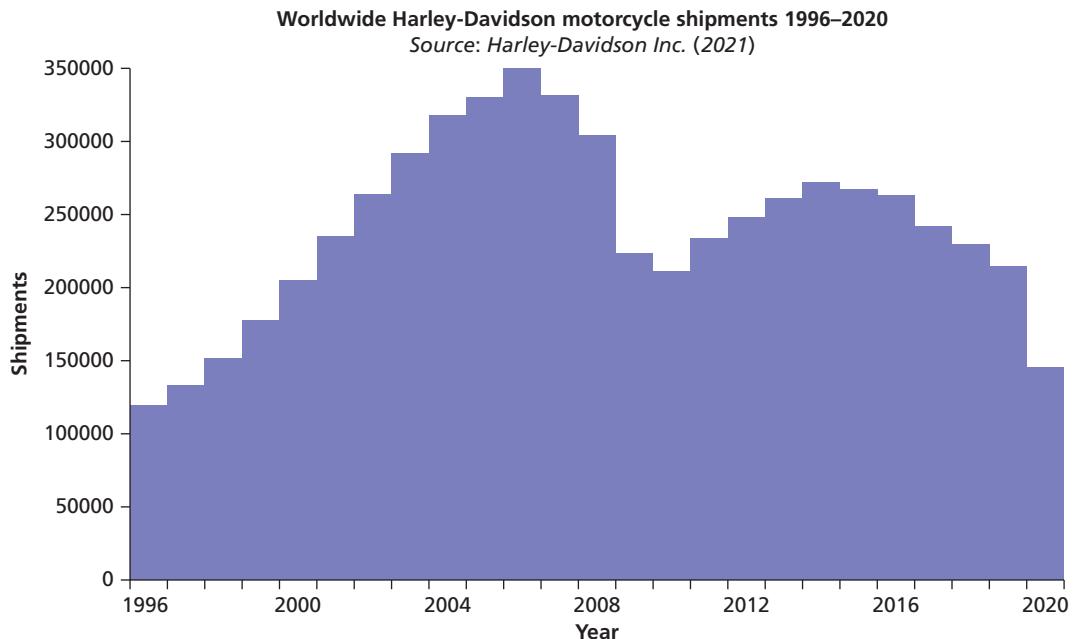


Figure 12.5 Histogram

Source: Adapted from Harley-Davidson Inc. (2021).

calculations required to draw histograms for unequal class intervals. Consequently, you may have to use a bar chart owing to the limitations of your analysis software.

Frequency polygons

Frequency polygons, although more frequently used to show distributions, are used occasionally to illustrate limits. Most analysis software treats them as a version of a line graph (Figure 12.6) in which the lines are extended to meet the horizontal axis, provided class widths are equal.

To show a trend

Line graphs

Trends can only be presented for variables containing numerical (and occasionally ranked) longitudinal data. The most suitable diagram for exploring the trend is a **line graph** (Kosslyn 2006) in which the data values for each time period are joined with a line to represent the trend. In Figure 12.6 the line graph reveals the rise and decline in the number of Harley-Davidson motorcycles shipped worldwide between 1996 and 2020. You can also use histograms (Figure 12.5) to show trends over continuous time periods and bar graphs to show trends between discrete time periods. The trend can also be calculated using time-series analysis (Section 12.12).

To show proportions or percentages

Pie charts

The most frequently used diagram to emphasise the proportion or share of occurrences is the pie chart, although bar charts have been shown to give equally good results

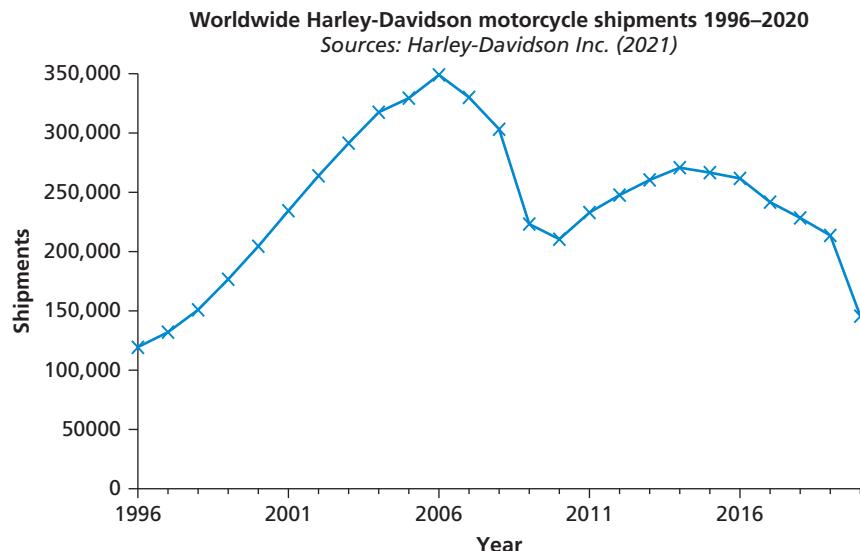


Figure 12.6 Line graph

Source: Adapted from Harley-Davidson Inc. (2021).

(Anderson et al. 2020). A **pie chart** is divided into proportional segments according to the share each has of the total value and the total value represented by the pie is noted (Box 12.10). For numerical and some categorical data, you will need to group data prior to drawing the pie chart, as it is difficult to interpret pie charts with more than six segments (Keen 2018).



Box 12.10 Focus on student research

Exploring and presenting data for individual variables

As part of audience research for his dissertation, Valentin asked people attending a play at a provincial theatre to complete a short paper questionnaire. This collected responses to 25 questions including:

- 3** How many plays (including this one) have you seen at this theatre in the past year?

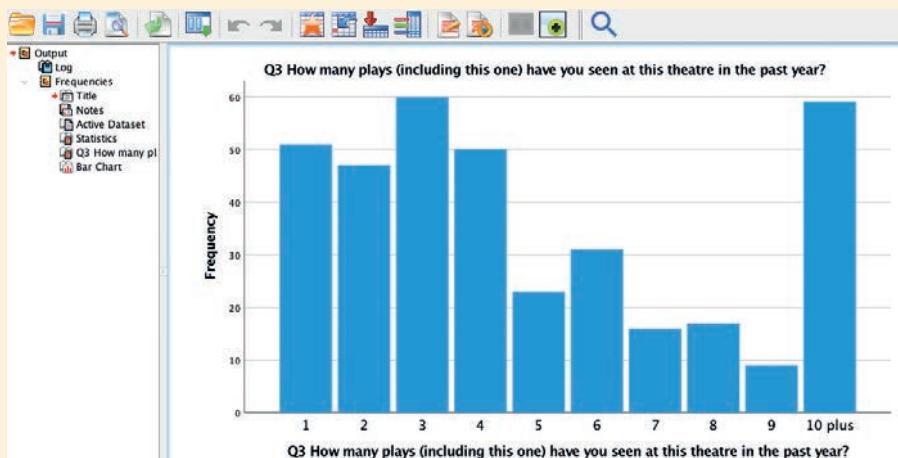
- 11** This play is good value for money.

| | |
|----------------------------|----------------------------|
| strongly disagree | disagree |
| <input type="checkbox"/> 1 | <input type="checkbox"/> 2 |
| agree | strongly agree |
| <input type="checkbox"/> 3 | <input type="checkbox"/> 4 |

- 24** How old are you?

| | |
|----------------------------|----------------------------|
| Under 18 | 18 to 34 |
| <input type="checkbox"/> 1 | <input type="checkbox"/> 2 |
| 35 to 64 | 65 and over |
| <input type="checkbox"/> 3 | <input type="checkbox"/> 4 |

Exploratory analyses were undertaken using analysis software and diagrams and tables generated. For



Question 3, which collected discrete (numerical) data, the aspects that were most important were the distribution of values and the highest and lowest numbers of plays seen. A bar graph, therefore, was drawn:

This emphasised that the most frequent number of plays seen by respondents was three and the least

frequent number of plays seen by the respondents was either nine or probably some larger number. It also suggested that the distribution was positively skewed towards lower numbers of plays seen.

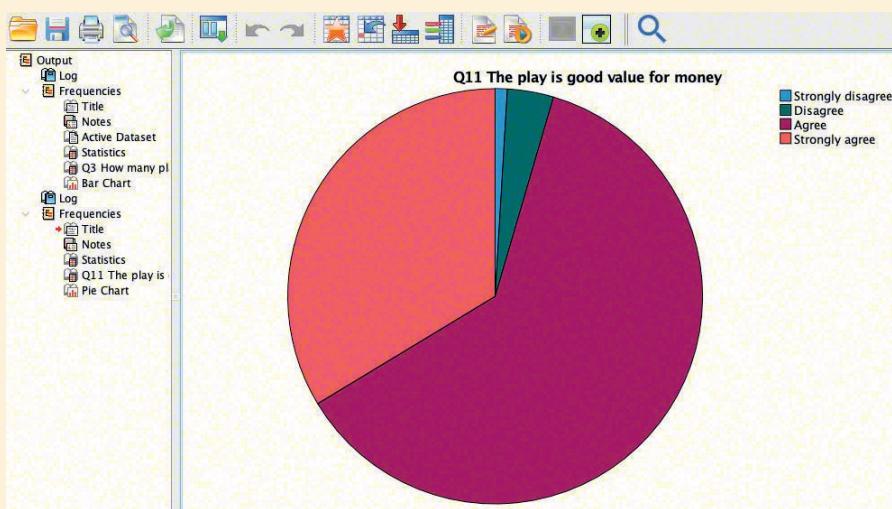
For Question 11 (ordinal categorical data), the most important aspect was the proportion of people





Box 12.10 Focus on student research (*continued*)

Exploring and presenting data for individual variables



agreeing and disagreeing with the statement. A pie chart was therefore drawn, although unfortunately the shadings were not similar for the two 'agree' categories and for the two 'disagree' categories.

This emphasised that the vast majority of respondents (95 per cent) agreed that the play was good value for money.

Question 24 collected data on each respondent's age. This question had grouped continuous (numerical) data into four unequal-width age groups meaning it was recorded as ordinal (categorical) data. For this analysis, the most important aspects were the specific number and percentage of respondents in each age category and so a table was constructed.

| Q24 How old are you? | | | | | |
|----------------------|-----------|---------|---------------|--------------------|-------|
| | Frequency | Percent | Valid Percent | Cumulative Percent | |
| Valid | Under 18 | 30 | 4.4 | 4.4 | 4.4 |
| | 18 to 34 | 144 | 20.9 | 21.0 | 25.4 |
| | 35 to 64 | 366 | 53.2 | 53.4 | 78.8 |
| | 65 plus | 145 | 21.1 | 21.2 | 100.0 |
| Total | | 685 | 99.6 | 100.0 | |
| Missing | System | 3 | .4 | | |
| Total | | 688 | 100.0 | | |

To show the distribution of values

Frequency polygons

Prior to using many statistical tests (Sections 12.8 to 12.11) it is necessary to establish the distribution of values for variables containing numerical data. For continuous data, this can be visualised by plotting a histogram or frequency polygon. For discrete data a bar graph or frequency polygon can be plotted. A **frequency polygon** is a line graph connecting the mid points of the bars of a histogram or bar graph (Figure 12.13). If your graph shows a bunching to the left and a long tail to the right, the data are **positively skewed** (Figure 12.7). If the converse is true, the data are **negatively skewed** (Figure 12.7). If your data are equally distributed either side of the highest frequency, then they are **symmetrically distributed**. A special form of the symmetric distribution, in which the data can be plotted as a bell-shaped curve, is known as **normal distribution** (Figure 12.7).

The other indicator of the distribution's shape is **kurtosis** – the pointedness or flatness of the distribution compared with normal distribution. If a distribution is more pointed or peaked, it is said to be leptokurtic, and the kurtosis value is positive. If a distribution is flatter, it is said to be platykurtic and the kurtosis value is negative. A distribution that is between the extremes of peakedness and flatness is said to be mesokurtic and has a kurtosis value of zero (Dancey and Reidy 2020).

Box plots

An alternative, often included in more advanced statistical analysis software, is the **box plot** (Figure 12.8). This provides a pictorial representation of the distribution of the data for a

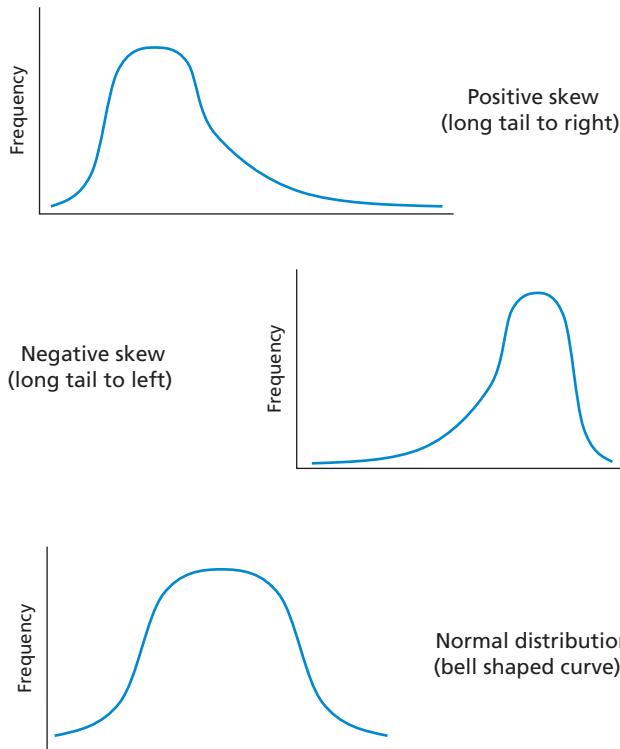


Figure 12.7 Frequency polygons showing distributions of values

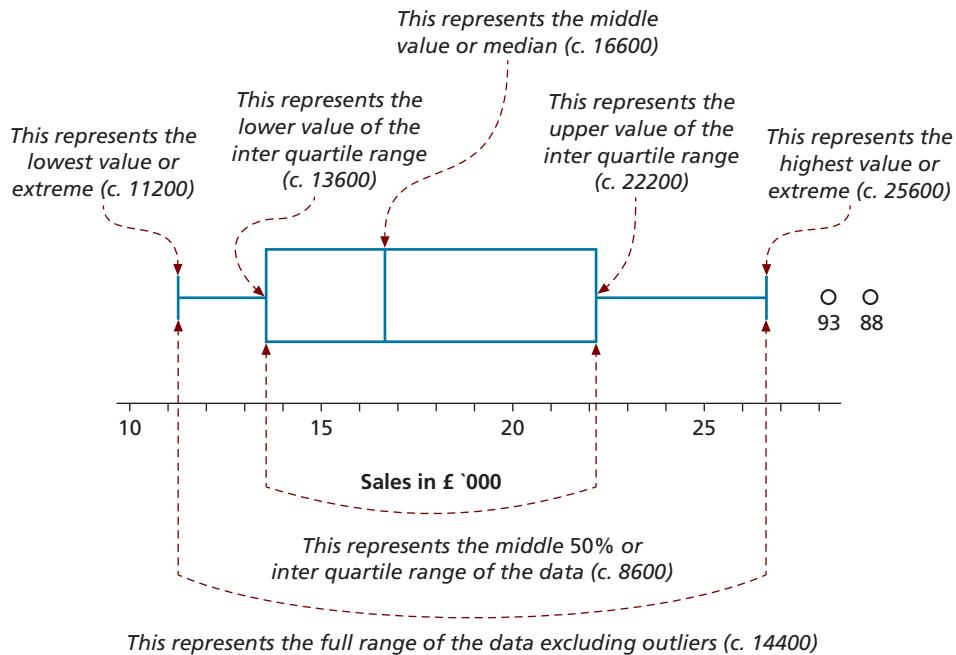


Figure 12.8 Annotated box plot

variable. The plot shows where the middle value or median is, how this relates to the middle 50 per cent of the data or inter-quartile range, and highest and lowest values or *extremes* (Section 12.5). It also highlights outliers, those values that are very different from the data. In Figure 12.8 the two outliers might be due to mistakes in data entry. Alternatively, they may be correct and emphasise that sales for these two cases (93 and 88) are far higher. In this example we can see that the data values for the variable are positively skewed as there is a long tail to the right.

12.7 Exploring and comparing two or more variables

To show interdependence and specific amounts

Contingency tables

As with individual variables, the best method of showing interdependence between variables so that any specific amount can be discerned easily is a table. This is known as a **contingency table** or as a **cross-tabulation** (Table 12.3). For variables where there are likely to be a large number of categories (or values for numerical data), you may need to group the data to prevent the table from becoming too large.

Most statistical analysis software allows you to add totals and row and column percentages when designing your table. Statistical analyses such as chi square can also be undertaken at the same time (Section 12.10).

To compare the highest and lowest values

Multiple bar graphs

Comparisons of variables that emphasise the highest and lowest rather than precise values are best explored using a **multiple bar graph**, also known as a **multiple bar chart** (Kosslyn

Table 12.3 Contingency table: Number of insurance claims by gender, 2022

| Number of claims* | Male | Female | Total |
|-------------------|--------------|--------------|--------------|
| 0 | 10032 | 13478 | 23510 |
| 1 | 2156 | 1430 | 3586 |
| 2 | 120 | 25 | 145 |
| 3 | 13 | 4 | 17 |
| Total | 12321 | 14937 | 27258 |

*No clients had more than three claims

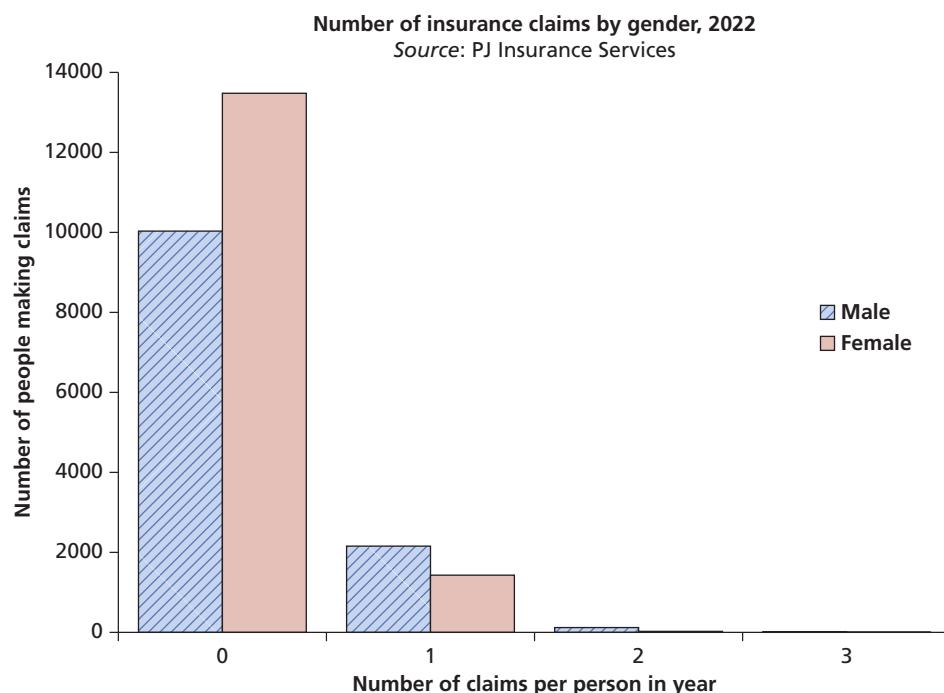
Source: PJ Insurance Services

2006), alternatively known as a **compound bar graph** or **compound bar chart**. As for a bar graph, continuous data – or data where there are many values or categories – need to be grouped. Within any multiple bar graph, you are likely to find it easiest to compare between adjacent bars. The multiple bar graph (Figure 12.9) has been drawn using the data in Table 12.3 to emphasise comparisons between males and females rather than between numbers of claims.

To compare proportions or percentages

Percentage component bar graphs

Comparison of proportions between variables uses either a **percentage component bar graph** (**percentage component bar chart** also known as a divided bar chart) or

**Figure 12.9** Multiple bar graph

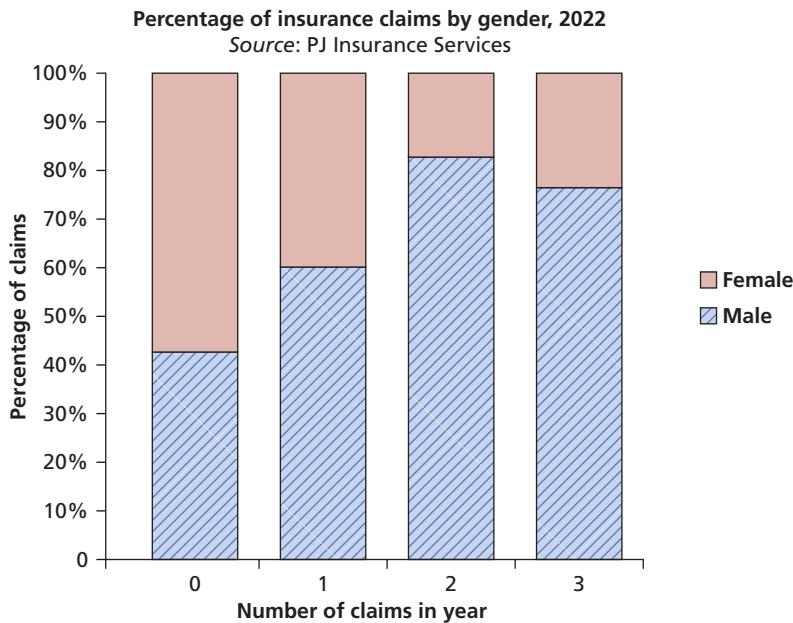


Figure 12.10 Percentage component bar graph

two or more pie charts. Either type of diagram can be used for all data types, provided that continuous data, and data where there are more than six values or categories, are grouped. Although percentage component bar graphs are more straightforward to draw than comparative pie charts when using most spreadsheets, comparative pie charts are often more appealing on infographics. Within your percentage component bar graphs, comparisons will be easiest between adjacent bars. The chart in Figure 12.10 has been drawn using the data in Table 12.3 to emphasise the proportions of males and females for each number of insurance claims in the year. Males and females, therefore, form a single bar.

To compare trends so the intersections are clear

Multiple line graphs

The most suitable diagram to compare trends for two or more numerical (or occasionally ranked) variables is a **multiple line graph** (Box 12.11) where one line represents each variable (Kosslyn 2006). You can also use multiple bar graphs in which bars for the same time period are placed adjacent to each other. Intersections in trends – that is, where values for two or more variables intersect – are shown by the place where the lines on a multiple line graph cross.

To compare the cumulative totals

Stacked bar graphs

Comparison of cumulative totals between variables uses a variation of the bar chart. A **stacked bar graph**, also known as a **stacked bar chart**, can be used for all data types provided that continuous data and data where there are more than six possible values

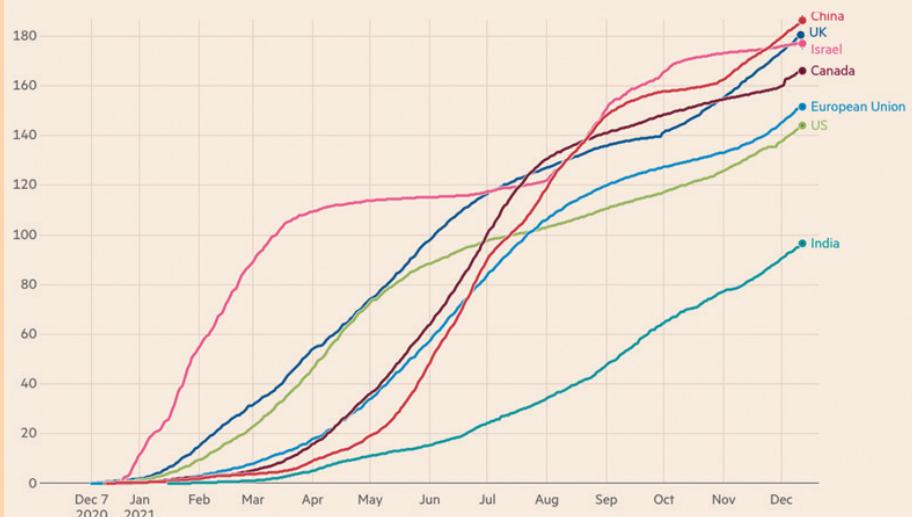


Box 12.11 Focus on research in the news

Vaccination doses administered in UK, Israel, US, European Union, Canada, China and India

Vaccination doses administered in UK, Israel, US, European Union, Canada, China and India

Cumulative doses administered per 100 residents



Some daily figures are estimates based on incomplete cumulative data.

Source: Our World in Data, World Health Organization, national sources, FT research.

Data updated December 13 2021 10:46am GMT. Interactive version: ft.com/covid-vaccine

Vaccine rollouts in advanced economies are largely outpacing those in emerging and developing economies even in countries with similar death rates. Officials at the World Health Organization have warned that the world is on the brink of ‘catastrophic moral failure’ as poor countries fall behind. Left unchecked, the virus could also mutate into strains that existing vaccines do not protect against.



Source: Abridged from ‘Covid-19 vaccine tracker: the global race to vaccinate’, FT Visual & Data Journalism team (2021) *Financial Times*, 13 December. Copyright © The Financial Times Ltd.

or categories are grouped. As with percentage component bar graphs, the design of the stacked bar graph is dictated by the totals you want to compare. For this reason, in Figure 12.11 (using data from Table 12.3) males and females have been stacked to give totals that can be compared for zero, one, two and three claims in a year.

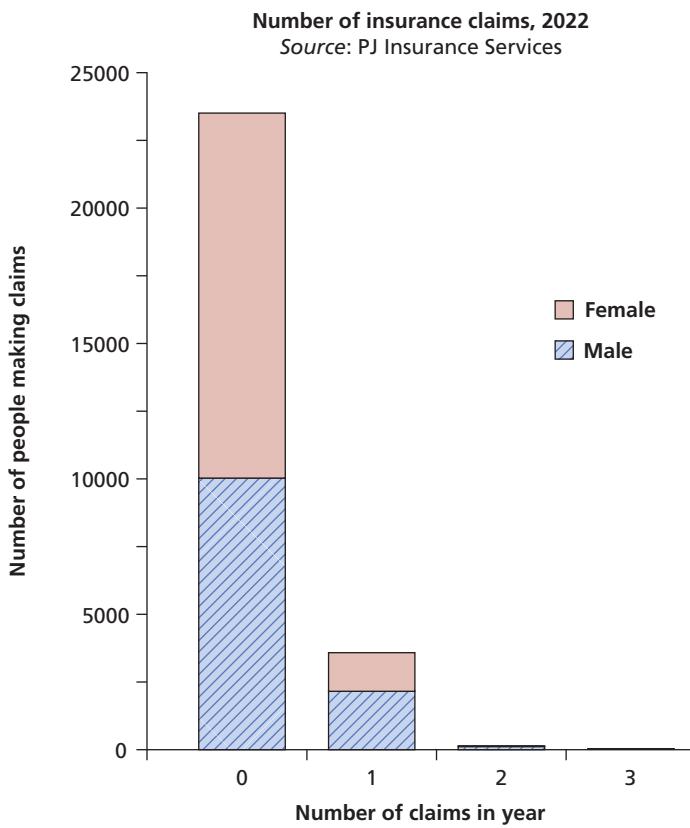


Figure 12.11 Stacked bar graph

To compare the proportions and cumulative totals

Comparative proportional pie charts

To compare both proportions of each category or value and the cumulative totals for two or more variables it is best to use **comparative proportional pie charts** for all data types. For each comparative proportional pie chart, the total area of the pie chart represents the total for that variable. By contrast, the angle of each segment represents the relative proportion of a category within the variable (Box 12.10). Because of the complexity of drawing comparative proportional pie charts, they are rarely used for exploratory data analysis, although they can be used to good effect in infographics and research reports.

To compare the distribution of values

Multiple frequency polygons, bar graphs and box plots

Often it is useful to compare the distribution of values for two or more variables. Plotting multiple frequency polygons (Box 12.11) or bar graphs (Figure 12.9) will enable you to compare distributions for up to three or four variables. After this your diagram is likely just to look a mess! An alternative is to use a diagram of multiple box plots, like the one

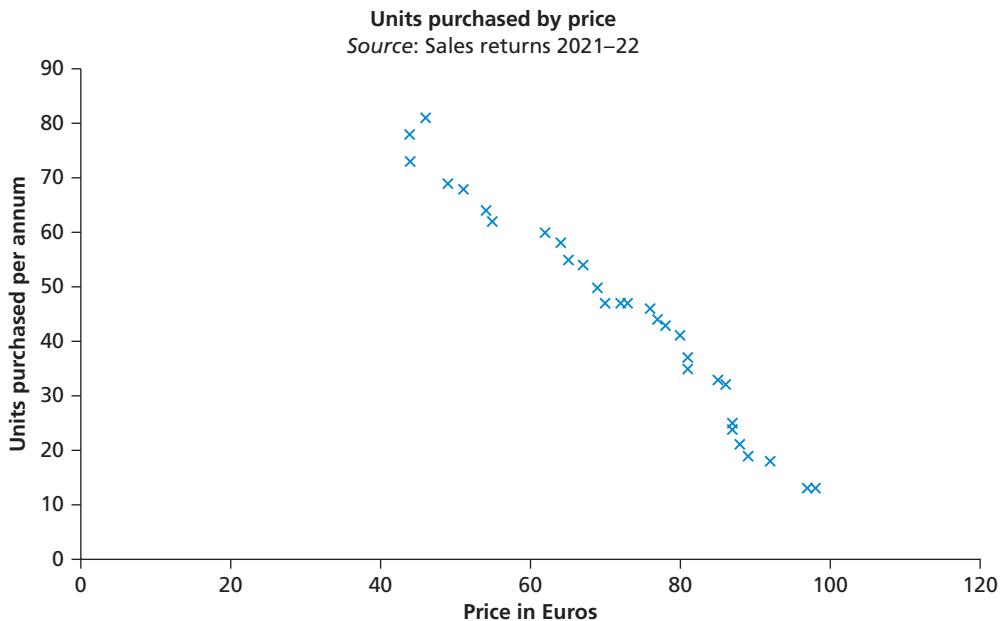


Figure 12.12 Scatter graph

in Figure 12.8. This provides a pictorial representation of the distribution of the data for the variables in which you are interested. These plots can be compared and are interpreted in the same way as the single box plot.

To show the interrelationships between cases for variables

Scatter graphs

You can explore possible interrelationships between ranked and numerical data variables by plotting one variable against another. This is called a **scatter graph** (also known as a **scatter plot**), and each cross (point) represents the values for one case (Figure 12.12). Convention dictates that you plot the **dependent variable** – that is, the variable that changes in response to changes in the other (**independent**) variable – against the vertical axis. The strength of the interdependence or relationship is indicated by the closeness of the points to an imaginary straight line. If as the values for one variable increase, so do those for the other then you have a positive relationship. If as the values for one variable decrease those for the other variable increase, then you have a negative relationship. Thus, in Figure 12.12 there is a negative relationship between the two variables. The strength of this relationship can be assessed statistically using techniques such as correlation or regression (Section 12.11).

12.8 Describing data using statistics

The exploratory data analysis approach (Section 12.5) emphasised the use of diagrams to understand your data. **Descriptive statistics** enable you to describe (and compare) a variable's data values numerically. Your research question(s) and objectives, although limited

by the type of data (Table 12.4), should guide your choice of statistics. Statistics to describe a variable focus on two aspects of the data values' distribution:

- the central tendency;
- the dispersion.

Table 12.4 Descriptive statistics by data type: a summary

| To calculate a measure of: | | Categorical | | Numerical | |
|----------------------------------|--|--------------------------|---------------------|---|----------|
| | | Nominal (Descriptive) | Ordinal (Ranked) | Continuous | Discrete |
| Central tendency that ... | ... represents the value that occurs <i>most frequently</i> | Mode | | | |
| | ... represents the <i>middle</i> value | | | Median | |
| | ... includes <i>all</i> data values (average) | | | Mean | |
| | ... includes all data values <i>other than those at the extremes</i> of the distribution | | | Trimmed mean | |
| Dispersion that ... | ... states the difference between the <i>highest and lowest</i> values | | | Range (data need not be normally distributed but must be placed in rank order) | |
| | ... states the difference within the <i>middle 50%</i> of values | | | Inter-quartile range (data need not be normally distributed but must be placed in rank order) | |
| | ... states the difference within <i>another fraction</i> of the values | | | Deciles or percentiles (data need not be normally distributed but must be placed in rank order) | |
| | ... describes the extent to which data values <i>differ from the mean</i> | | | Variance, or more usually, the standard deviation (data should be normally distributed) | |
| | ... compares the extent to which data values <i>differ from the mean</i> between variables | | | Coefficient of variation (data should be normally distributed) | |
| | ... allows the <i>relative extent</i> that data values differ to be compared | | | Index numbers | |

Source: © Mark Saunders, Philip Lewis and Adrian Thornhill 2022

When describing data for both samples and populations quantitatively it is usual to provide some general impression of values that could be seen as common, middling or average. These are termed measures of **central tendency** and are discussed in virtually all statistics textbooks. The three main ways of measuring the central tendency most used in business research are the:

- value that occurs most frequently (mode);
- middle value or mid-point after the data have been ranked (median);
- value, often known as the average, that includes all data values in its calculation (mean).

However, as we saw in Box 12.1, beware: if you have used numerical codes, most analysis software can calculate all three measures whether or not they are appropriate!

As well as describing the central tendency for a variable, it is important to describe how the data values are dispersed around the central tendency. As you can see from Table 12.4, this is only possible for numerical data. Two of the most frequently used ways of describing the dispersion are the:

- difference within the middle 50 per cent of values (inter-quartile range);
- extent to which values differ from the mean (standard deviation).

Although these **dispersion measures** are suitable only for numerical data, most statistical analysis software will also calculate them for categorical data if you have used numerical codes.

These measures of central tendency and dispersion are summarised in Table 12.4. Those most pertinent to your research question(s) and objectives will eventually be quoted in your project report as support for your arguments.

To describe the central tendency using the value that occurs most frequently

Modes

The **mode** is the value that occurs most frequently. For descriptive data, the mode is the only measure of central tendency that can be interpreted sensibly. You might read in a report that the most common (modal) colour of motor vehicles sold last year was silver, or that the two equally most popular makes of motorcycle in response to a questionnaire were Honda and Yamaha. In such cases where two categories occur equally most frequently, this is termed bimodal. The mode can be calculated for variables where there are likely to be a large number of categories (or values for numerical data), although it may be less useful. One solution is to group the data into suitable categories and to quote the most frequently occurring or **modal group**.

To describe the central tendency using the middle value

Medians

If you have quantitative data, it is also possible to calculate the middle or **median** value by ranking all the values in ascending order and finding the mid-point (or **50th percentile**) in the distribution. For variables that have an even number of data values, the median will occur halfway between the two middle data values. The median has the advantage that it is not affected by extreme values in the distribution (Box 12.12).



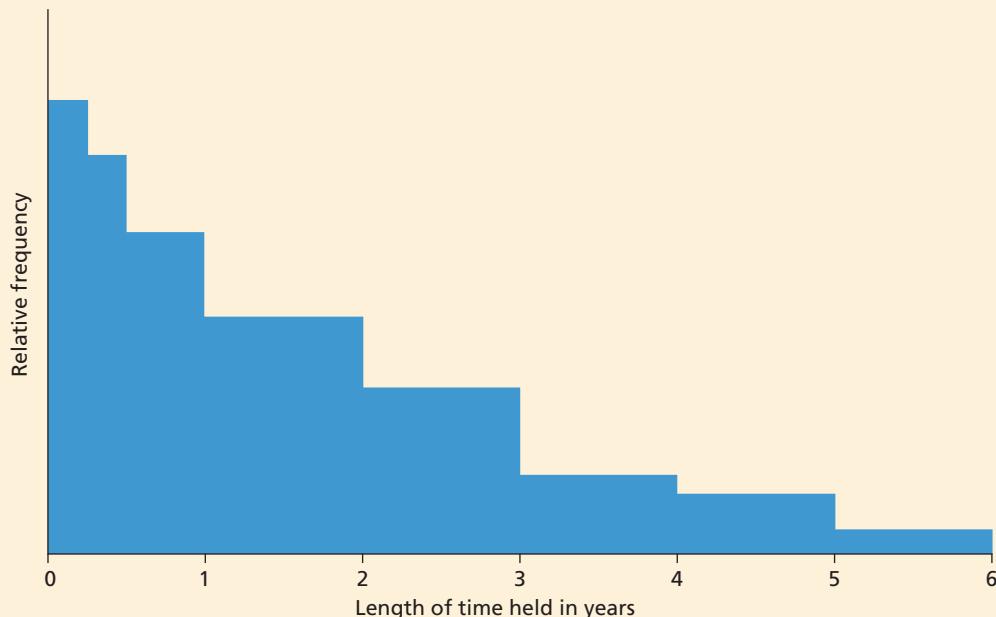
Box 12.12 Focus on student research

Describing the central tendency

As part of her research project, Kylie had obtained secondary data from the service department of her organisation on the length of time for which their customers had held service contracts.

| Length of time held contract | Number of customers |
|------------------------------|---------------------|
| < 3 months | 50 |
| 3 to < 6 months | 44 |
| 6 months to < 1 year | 71 |
| 1 to < 2 years | 105 |
| 2 to < 3 years | 74 |
| 3 to < 4 years | 35 |
| 4 to < 5 years | 27 |
| 5+ years | 11 |

As part of her exploratory analysis, she drew a histogram. This revealed a positively skewed distribution (long tail to the right).



From the table, the largest single group of customers were those who had contracts for 1 to less than 2 years. This was the modal time period (most commonly occurring). However, the usefulness of this statistic is limited owing to the variety of class widths. By definition, half of the organisation's customers will have held contracts below the median time period (approximately 1 year 5 months) and half above it. As there are 11 customers who have held service contracts for over 5 years, the mean time period (approximately 1 year 9 months) is pulled towards longer times. This is represented by the skewed shape of the distribution.

Kylie needed to decide which of these measures of central tendency to include in her research report. As the mode made little sense, she quoted the median and mean when interpreting her data:

The length of time for which customers have held service contracts is positively skewed. Although the mean length of time is approximately 1 year 9 months, half of customers have held service contracts for less than 1 year 5 months (median). Grouping of these data means that it is not possible to calculate a meaningful mode.

To describe the central tendency by including all data values

Means

The most frequently used measure of central tendency is the **mean** (average in everyday language), which includes all data values in its calculation. However, it is usually only possible to calculate a meaningful mean using numerical data.

The value of your mean is unduly influenced by extreme data values in skewed distributions (Section 12.6, Box 12.13). In such distributions the mean tends to get drawn towards the long tail of extreme data values and may be less representative of the central tendency. For this and other reasons Anderson et al. (2020) suggest that the median may be a more useful descriptive statistic. Alternatively, where the mean is affected by extreme data values (outliers) these may be excluded and a **trimmed mean** calculated. This excludes a certain proportion (for example, 5 per cent) of the data from both ends of the distribution, where the outliers are located. Because the mean is the building block for many of the statistical tests used to explore relationships (Section 12.9), it is usual to include it as at least one of the measures of central tendency for numerical data in your report.



Box 12.13 Focus on management research

Accounting for modern slavery

In a recent paper in the *Accounting, Auditing and Accountability Journal*, Christ et al. (2019) analyse the 100 top Australian companies' voluntary practices in supply chains disclosures about modern slavery. Looking at these companies' annual and standalone reports and websites they reveal that the volume of disclosures overall is low and, with 37 of the 100 companies having disclosed a modern slavery statement. A wide range of themes of modern slavery were disclosed with themes of bribery and corruption and human rights issues the most dominant.

Within their analysis of supply chain themes disclosed Christ and colleagues provide details of the overall mean and maximum number of sentences relating to supply chain disclosures and a breakdown for a range of different supply chain modern slavery

themes. Overall disclosures range from a total sample mean number of 2.62 sentences in annual reports to a total sample mean number of 69.04 sentences on companies' websites, standalone reports having a total sample mean of 7.38 sentences.

Within annual reports, the theme with the highest mean number of sentences was 'code of conduct' (mean = 0.76 sentences) and 'modern slavery in general' (mean = 0.44 sentences), whereas for websites it was 'bribery and corruption' (mean = 17.73 sentences), modern slavery in general (mean = 12.08 sentences) and 'human rights' (mean = 7.32 sentences). All themes were reported at greater length in websites than in annual reports. The maximum number of sentences on a single website (179) related to 'bribery and corruption' whereas the maximum number of sentences in an annual report related to 'code of conduct' (17) or 'modern slavery in general' (17).

Based on their analysis, Christ and colleagues conclude that, although there appear to be habitual processes encouraging disclosure about modern slavery, legislation is required to encourage further engagement.

To describe the dispersion by stating the difference between values

Range and quartiles

To get a quick impression of the distribution of data values for a variable you could simply calculate the difference between the lowest and the highest values once they have been ranked in ascending order – that is, the **range**. However, this statistic is rarely used in research reports as it represents only the extreme values.

A more frequently used statistic is the **inter-quartile range**. As we discussed earlier, the median divides the range into two. The range can be further divided into four equal sections called **quartiles**. The **lower quartile** is the value below which a quarter of your data values will fall; the **upper quartile** is the value above which a quarter of your data values will fall. As you would expect, the remaining half of your data values will fall between the lower and upper quartiles. The difference between the upper and lower quartiles is the inter-quartile range (Anderson et al. 2020). As a consequence, it is concerned only with the middle 50 per cent of data values and ignores extreme values.

Percentiles and deciles

You can also calculate the range for other fractions of a variable's distribution. One alternative is to divide your distribution using **percentiles**. These split your ranked distribution into 100 equal parts. Obviously, the lower quartile is the 25th percentile and the upper quartile the 75th percentile. However, you could calculate a range between the 10th and 90th percentiles so as to include 80 per cent of your data values. Another alternative is to divide the range into 10 equal parts called **deciles**.

To describe the dispersion by comparing the extent by which values differ from the mean

Standard deviation

Conceptually and statistically in research it is important to look at the extent to which the data values for a variable are spread around their mean, as this is what you need to know to assess its usefulness as a typical value for the distribution. If your data values are all close to the mean, then the mean is more typical than if they vary widely. To describe the extent of spread of numerical data you use the **standard deviation**. If your data are a sample (Section 7.1), this is calculated using a slightly different formula than if your data are a population, although if your sample is larger than about 30 cases there is little difference in the two statistics.

Coefficient of variation

You may need to compare the relative spread of data between distributions of different magnitudes (e.g. one may be measured in hundreds of tonnes, the other in billions of tonnes). To make a meaningful comparison you will need to take account of these different magnitudes. A common way of doing this is:

- 1 to divide the standard deviation by the mean;
- 2 then to multiply your answer by 100.

This results in a statistic called the **coefficient of variation** (Black 2020). The values of this statistic can then be compared. The distribution with the largest coefficient of variation has the largest relative spread of data (Box 12.14).



Box 12.14 Focus on student research

Describing variables and comparing their dispersion

Cathy was interested in the total value of transactions at the main and sub-branches of a major bank. The mean value of total transactions at the main branches

was approximately five times as high as that for the sub-branches. This made it difficult to compare the relative spread in total value of transactions between the two types of branches. By calculating the coefficients of variation, Cathy found that there was relatively more variation in the total value of transactions at the main branches than at the sub-branches. This is because the coefficient of variation for the main branches was larger (23.62) than the coefficient for the sub-branches (18.08).

| | A | B | C | D |
|---|-------------|------------------------------|--------------------|--------------------------|
| 1 | Branch type | Mean total transaction value | Standard deviation | Coefficient of variation |
| 2 | Main | £6,000,000 | £1,417,000 | 23.62 |
| 3 | Sub | £1,200,000 | £217,000 | 18.08 |
| 4 | | | | |

Index numbers

Alternatively, as discussed at the start of the chapter in relation to the Economist's rEV Index, you may wish to compare the relative extent to which values differ. One way of doing this is to use **index numbers** and consider the relative differences rather than actual data values. Such indices compare each data value against a base data value that is normally given the value of 100, differences being calculated relative to this value. An index number greater than 100 represents a larger or higher data value relative to the base value and an index less than 100, a smaller or lower data value.

To calculate an index number for each case for a data variable you use the following formula:

$$\text{Index number for case} = \frac{\text{data value for case}}{\text{base data value}} \times 100$$

We discuss index numbers further when we consider examining trends (Section 12.13).

12.9 Statistical tests' assumptions and hypothesis testing

When examining associations, differences, relationships, predictions and trends using statistics, it is important to ensure that each data variable meets the assumptions of the specific statistic you intend use. These assumptions can include:

- suitability of statistical test for the data type(s);
- whether the data variable is normally distributed;
- sample size required.

Providing these assumptions are met, you should usually be able to assess the significance of your findings using hypothesis testing. As you do this, it will be important to consider the:

- types of hypotheses;
- statistical and practical significance of your findings;
- need to minimise errors when making inferences, in particular by wrongly rejecting a null hypothesis and saying something is true when it is not (a Type I error).

Statistical tests' assumptions

Suitability of the statistical test for the data type

The suitability of the two main groups of statistical tests, non-parametric and parametric, depends upon data type. **Non-parametric statistics** are designed primarily for use with categorical (dichotomous, nominal and ordinal) data where there is no distributional model and so we cannot use statistics to estimate parameters. In contrast, **parametric statistics** are used with numerical (interval and ratio) data. Although parametric statistics are considered more powerful because they use numerical data, a number of assumptions about the actual data being used need to be satisfied if they are not to produce spurious results (Blumberg et al. 2014). These include the:

- data cases selected for the sample should be independent – in other words the selection of any one case for your sample should not affect the probability of any other case being included in the same sample;
- data cases should be drawn from normally distributed populations (Section 12.6 and later in this section);
- populations from which the data cases are drawn should have equal variances (Sections 12.6 and 12.10);
- data used should be numerical.

The normal distribution

As we have already noted, parametric tests assume that the numerical data cases in your sample are drawn from normally distributed populations. This means that the data values for each quantitative variable should also be normally distributed, being clustered around the variable's mean in a symmetrical pattern forming a bell-shaped frequency distribution (Figure 12.13). Fortunately, it is relatively easy to check if data values for a particular variable are distributed normally, both using graphs and statistically.

In Section 12.6 we looked at a number of different types of graphs including histograms (Figure 12.5), box plots (Figure 12.8) and frequency polygons (Figure 12.7). All of these can be used to assess visually whether the data values for a particular numerical variable are clustered around the mean in a symmetrical pattern, and so normally distributed. For normally distributed data, the value of the mean, median and mode are also likely to be the same.

Another way of testing for normality is to use statistics to establish whether the distribution as a whole for a variable differs significantly from a comparable normal distribution. Fortunately, this is relatively easy to do in statistical software such as IBM SPSS Statistics using the **Kolmogorov–Smirnov test** and the **Shapiro–Wilk test** (Box 12.15), as the software also calculates a comparable normal distribution automatically. For both these tests the calculation consists of the test statistic (labelled D and W respectively), the degrees of

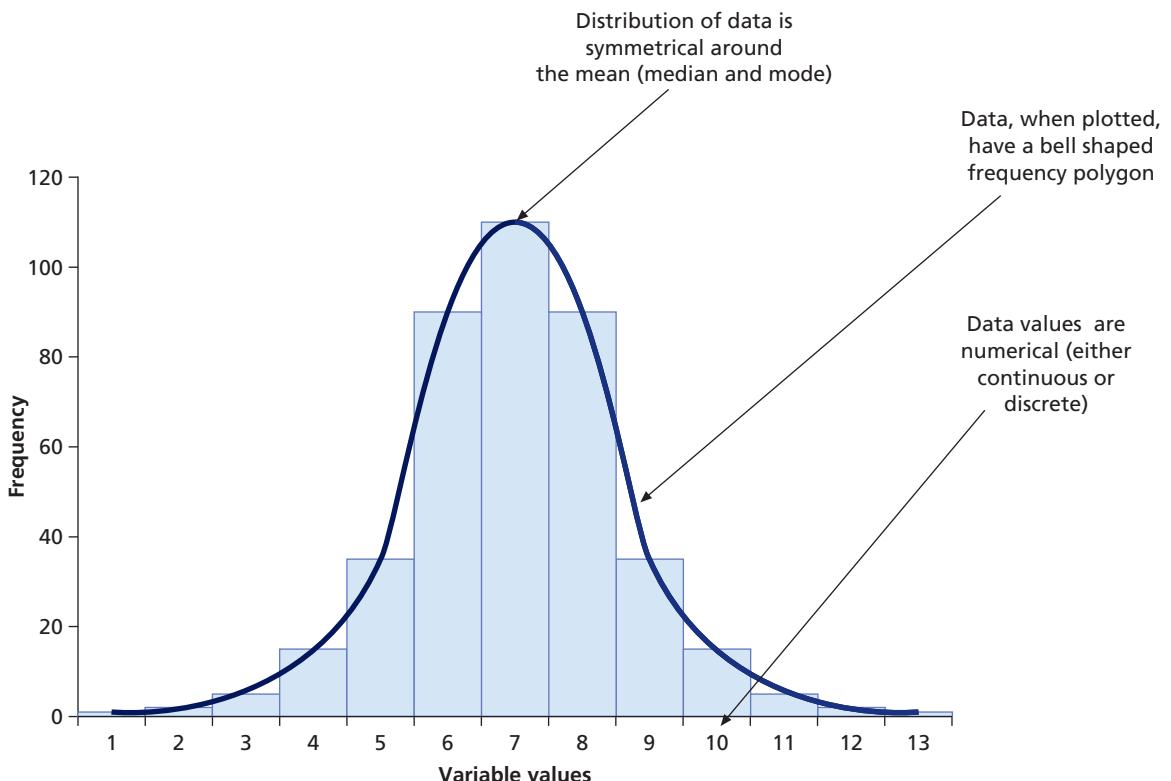


Figure 12.13 Annotated frequency polygon showing the normal distribution

freedom¹ (df) and, based on this, the probability (p -value). The p -value is the probability of the data for your variable, or data more extreme, occurring by chance alone from a comparable normal distribution for that variable if there really was no difference. For either statistic, a probability of 0.05 means there is a 5 per cent likelihood of the actual data distribution or one more extreme occurring by chance alone from a comparable normal distribution if there was no real difference. Therefore, a probability of 0.05 or lower² for either statistic means that these data are unlikely to be normally distributed. When interpreting probabilities from software packages, beware: owing to statistical rounding of numbers a probability of 0.000 does not mean zero but that it is less than 0.001 (Box 12.15). If the probability is greater than 0.05, then this is interpreted as the data being likely to be normally distributed. However, with very large samples it is easy to obtain statistically significant differences between a sample variable and a comparable normal distribution when actual differences are quite small. For this reason, it is often helpful to also use a graph to make an informed decision.

¹Degrees of freedom are the number of values free to vary when computing a statistic. The number of degrees of freedom for a contingency table of at least two rows and two columns of data is calculated from: (number of rows in the table–1) × (number of columns in the table–1).

²A probability of 0.05 means that the probability of your test result or one more extreme occurring by chance alone, if there really was no difference, is 5 in 100, that is 1 in 20.

Sample size

In addition, as we will discuss later, you need to ensure that your sample size is sufficiently large to meet the requirements of the statistic you are using (see also Section 7.2). If the assumptions are not satisfied, it is often still possible to use non-parametric statistics.

Hypothesis testing and statistical significance

Assessing the statistical significance of relationships and differences between variables usually involves testing a hypothesis. A **hypothesis** is a tentative, usually testable, explanation that there is an association, difference or relationship between two or more variables.

Types of hypotheses

As part of your research project, you might collect sample data to examine the association between two variables. You will phrase this as a testable explanation that puts forward the absence of that relationship (termed a **null hypothesis**) such as: ‘there is no association between . . .’ Once you have entered data into the analysis software, chosen the statistic and clicked on the appropriate icon, an answer will appear. With most statistical analysis software this consists of a test statistic, the degrees of freedom (*df*) and, based on these, the statistical significance (*p*-value). This is the probability that the value of the test statistic summarising a specific aspect of your data would be equal to or more extreme than its actual observed value, given the specified assumptions of that test (Wasserstein and Lazar 2016).

If the probability of your test statistic value or one more extreme having occurred is less than a prescribed significance value (usually $p < 0.05$ or lower³), this is usually interpreted as casting doubt on or providing evidence against your null hypothesis and the associated underlying assumptions. This means your data are more likely to support the explanation expressed in your hypothesis; in this example a testable statement such as: ‘There is an association between . . .’ Statisticians refer to this as rejecting the null hypothesis and accepting the hypothesis, often abbreviating the terms null hypothesis to H_0 and hypothesis to H_1 . Consequently, rejecting a null hypothesis could mean casting doubt on an explanation such as ‘there is no difference between . . .’ or ‘there is no relationship between . . .’ and accepting an explanation such as ‘there is a difference between . . .’ or ‘there is a relationship between . . .’ However, conclusions and policy decisions should not be based just on whether the *p*-value passes a specific threshold. Contextual factors such as the research design, quality of data and other external evidence are also important in interpreting the findings (Wasserstein and Lazar 2016). If the probability of obtaining the test statistic or one more extreme by chance alone is greater than or equal to a prescribed value (usually $p = 0.05$), this is normally interpreted as your data being compatible with the explanation expressed by your null hypothesis and its associated underlying assumptions. This indicates the null hypothesis can be accepted and is referred to by statisticians as failing to reject the null hypothesis. There may still be a relationship between the variables under such circumstances, but you cannot make the conclusion with any certainty. Remember, when interpreting probabilities from software packages, beware: owing to statistical rounding of numbers a probability of 0.000 does not mean zero, but that it is less than 0.001 (Box 12.15).

³A probability of 0.05 means that the probability of your test result or one more extreme occurring by chance alone, if there really was no difference in the population from which the sample was drawn (in other words if the null hypothesis was true), is 5 in 100, that is 1 in 20.



Box 12.15 Focus on student research

Testing for normality

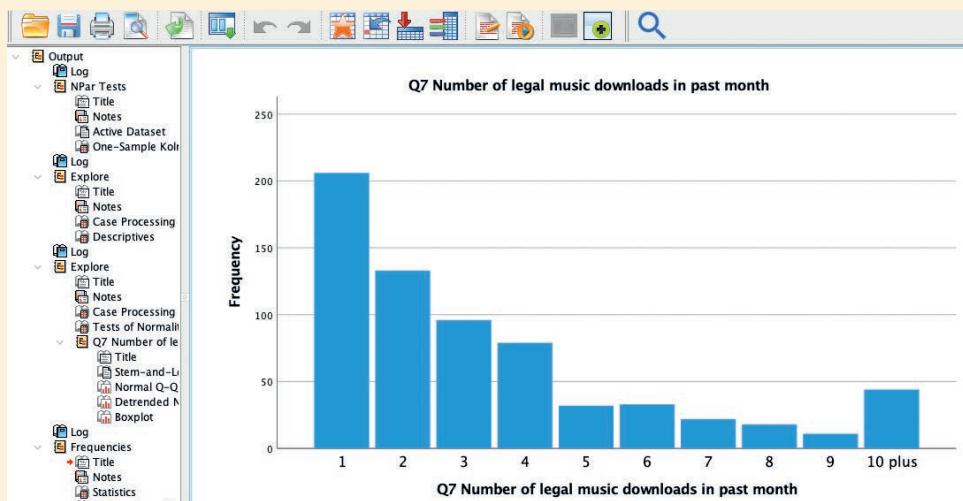
As part of his research project, Osama had collected quantitative data about music piracy and illegal

downloading of music from student respondents. Before undertaking his statistical analysis, Osama decided to test his quantitative variables for normality using the Kolmogorov–Smirnov test and the Shapiro–Wilk test. The output from IBM SPSS Statistics for one of his data variables, ‘number of legal music downloads made in the past month’, follows:

| Tests of Normality | | | | | | |
|--|---------------------------------|-----|-------|-----------|-----|-------|
| | Kolmogorov-Smirnov ^a | df | Sig. | Statistic | df | Sig. |
| Q7 Number of legal music downloads in past month | .201 | 674 | <.001 | .815 | 674 | <.001 |
| a. Lilliefors Significance Correction | | | | | | |

This calculated the significance (Sig.) for both the Kolmogorov–Smirnov test and the Shapiro–Wilk test as ‘000’, meaning that for this variable the likelihood of the actual distribution or one more extreme differing from a normal distribution occurring by chance alone was less

than 0.001. Consequently, the data values for variable ‘Number of legal music downloads in past month’ were not normally distributed, reducing his choice of statistics for subsequent analyses. This was confirmed by a bar chart showing the distribution of the data for the variable:



Osama reported the outcome of this analysis in his project report, quoting the test statistics ‘D’ and ‘W’ and their associated degrees of freedom ‘df’ and probabilities ‘p’ in brackets:

Tests for normality revealed that data for the variable ‘number of legal music downloads in the past month’ cast considerable doubt on the data being normally distributed [D = 0.201, df = 674, p < 0.001; W = 0.815, df = 674, p < 0.001].

The hypothesis and null hypothesis we have just stated are often termed **non-directional**. This is because they refer to a difference rather than also including the nature of the difference. A **directional hypothesis** includes within the testable statement the direction of the difference, for example ‘larger’. This is important when interpreting the probability of obtaining the test result, or one more extreme, by chance. Statistical software (Box 12.19) often states whether this probability is one-tailed or two-tailed. Where you have a directional hypothesis such as when the direction of the difference is larger, you should use the one-tailed probability. Where you have a non-directional hypothesis and are only interested in the difference, you should use the two-tailed probability.

While considering hypothesis testing, it is worth mentioning that some quantitative analyses, when written up, do not specify actual hypotheses. Rather, the theoretical underpinnings of the research and the research questions provide the context within which the probability of relationships between variables occurring by chance alone is tested. Thus, although hypothesis testing has taken place, statistical significance is often only discussed in terms of the probability (*p*-value) of the test statistic value or one more extreme occurring by chance.

Statistical and practical significance

The probability of a test statistic value or one more extreme occurring by chance is determined in part by your sample size (Section 7.5). One consequence of this is that it is very difficult to obtain a low *p*-value for a test statistic with a small sample. Conversely, by increasing your sample size, less obvious relationships and differences will be found to be statistically significant until, with extremely large samples, almost any relationship or difference will be significant (Anderson 2003). This is inevitable as your sample is becoming closer in size to the population from which it was selected. You therefore need to remember that small populations can make statistical tests insensitive, while very large samples can make statistical tests overly sensitive. There are two consequences to this:

- if you expect a difference, relationship or association will be small, you need to have a larger sample size;
- if you have a large sample and the difference, relationship or association has statistical significance, you need also to assess the practical significance of this relationship.

Both these points are crucial as it is not unusual for a test statistic to be statistically significant but trivial in the real world. Fortunately, it is relatively straightforward to assess the practical significance of something that is statistically significant by calculating an appropriate **effect size index**. These indices measure the size of either differences between groups (the *d* family of statistical tests) or association between groups (the *r* family of statistical tests) and an excellent discussion can be found in Ellis (2010).

Minimising errors when making inferences

Inevitably, errors can occur when making inferences from samples. Statisticians refer to these as Type I and Type II errors. Blumberg et al. (2014) use the analogy of legal decisions to explain Type I and Type II errors. In their analogy they equate a Type I error to a person who is innocent being unjustly convicted and a Type II error to a person who is guilty of a crime being unjustly acquitted. In business and management research we would say that an error made by wrongly rejecting a null hypothesis and therefore accepting the hypothesis is a **Type I error**. Type I errors might involve you concluding that two variables are related when they are not, or incorrectly concluding that a sample statistic exceeds the value that would be expected by chance alone. This means you are rejecting your null hypothesis when you should not. The term ‘**statistical significance**’ discussed earlier

| | | Likelihood of making a | |
|-----------------------|------|------------------------|------------------|
| | | Type I error | Type II error |
| Significance level at | 0.05 | Increased | Decreased |
| | 0.01 | Decreased | Increased |

Figure 12.14 Type I and Type II errors

therefore refers to the probability of making a Type I error. A **Type II error** involves the opposite occurring. In other words, you fail to reject your null hypothesis when it should be rejected. This means that Type II errors might involve you in concluding that two variables are not related when they are, or that a sample statistic does not exceed the value that would be expected by chance alone.

Given that a Type II error is the inverse of a Type I error, it follows that if we reduce our likelihood of making a Type I error by setting the significance level to 0.01 rather than 0.05, we increase our likelihood of making a Type II error by a corresponding amount. This is not an insurmountable problem, as researchers usually consider Type I errors more serious and prefer to take a small likelihood of saying something is true when it is not (Figure 12.14). It is therefore generally more important to minimise Type I than Type II errors.

Assessing statistical significance

In examining associations, differences, relationships, predictions and trends you are likely to ask questions such as: 'Do these groups differ?', 'Are these variables related?', or 'Can I predict this variable from another?' or 'What is the trend?'. Through statistical analysis you will establish the probability of the test statistic summarising what you have found in your data, or a finding more extreme, occurring. This process of assessing the statistical significance of findings from a sample is known as **significance testing**, the classical approach to significance testing being **hypothesis testing**. Significance testing can therefore be thought of as assessing the possibility that your result could be due to random variation in your sample.

The way in which statistical significance is assessed using both non-parametric and parametric statistics can be thought of as answering one from a series of questions, dependent on the data type:

- Is the independence or association statistically significant?
- Are the differences statistically significant?
- What is the strength of the relationship and is it statistically significant?
- Are the predicted values statistically significant?

When assessing significance each question will usually be phrased as a hypothesis about the association, difference, relationship, or prediction regarding two or more variables. The questions and associated statistics are summarised in Table 12.5 along with statistics used to help examine trends.

Chapter 12 Analysing data quantitatively

Table 12.5 Statistics to examine associations, differences, relationships, predictions and trends by data type: A summary

| | Categorical | | Numerical | | | |
|---|---|--|--|----------|--|--|
| | Nominal (Descriptive) | Ordinal (Ranked) | Continuous | Discrete | | |
| To test <i>normality</i> of a distribution | | | Kolmogorov–Smirnov test, Shapiro–Wilk test | | | |
| To test whether two variables are <i>independent</i> | Chi square (data may need grouping) | | Chi square if variable grouped into discrete classes | | | |
| To test whether two variables are <i>associated</i> | Cramer's V and Phi (both variables must be dichotomous) | | | | | |
| To test whether two groups (categories) are <i>different</i> | | Kolmogorov–Smirnov (data may need grouping) or Mann–Whitney <i>U</i> test | Independent <i>t</i> -test or paired <i>t</i> -test (often used to test for changes over time) or Mann–Whitney <i>U</i> test (where data skewed or a small sample) | | | |
| To test whether three or more groups (categories) are <i>different</i> | | | | | | |
| To assess the <i>strength of relationship</i> between two variables | | Spearman's rank correlation coefficient (Spearman's rho) or Kendall's rank order correlation coefficient (Kendall's tau) | Pearson's product moment correlation coefficient (PMCC) | | | |
| To assess the <i>strength of a relationship</i> between one dependent and one independent variable | | | | | | |
| To assess the <i>strength of a relationship</i> between one dependent and two or more independent variables | | | | | | |
| To <i>predict</i> the value of a dependent variable from one or more independent variables | | | | | | |
| To <i>establish the trend</i> (explore relative change) over time | | | | | | |
| To <i>compare trends</i> (relative changes) over time | | | | | | |
| To <i>determine the trend</i> over time of a series of data | | | | | | |

Source: © Mark Saunders, Philip Lewis and Adrian Thornhill 2022

12.10 Examining associations and differences

To test whether two variables are independent or associated (using categorical data)

Chi square test

Often descriptive or numerical data will be summarised as categorical data using a two-way contingency table (such as Table 12.3). The **chi square test** (χ^2) enables you to find out how likely it is that the two variables are independent. It is based on a comparison of the observed values in the table with what might be expected if the two distributions were entirely independent. Therefore, you are assessing the likelihood of the data in your table, or data more extreme, occurring by chance alone by comparing it with what you would expect if the two variables were independent of each other. This could be phrased as the null hypothesis: ‘there is no dependence . . .’.

The test relies on:

- the categories used in the contingency table being mutually exclusive, so that each observation falls into only one category or class interval;
- no more than 25 per cent of the cells in the table having expected values of less than 5. For contingency tables of two rows and two columns, no expected values of less than 10 are preferable (Dancey and Reidy 2020).

If the latter assumption is not met, the accepted solution is to combine rows and columns where this produces meaningful data.

Most statistical analysis software calculates the chi square statistic, degrees of freedom⁴ and the *p*-value automatically. However, if you are using a spreadsheet, you will usually need to look up the probability in a ‘critical values of chi square’ table using your calculated chi square value and the degrees of freedom. There are numerous copies of this table online. A probability of 0.05 means that there is only a 5 per cent likelihood of the data in your table or data more extreme occurring by chance alone and is usually considered statistically significant. Therefore, a probability of 0.05 or smaller means you can be at least 95 per cent certain that the dependence between your two variables represented by the data in the table could not have occurred by chance alone.

Cramer's V

Some software packages, such as IBM SPSS Statistics, calculate the statistic **Cramer's V** alongside the chi square statistic (Box 12.16). If you include the value of Cramer's V in your research report, it is usual to do so in addition to the chi square statistic. Whereas the chi square statistic gives the probability of data in a table, or data more extreme, occurring by chance alone, Cramer's V measures the association between the two variables within the table on a scale where 0 represents no association and 1 represents perfect association. Because the value of Cramer's V is always between 0 and 1, the relative strengths of associations between different pairs of variables that are considered statistically significant can be compared.

⁴ Degrees of freedom are the number of values free to vary when computing a statistic. The number of degrees of freedom for a contingency table of at least two rows and two columns of data is calculated from (number of rows in the table–1) × (number of columns in the table–1).



Box 12.16 Focus on student research

Testing whether two variables are independent or associated

As part of his research project, John wanted to find out whether there was a significant dependence between salary grade of respondent and

gender. Earlier analysis using IBM SPSS Statistics had indicated that there were 385 respondents in his sample with no missing data for either variable. However, it had also highlighted there were only 14 respondents in the five highest salary grades (GC01 to GC05).

Bearing in mind the assumptions of the chi square test, John decided to combine salary grades GC01 through GC05 to create a combined grade GC01–5 using IBM SPSS Statistics:

| Grade (current) | *Gender | | Total |
|-----------------|---------|--------|-------|
| | Male | Female | |
| GC01-5 | 14 | 2 | 16 |
| GC06 | 19 | 4 | 23 |
| GC07 | 61 | 11 | 72 |
| GC08 | 65 | 25 | 90 |
| GC09 | 97 | 87 | 184 |
| Total | 256 | 129 | 385 |

He then used his analysis software to undertake a chi square test and calculate Cramer's V.

| | Value | df | Asymptotic Significance (2-sided) |
|--------------------|---------------------|----|-----------------------------------|
| Pearson Chi-Square | 33.587 ^a | 4 | <.001 |
| Likelihood Ratio | 35.279 | 4 | <.001 |
| N of Valid Cases | 385 | | |

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 5.36.

| Symmetric Measures | | | |
|--------------------|-------|--------------------------|--|
| | Value | Approximate Significance | |
| Nominal by Nominal | | | |
| Phi | .295 | <.001 | |
| Cramer's V | .295 | <.001 | |
| N of Valid Cases | 385 | | |

As can be seen, this resulted in an overall chi square value of 33.59 with 4 degrees of freedom (*df*).

The significance of .000 (Asymp. Sig. – two sided) meant that the probability of the values in his table or values more extreme occurring by chance alone was less than 0.001. He therefore concluded that the

gender and grade were extremely unlikely to be independent and quoted the statistic in his project report:

$$\chi^2 = 33.59, df = 4, p < 0.001^*$$

The Cramer's V value of .295, significant at the 0.001 level (Approx. Sig.), showed that the association

between gender and salary grade, although weak, could be considered significant. This indicated that men (coded 1 whereas females were coded 2) were more likely to be employed at higher salary grades GC01–5 (coded using lower numbers). John also quoted this statistic in his project report:

$$[V_c = 0.295, p < 0.001]$$

To explore this association further, John examined the cell values in relation to the row and column totals. Of males, 5 per cent were in higher salary grades (GC01–5) compared to less than 2 per cent of females. In contrast, only 38 per cent of males were in the lowest salary grade (GC09) compared with 67 per cent of females.

*You will have noticed that the computer printout in this box does not have a zero before the decimal point. This is because most software packages follow the North American convention of not placing a zero before the decimal point.

Phi

An alternative statistic used to measure the association between two variables is **Phi**. This statistic measures the association on a scale between –1 (perfect negative association), through 0 (no association) to 1 (perfect association). However, unlike Cramer's V, using Phi to compare the relative strengths of associations between pairs of variables considered statistically significant can be problematic. This is because, although values of Phi will only range between –1 and 1 when measuring the association between two dichotomous variables, they may exceed these extremes when measuring the association for categorical variables where at least one of these variables has more than two categories. For this reason, we recommend that you use Phi only when comparing pairs of dichotomous variables.

To test whether two groups are different (using ranked data)

Kolmogorov-Smirnov two-sample test

Sometimes you need to establish whether the distribution of an observed set of values for each category of a variable differs from a specified distribution other than the normal distribution, for example whether your sample differs from the population from which it was selected. The **Kolmogorov-Smirnov two-sample test** enables you to establish this for ranked data (Corder and Foreman 2014). It is based on a comparison of the cumulative proportions of the observed values in each category of your sample with the cumulative proportions in the same categories for a second 'sample' such as the population from which it was selected. Therefore, you are testing the likelihood of the distribution of your observed values differing from that of the specified population by chance alone.

The Kolmogorov-Smirnov two-sample test calculates a k_s statistic and an associated probability that the distribution in the first sample or one more extreme differs from the distribution in the second sample by chance (Corder and Foreman 2014). Although the two-sample test statistic is not often found in analysis software other than for comparisons with a normal distribution (Sections 12.6 and 12.9), it is easily accessible online (Box 12.17). A test statistic with a p -value of 0.05 means that there is only a 5 per cent likelihood that



Box 12.17 Focus on student research

Testing the representativeness of a sample

Jaimie's research question was: 'To what extent are my organisation's espoused customer service values evident in customer facing employees' views of the service they provide to customers?' As part of her research, she emailed a link to a Web questionnaire to the 217 employees in the organisation where she worked and 94 of these responded. The responses from each category of employee in terms of their seniority within the organisation's hierarchy were as shown in the table below.

Using an online Kolmogorov-Smirnov two-sample test calculator (SciStatCalc 2013) Jaimie calculated a Kolmogorov-Smirnov test statistic (ks) of 0.632 with a p -value of 0.819. This meant that the probability of the distribution in her sample (or one more extreme) differing from that of the organisation's employees having occurred by chance alone was 0.819; in other words, more than 80 per cent. She concluded that those employees who responded were unlikely to differ significantly from the total population in terms of their seniority within the organisation's hierarchy. This was stated in her research report.

Statistical analysis revealed the sample selected was very unlikely to differ significantly from all employees in terms of their seniority within the organisation's hierarchy [$ks = .632$, $p = .819$].

| | | Shop floor workers | Technicians | Supervisors | Quality managers | Management team | Total |
|------------------------|---------|--------------------|-------------|-------------|------------------|-----------------|-------|
| Respondents | Number | 48 | 29 | 8 | 6 | 3 | 94 |
| | Percent | 51.1 | 30.9 | 8.5 | 6.4 | 3.2 | 100 |
| Total employees | Number | 112 | 68 | 22 | 14 | 1 | 217 |
| | Percent | 51.6 | 31.3 | 10.1 | 6.5 | 0.5 | 100 |

the distribution in the sample or one more extreme differs from that in the second sample by chance alone and, is usually considered statistically significant. Therefore, a probability of 0.05 or smaller means you can be at least 95 per cent certain that the difference between your two distributions is unlikely to be explained by chance factors alone.

To test whether two groups are different (using numerical data)

t-tests

If a numerical variable can be divided into two distinct groups using a descriptive variable, you can assess the likelihood of these groups being different using an **independent groups t-test** (Box 12.17). This compares the difference in the means of the two groups using a measure of the spread of the scores. If the likelihood of an observed difference or one greater between these two groups occurring by chance alone is low, this is represented by a large t statistic with a low probability (p -value). A p -value of 0.05 or less is usually termed statistically significant.

Alternatively, you might have numerical data for two variables that measure the same feature but under different conditions. Your research could focus on the effects of an

intervention such as employee counselling. Consequently, you would have pairs of data that measure work performance before and after counselling for each case. To assess the likelihood of any difference or one greater between your two variables (each half of the pair) occurring by chance alone, you would use a **paired t-test**. Although the calculation of this is slightly different, your interpretation would be the same as for the independent groups *t*-test.

Both forms of ***t*-test** assumes that the data are normally distributed (Section 12.6 and 12.9), and this can be ignored without too many problems for sufficiently large samples, this often being defined as less than 100 (Lumley et al. 2002) and by some as less than 30 (Hays 1994). The assumption that the data for the two groups have the same variance (standard deviation squared) can also be ignored provided that the two samples are of similar size (Hays 1994).

Mann–Whitney *U* test

If the data are skewed or the sample size is small, the most appropriate statistical test is the **Mann–Whitney *U* Test**. This test is the non-parametric equivalent of the independent groups *t*-test (Dancey and Reidy 2020). Consequently, if the likelihood of a difference or one greater between these two groups occurring by chance alone is low, this will be represented by a large *U* statistic with a probability less than 0.05. This is termed statistically significant.

To test whether three or more groups are different (using numerical data)

One-way analysis of variance (ANOVA)

Where a numerical variable can be divided into three or more distinct groups using a descriptive variable, you can assess the likelihood of these groups being different occurring by chance alone by using **one-way analysis of variance** or one-way **ANOVA** (Box 12.18). As you can gather from its name, ANOVA analyses the **variance**, that is, the spread of data values, within and between groups of data by comparing means. The *F* ratio or *F* statistic represents these differences. If the likelihood of the observed difference or one greater between groups occurring by chance alone is low, this will be represented by a large *F* ratio with a probability of less than 0.05. This is usually considered statistically significant.



Box 12.18 Focus on management research

Testing whether groups are different

A vast body of research supports the benefits of smiling leading to the belief that the larger the smile the better for business. However, there is also evidence that, although broad smiles enhance warmth judgements of the person smiling, they also signal that the smiler is less competent than an individual who is smiling

only slightly. Drawing on this, research by Wang et al. (2017) argues that while a broad as opposed to a slight smile conveys a marketer is friendly and sociable, the broad smile also suggests that a marketer may lack competence. In their paper titled, 'Smile big or not? Effects of smile intensity on perceptions of warmth and competence' in the *Journal of Consumer Research* they expressed this as a hypothesis:

H¹: *Compared to a slight smile, a broad smile will lead to higher perceptions of the marketer's warmth, but lower perceptions of the marketer's competence.* (Wang et al. 2017: 789).





Box 12.18 Focus on management research (continued)

Testing whether groups are different

To test this hypothesis, they selected two images of the same person from a database of digital morphed photographs of facial expressions of different emotions at five different levels of intensity; one of a slight and one of a broad smile. These two photographs were consistent in other appearance cues such as head orientation, brow position and gaze orientation.

Next, they collected data from a sample of 123 adults from Amazon's Mechanical Turk (Mturk) who were each told that the purpose of the research was to examine people's first impressions. Each respondent was shown one of the two photographs and asked to report their warmth and competence perceptions. Warmth was measured using a scale comprising four questions relating to whether the person in the photograph was (i) warm, (ii) kind, (iii) friendly and (iv) sincere. Competence was measured using a scale comprising four questions relating to whether the person in the photograph was (i) competent, (ii) intelligent, (iii) capable and (iv) skilful. All these questions were scored 1 = 'not at all', through to 7 = 'very much so'. To

ensure that the manipulation of the smile had not affected the variables, respondents were also asked questions about the authenticity of the smile and the attractiveness of the person.

Independent sample *t*-tests revealed that the ratings of smile intensity were significantly higher when the person was smiling broadly ($t = 2.60, p = .01$). Ratings of the person's perceived authenticity and attractiveness did not appear to differ significantly between broad and slight smiles, the *t* statistic not being reported in the paper.

Subsequently, Wang and colleagues tested their hypothesis regarding the differential effect of smile intensity on perceptions of warmth and competence by calculating ANOVA (analysis of variance) statistics. This revealed that judgements of warmth were significantly higher for a broad smile than for a slight smile ($F(1,121) = 23.28, p < .001$). However, competence judgements were significantly lower for a broad smile than for a slight smile ($F(1,121) = 6.29, p = .01$). This provided support for their hypothesis arguing that individuals displaying broad smiles tend to be judged as warmer but less competent than those displaying slight smiles.

Subsequent research reported in the same paper investigated the impact on perceptions of smiles of different consumption contexts looking at the marketer's persuasive intent, perceived purchased risk and regulatory frameworks.

The following assumptions need to be met before using one-way ANOVA. More detailed discussion is available in Hays (1994) and Dancey and Reidy (2020).

- Each data value is independent and does not relate to any of the other data values. This means that you should not use one-way ANOVA where data values are related in some way, such as the same case being tested repeatedly.
- The data for each group are normally distributed (Sections 12.6 and 12.9). This assumption is not particularly important provided that the number of cases in each group is large (30 or more).
- The data for each group have the same variance (standard deviation squared). However, provided that the number of cases in the largest group is not more than 1.5 times that of the smallest group, this appears to have very little effect on the test results.

12.11 Assessing the strength of relationships

If your data set contains ranked or numerical data, it is likely that, as part of your exploratory data analysis, you will already have plotted the relationship between cases for these ranked or numerical variables using a scatter graph (Figure 12.12). Such relationships

might include those between weekly sales of a new product and those of a similar established product, or age of employees and their length of service with the company. These examples emphasise the fact that your data can contain two sorts of relationship:

- those where a change in one variable is accompanied by a change in another variable but it is not clear which variable caused the other to change, a **correlation**;
- those where a change in a (dependent) variable is caused by a change in another (independent) variable(s), a **causal relationship**.

To assess the strength of relationship between pairs of variables (using numerical data)

Correlation coefficients

A **correlation coefficient** enables you to quantify the strength of the linear relationship between two ranked or numerical variables. This coefficient (usually represented by the letter r) can take on any value between $+1$ and -1 (Figure 12.15). A value of $+1$ represents a perfect **positive correlation**. This means that the two variables are precisely related and that as values of one variable increase, values of the other variable will increase. By contrast, a value of -1 represents a perfect **negative correlation**. Again, this means that the two variables are precisely related; however, as the values of one variable increase those of the other decrease. Correlation coefficients between $+1$ and -1 represent weaker positive and negative correlations, a value of 0 meaning the variables are perfectly independent. Within business research it is extremely unusual to obtain perfect correlations.

For data collected from a sample, you will need to know the probability of your correlation coefficient or one more extreme (larger) having occurred by chance alone. Most analysis software calculates this probability automatically (Box 12.19). As outlined earlier, if this probability is very low (usually less than 0.05) then the relationship is usually considered statistically significant. In effect you are rejecting the null hypothesis, that is a statement such as: ‘there is no correlation between . . .’ and accepting a hypothesis such as: ‘there is a correlation between . . .’. If the probability is greater than 0.05 then your relationship is usually considered not statistically significant.

Pearson's product moment correlation coefficient (PMCC)

If both your variables contain numerical data, you should use **Pearson's product moment correlation coefficient** (PMCC) to assess the strength of relationship (Box 12.19). Where these data are from a sample then the sample should have been selected at random and the data should be normally distributed. If one or both of your variables contain ranked data you cannot use PMCC, but will need to use a correlation coefficient that is calculated using ranked data.

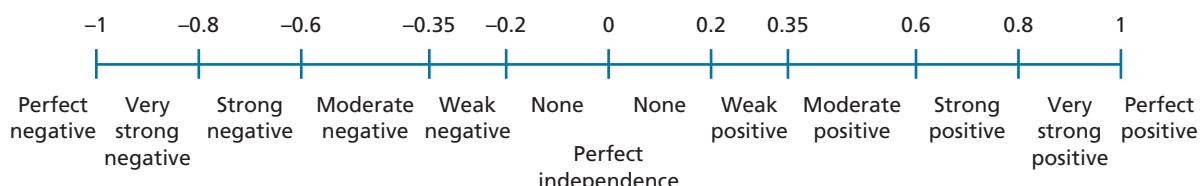


Figure 12.15 Interpreting a correlation coefficient

Source: Developed from earlier editions, Hair et al. (2018)



Box 12.19 Focus on student research

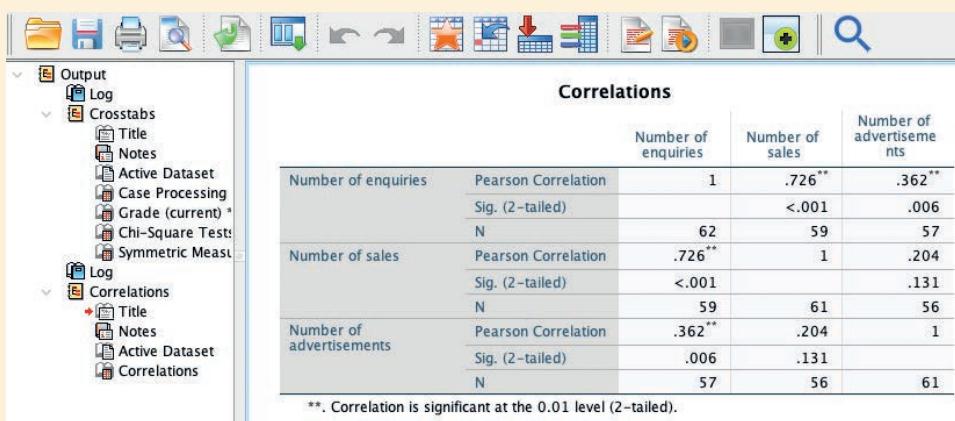
Assessing the strength of relationship between pairs of variables

As part of his research project, Hassan obtained data from a company on the number of television advertisements, number of enquiries and number of sales of their product. These data were entered into the analysis software. He wished to discover whether there

were any relationships between the following pairs of these variables:

- number of television advertisements and number of enquiries;
- number of television advertisements and number of sales;
- number of enquiries and number of sales.

As the data were numerical, he used the statistical analysis software to calculate Pearson's product moment correlation coefficients for all pairs of variables. The output was the correlation matrix:



| | | Number of enquiries | Number of sales | Number of advertisements |
|--------------------------|---------------------|---------------------|-----------------|--------------------------|
| Number of enquiries | Pearson Correlation | 1 | .726** | .362** |
| | Sig. (2-tailed) | | <.001 | .006 |
| | N | 62 | 59 | 57 |
| Number of sales | Pearson Correlation | .726** | 1 | .204 |
| | Sig. (2-tailed) | <.001 | | .131 |
| | N | 59 | 61 | 56 |
| Number of advertisements | Pearson Correlation | .362** | .204 | 1 |
| | Sig. (2-tailed) | .006 | .131 | |
| | N | 57 | 56 | 61 |

**. Correlation is significant at the 0.01 level (2-tailed).

Hassan's matrix is symmetrical because correlation implies only a relationship rather than a cause-and-effect relationship. Cell values in the matrix include the correlation coefficient. Thus, the correlation between the number of advertisements and the number of enquiries is 0.362. This coefficient shows that there is a weak to moderate positive relationship between the number of television advertisements and the number of enquiries. The (**) highlights that the probability of this correlation coefficient or one more extreme occurring by chance alone is less than or equal to 0.01 (1 per cent). This correlation coefficient is therefore usually considered statistically significant. The two-tailed significance for each correlation, rather than a one-tailed significance, is stated as correlation does not

test the direction of a relationship, just whether they are related.

Using the data in this matrix Hassan concluded that:

There is a significant strong positive relationship between the number of enquiries and the number of sales ($r(59) = .726, p < 0.001$) and a significant but weak to moderate relationship between the number of television advertisements and the number of enquiries ($r(57) = .362, p = 0.006$). However, there appears to be no significant relationship between the number of television advertisements and the number of sales ($r(56) = .204, p = 0.131$).

Spearman's and Kendall's rank correlation coefficients

Rank correlation coefficients represent the degree of agreement between the two sets of rankings. Before calculating the rank correlation coefficient, you will need to ensure that the data for both variables are ranked. Where one of the variables is numerical this will necessitate converting these data to ranked data. Subsequently, you have a choice of rank correlation coefficients. The two used most widely in business and management research are **Spearman's rank correlation coefficient** (Spearman's ρ , the Greek letter rho) and **Kendall's rank correlation coefficient** (Kendall's τ , the Greek letter tau). Where data are being used from a sample, both these rank correlation coefficients assume that the sample is selected at random, and the data are ranked (ordinal). Given this, it is not surprising that whenever you can use Spearman's rank correlation coefficient you can also use Kendall's rank correlation coefficient. However, if your data for a variable contain tied ranks, Kendall's rank correlation coefficient is generally considered to be the more appropriate of these coefficients to use. Although each of the correlation coefficients discussed uses a different formula in its calculation, the resulting coefficient is interpreted in the same way as PMCC.

To assess the strength of a cause-and-effect relationship between dependent and independent variables (using numerical data)

Coefficients of determination and multiple determination

The **coefficient of determination** enables you to assess the strength of relationship between a numerical dependent variable and one numerical independent variable; and the **coefficient of multiple determination** enables you to assess the strength of relationship between a numerical dependent variable and two or more independent variables. Once again, where these data have been selected from a sample, the sample must have been selected at random. For a dependent variable and one (or perhaps two) independent variables you will have probably already plotted this relationship on a scatter graph. If you have more than two independent variables this is unlikely as it is very difficult to represent four or more scatter graph axes visually!

The coefficient of determination (represented by r^2) and the coefficient of multiple determination (represented by R^2) can both take on any value between 0 and +1. They measure the proportion of the variation in a dependent variable (amount of sales) that can be explained statistically by the independent variable (marketing expenditure) or variables (marketing expenditure, number of sales staff, etc.). This means that if all the variation in amount of sales can be explained by the marketing expenditure and the number of sales staff, the coefficient of multiple determination will be 1. If 50 per cent of the variation can be explained, the coefficient of multiple determination will be 0.5, and if none of the variation can be explained, the coefficient will be 0 (Box 12.20). Within our research we have rarely obtained a coefficient above 0.8.

For a dependent variable and two or more independent variables you will have probably already plotted this relationship on a scatter graph. The process of calculating the coefficient of determination and regression equation using one independent variable is normally termed **regression analysis**. Calculating a coefficient of multiple determination



Box 12.20 Focus on student research

Assessing a cause-and-effect relationship

As part of her research project, Arethea wanted to assess the relationship between employees' annual salaries and the number of years each had been employed by an organisation. She believed that an employee's annual salary would be dependent

on the number of years for which she or he had been employed (the independent variable). Arethea entered these data into her analysis software and calculated a coefficient of determination (r^2) of 0.37.

As she was using data for all employees of the firm (the total population) rather than a sample, the probability of her coefficient occurring by chance alone was 0. She therefore concluded that 37 per cent of the variation in current employees' salary could be explained by the number of years they had been employed by the organisation.

and regression equation using two or more independent variables is termed **multiple regression analysis**, and we advise you to use statistical analysis software and consult a detailed statistics textbook that also explains how to use the software, such as Field (2018). For sample data most statistical analysis software will automatically calculate the significance of the coefficient of multiple determination or one more extreme occurring by chance. A very low p -value (usually less than 0.05) means that your coefficient or one more extreme is unlikely to have occurred by chance alone.

12.12 Making predictions

To predict the value of a variable from one or more other variables

Calculating the regression equation

Regression analysis can also be used to predict the values of a dependent variable given the values of one or more independent variables by calculating a **regression equation** (Box 12.21). You may wish to predict the amount of sales for a specified marketing expenditure and number of sales staff. You would represent this as a regression equation:

$$AoS_i + \alpha + \beta_1 ME_i + \beta_2 NSS_i$$

where:

- AoS is the amount of sales (the dependent variable).
- ME is the marketing expenditure (an independent or predictor variable).
- NSS is the number of sales staff (an independent or predictor variable).
- α is the regression coefficients.
- β_1 and β_2 are the beta coefficients.



Box 12.21 Focus on student research

Forecasting the number of road injury accidents

As part of her research project, Nimmi had obtained data on the number of road injury accidents and the number of drivers breath tested for alcohol in 39 police force areas. In addition, she obtained data on the total population (in thousands) for each of these areas from the most recent census. Nimmi wished to find out if it was possible to predict the number of road injury

accidents (RIA_i) in each police area (her dependent variable) using the number of drivers breath tested (BT_i) and the total population in thousands (POP_i) for each of the police force areas (independent variables). This she represented as an equation:

$$RIA_i + \alpha + \beta_1 BT_i + \beta_2 POP_i$$

Nimmi entered her data into the analysis software and undertook a multiple regression analysis. She scrolled down the output file and found the table headed 'Coefficients'. Nimmi substituted the 'unstandardised coefficients' into her regression equation (after rounding the values):

$$RIA_i = -30.689 + 0.011 BT_i + 0.127 POP_i$$

| Model | Coefficients ^a | | | |
|-------|--------------------------------------|------------|------|--------|
| | B | Std. Error | Beta | t |
| 1 | (Constant) -30.689 | 11.798 | | -2.601 |
| | Number of breath tests .011 | .005 | .184 | 2.206 |
| | Population of area in thousands .127 | .013 | .803 | 9.632 |

a. Dependent Variable: Number of injury accidents

This meant she could now predict the number of road injury accidents for a police area of different populations for different numbers of drivers breath tested for alcohol. For example, the number of road injury accidents for an area of 500,000 population in which 10,000 drivers were breath tested for alcohol can now be estimated:

$$\begin{aligned} & -30.689 + (0.011 \times 10000) + (0.127 \times 500) \\ & = -30.689 + 110 + 49 + 63.5 \\ & = 81.8 \end{aligned}$$

In order to check the usefulness of these estimates, Nimmi scrolled back up her output and looked at the results of R^2 , t -test and F -test.

The R^2 and adjusted R^2 values of 0.965 and 0.931 respectively both indicated that there was a high degree of goodness of fit of her regression model. It also meant that over 90 per cent of variance in the dependent variable (the number of road

injury accidents) could be explained by the regression model. The F -test result was 241.279 with a significance ('Sig.') of .000. This meant that the probability of these or more extreme results occurring by chance was less than 0.001. This she interpreted as a significant relationship between the number of road injury accidents in an area and the population of the area, and the number of drivers breath tested for alcohol.

The t -test results for the individual regression coefficients (shown in the first extract) for the two independent variables were 9.632 and 2.206. Once again, the probability of both these or more extreme results occurring by chance was less than 0.05, being less than 0.001 for the independent variable population of area in thousands and 0.034 for the independent variable number of breath tests. This means that the regression coefficients for these variables were both considered significant at the $p < 0.05$ level.





Box 12.21 Focus on student research (continued)

Forecasting the number of road injury accidents

The screenshot shows the SPSS interface with the 'Model Summary' and 'ANOVA' tables selected. The 'Model Summary' table displays the following data:

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|-------|-------------------|----------|-------------------|----------------------------|
| 1 | .965 ^a | .931 | .927 | 43.42389 |

a. Predictors: (Constant), Population of area in thousands, Number of breath tests

The 'ANOVA' table displays the following data:

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|-------|------------|----------------|----|-------------|---------|--------------------|
| 1 | Regression | 909927.069 | 2 | 454963.535 | 241.279 | <.001 ^b |
| | Residual | 67882.828 | 36 | 1885.634 | | |
| | Total | 977809.897 | 38 | | | |

a. Dependent Variable: Number of injury accidents
b. Predictors: (Constant), Population of area in thousands, Number of breath tests

This equation can be translated as stating:

$$\text{Amount of sales}_i = \text{value} + (\beta_1 * \text{Marketing expenditure}_i) + (\beta_2 * \text{Number of sales staff}_i)$$

Using regression analysis, you would calculate the values of the constant coefficient a and the slope coefficients β_1 and β_2 from data you had already collected on amount of sales, marketing expenditure and number of sales staff. A specified marketing expenditure and number of sales staff could then be substituted into the regression equation to predict the amount of sales that would be generated. When calculating a regression equation, you need to ensure the following assumptions are met:

- the relationship between dependent and independent variables is linear. **Linearity** refers to the degree to which the change in the dependent variable is related to the change in the independent variables. Linearity can easily be examined through residual plots (these are usually drawn by the analysis software). Two things may influence the linearity. First, individual cases with extreme values on one or more variables (outliers) may violate the assumption of linearity. It is, therefore, important to identify these outliers and, if appropriate, exclude them from the analysis. Second, the values for one or more variables may violate the assumption of linearity. For these variables the data values may need to be transformed. Techniques for this can be found in other, more specialised books on multivariate data analysis, for example Hair et al. (2018).
- the extent to which the data values for the dependent and independent variables have equal variances (this term was explained earlier in Section 12.4), also known as **homoscedasticity**. Again, analysis software usually contains statistical tests for equal variance. For example, the Levene test for homogeneity of variance measures the equality of variances for a single pair of variables. If **heteroscedasticity** (that is, unequal variances)

- exists, it may still be possible to carry out your analysis. Further details of this can again be found in more specialised books on multivariate analysis, such as Hair et al. (2018).
- absence of correlation between two or more independent variables (**collinearity** or **multicollinearity**), as this makes it difficult to determine the separate effects of individual variables. The simplest diagnostic is to use the correlation coefficients, extreme collinearity being represented by a correlation coefficient of 1. The rule of thumb is that the presence of high correlations (generally 0.90 and above) indicates substantial collinearity (Hair et al. 2018). Other common measures include the tolerance value and its inverse – the **variance inflation factor** (VIF). Hair et al. (2018) recommend a very small tolerance value (0.10 or below) or a large VIF value (10 or above) indicates high collinearity.
 - the errors for the independent variables are normally distributed (Sections 12.6 and 12.9). The **residuals** are the ‘errors’ (differences) between each predicted value for the dependent variable when using the regression equation and the associated observed value for the dependent variable. A simple diagnostic tool is to either draw a histogram or frequency polygon of the residual values and look for a normal distribution (Figures 12.7 and 12.17). Alternatively, you can use a scatter graph (Figure 12.12) to plot the residual values against the predicted values and look for a diagonal line running from bottom left to top right.
 - if your data are a sample, rather than a population, you also need to estimate the number of cases required in your sample. For regression analysis a widely used formula to estimate the number needed to satisfy the analysis’ assumptions is:

$$\text{Sample size} = 50 + (8 \times \text{number of independent (predictor) variables})$$

Consequently, for a regression analysis with two independent variables the sample size can be estimated as:

$$\begin{aligned}\text{sample size} &= 50 + (8 \times 2) \\ &= 50 + 16 = 66\end{aligned}$$

However, this is an approximation and will overestimate the sample size required as the number of independent variables increases (Green 1991).

To measure the goodness of fit of the regression equation

Coefficients of determination and multiple determination

The coefficient of determination, r^2 (discussed earlier), can be used as a measure of how good a predictor your regression equation is likely to be. If your equation is a perfect predictor then the coefficient of determination will be 1. If the equation can predict only 50 per cent of the variation, then the coefficient of determination will be 0.5, and if the equation predicts none of the variation, the coefficient will be 0. The coefficient of multiple determination (R^2) indicates the degree of the goodness of fit for your estimated multiple regression equation. It can be interpreted as how good a predictor your multiple regression equation is likely to be. It represents the proportion of the variability in the dependent variable that can be explained by your multiple regression equation. This means that when multiplied by 100, the coefficient of multiple determination can be interpreted as the percentage of variation in the dependent variable that can be explained by the estimated regression equation. The adjusted R^2 statistic (which takes into account the number of independent variables in your regression equation) is preferred by some researchers as it helps avoid overestimating the impact of adding an independent variable on the amount of variability explained by the estimated regression equation.

t-tests and *F*-tests

The *t*-test and *F*-test are used to work out the probability of the relationship represented by your regression analysis or one more extreme having occurred by chance. In simple linear regression (with one independent and one dependent variable), the *t*-test and *F*-test will give you the same answer. However, in multiple regression, the *t*-test is used to find out the probability of the relationship between each of the individual independent variables and the dependent variable or one more extreme occurring by chance. In contrast, the *F*-test is used to find out the overall probability of the relationship or one more extreme between the dependent variable and all the independent variables occurring by chance. The *t* distribution table and the *F* distribution table are used to determine whether a *t*-test or an *F*-test is significant by comparing the results with the *t* distribution and *F* distribution respectively, given the degrees of freedom and the predefined significance level.

12.13 Examining trends

When examining longitudinal data, the first thing we recommend you do is to draw a line graph to obtain a visual representation of the trend (Figure 12.6). Subsequently, statistical analyses can be undertaken. Three of the more common uses of such analyses are:

- to explore the trend or relative change for a single variable over time;
- to compare trends or the relative change for variables measured in different units or of different magnitudes;
- to determine the long-term trend and forecast future values for a variable (time series analysis).

To establish the trend

Index numbers

To answer some research question(s) and meet some objectives you may need to establish the trend for one variable. One way of doing this is to use **index numbers** to compare the relative magnitude for each data value (case) over time rather than using the actual data value. Index numbers are also widely used in business publications and by organisations. Various share indices (Box 12.22), such as the *Financial Times* FTSE 100, and the Nasdaq Composite Index are well-known examples.

Although such indices can involve quite complex calculations, they all compare change over time against a base period. The **base period** is normally given the value of 100 (or 1000 in the case of many share indices, including the FTSE 100) and change is calculated relative to this. Thus, a value greater than 100 would represent an increase relative to the base period, and a value less than 100 a decrease.

To calculate simple index numbers for each case of a longitudinal variable you use the following formula:

$$\text{Index number for case} = \frac{\text{date value for case}}{\text{base period data value}} \times 100$$

Thus, if a company's sales were 125,000 units in 2021 (base period) and 150,000 units in 2022, the index number for 2021 would be 100 and for 2022 it would be 120.



Box 12.22 Focus on research in the news

Stocks dip as traders weigh Covid curbs and monetary policy direction

Nasdaq drops 1.7 per cent while Tesla, Peloton and BuzzFeed fall sharply

Global stocks dipped on Thursday, as traders weighed new restrictions aimed at tackling the spread of the Omicron coronavirus variant along with questions about the direction of monetary policy.

The technology focused Nasdaq Composite index closed 1.7 per cent lower, hitting a session low just before the bell. Shares in exercise bike company Peloton fell 11.4 per cent, while Tesla ceded 6.1 per cent. Media company BuzzFeed, which went public earlier this week, gave up 23.6 per cent.

Wall Street's blue-chip S&P 500 index slipped 0.7 per cent, also hitting session lows near the close. The equity gauge had ended the previous session within reach of its all-time closing high, wiping out almost all of its losses sustained in volatile trading since the emergence of Omicron rattled markets in late November.

European equities ended the day lower. The regional Stoxx 600 index dipped by just under 0.1 per cent, while London's FTSE 100 dropped 0.2 per cent.

The UK moved to implement its plan B restrictions on Wednesday evening, including guidance to work from home and mandatory mask-wearing for most indoor venues. Denmark tightened its virus control measures, following similar moves by other EU nations including Germany, Italy and Poland.

Meanwhile, economists surveyed by Refinitiv expect data to be released on Friday to show US consumer prices rose 6.8 per cent in the year to November. 'There was a post-Omicron rally but I think it is still too soon to interpret the recent data,' said Lale Akoner, senior market strategist at BNY Mellon. 'If there is a big inflation uptick, then I do believe the market is going to get nervous about the Fed hiking [interest rates] earlier and faster than has been priced in so far.'



Source: Abridged from: 'Stocks dip as traders weigh Covid curbs and monetary policy direction', Naomi Rovnick and Kate Duguid (2021) *Financial Times*, 9 December. Copyright © The Financial Times Ltd

To compare trends between two or more variables

Index numbers

To answer some other research question(s) and to meet the associated objectives you may need to compare trends between two or more variables measured in different units or at different magnitudes. For example, to compare changes in prices of fossil fuels such as oil and coal over time is difficult as the prices are recorded for different units (litres and tonnes). One way of overcoming this is to use index numbers) and compare the relative changes in the value of the index rather than actual figures. The index numbers for each variable are calculated in the same way as outlined earlier.

To determine the trend and forecast

Moving averages

The trend can be estimated by drawing a freehand line through the data on a line graph. However, these data are often subject to variations such as seasonal fluctuations, and so this method is not very accurate. A straightforward way of overcoming this is to calculate a moving average for the time series of data values. Calculating a **moving average** involves replacing each value in the time series with the mean of that value and those values directly preceding and following it (Anderson et al. 2020). This smooths out the variation in the data so that you can see the trend more clearly. The calculation of a moving average is relatively straightforward using either a spreadsheet or statistical analysis software.

Once the trend has been established, it is possible to forecast future values by continuing the trend forward for time periods for which data have not been collected. This involves calculating the **long-term trend** – that is, the amount by which values are changing in each time period after variations have been smoothed out. Once again, this is relatively straightforward to calculate using analysis software.

Forecasting using regression analysis

Forecasting can also be undertaken using other statistical methods, including regression analysis. If you are using regression for your time-series analysis, the **Durbin–Watson statistic** can be used to discover whether the value of your dependent variable at time t is related to its value at the previous time period, commonly referred to as $t - 1$. This situation, known as **autocorrelation** or **serial correlation**, is important as it means that the results of your regression analysis are less likely to be reliable. The Durbin–Watson statistic ranges in value from zero to 4. A value of 2 indicates no autocorrelation. A value towards zero indicates positive autocorrelation. Conversely, a value towards 4 indicates negative autocorrelation. More detailed discussion of the Durbin–Watson test can be found in other, more specialised books on multivariate data analysis, for example Hair et al. (2018).

12.14 Summary

- For data to be analysed quantitatively it must either already be quantified or able to be transformed into quantitative data.
- Non-numerical data such as text, voice and visual data can be quantified by classifying into sets or categories.
- Data for quantitative analysis comprise categorical and numerical data.
- Categorical data are either descriptive (dichotomous, or nominal) or ordinal (rank).
- Numerical data can be divided into either continuous and discrete data, or interval or ratio data.
- The data type, and associated precision of measurement, will constrain the tables, graphs, and statistical analysis techniques you can use.
- All data should, with few exceptions, be recorded using numerical codes to facilitate quantitative analyses.
- Where possible, you should use existing coding schemes to enable comparisons.
- For primary data you should include pre-set codes on the data collection form to minimise coding after collection. For variables where responses are not known, you will need to develop a codebook after data have been collected for the first 50 to 100 cases.

- You should enter codes for all data values, including missing data.
- Data are prepared for analysis as a data matrix in which each column usually represents a variable and each row a case. Your first variable should be a unique identifier to facilitate error checking.
- Your initial analysis should explore data using both tables and graphs. Your choice of table or graph will be influenced by your research question(s) and objective(s), the aspects of the data you wish to emphasise, and the data type.
- This may involve using:
 - tables to show specific amounts;
 - bar graphs, multiple bar graphs, histograms and, occasionally, pictograms and word clouds to show (and compare) highest and lowest amounts and relative distributions;
 - line graphs to show trends;
 - pie charts and percentage component bar graphs to show proportions or percentages;
 - box plots to show distributions;
 - multiple line graphs to compare trends and show intersections;
 - scatter graphs to show relationships between variables.
- Your choice of statistics will be influenced by your research question(s) and objective(s), statistical assumptions such as suitability for the data type, whether the data are normally distributed and sample size.
- Your analysis will use statistics to describe, examine associations and differences, examine relationships, make predictions, and examine and determine trends:
 - the mean, median or mode to describe the central tendency;
 - the inter-quartile range or the standard deviation to describe the dispersion;
 - Kolmogorov–Smirnov or Shapiro–Wilk to test for normality;
 - chi square to test whether two variables are independent;
 - Cramer's V and phi to test whether two variables are associated;
 - Kolmogorov–Smirnov to test whether the values differ from a specified population;
 - t-tests and ANOVA to test whether groups are different;
 - correlation, and coefficients of determination and multiple determination (regression), to assess the strength of relationships between variables;
 - regression analysis to predict values.
- Longitudinal data may necessitate selecting different statistical techniques such as:
 - index numbers to establish a trend or to compare trends between two or more variables measured in different units or at different magnitudes;
 - moving averages and regression analysis to determine the trend and forecast.

Self-check questions

Help with these questions is available at the end of the chapter.

- 12.1** The following secondary data have been obtained from the Park Trading Company's audited annual accounts:

| Year end | Income | Expenditure |
|----------|----------|-------------|
| 2013 | 11000000 | 9500000 |
| 2014 | 15200000 | 12900000 |
| 2015 | 17050000 | 14000000 |

| Year end | Income | Expenditure |
|----------|----------|-------------|
| 2016 | 17900000 | 14900000 |
| 2017 | 19000000 | 16100000 |
| 2018 | 18700000 | 17200000 |
| 2019 | 17100000 | 18100000 |
| 2020 | 17700000 | 19500000 |
| 2021 | 19900000 | 20000000 |

- a Which are the variables, and which are the cases?
- b Sketch a possible data matrix for these data for entering into a spreadsheet.

12.2 a How many variables will be generated from the following request?

Please list up to three things you like about your current role

Click to write 1

Click to write 2

Click to write 3

Source: Copyright © 2021 Qualtrics LLC. Used With Permission.

- b How would you go about devising a coding scheme for this question from a Web questionnaire that has been returned by over 600 employees?

12.3 a Illustrate the data from the Park Trading Company's audited annual accounts (Question 12.1) to show trends in income and expenditure.

- b What does your diagram emphasise?
- c What diagram would you use to emphasise the years with the lowest and highest income?

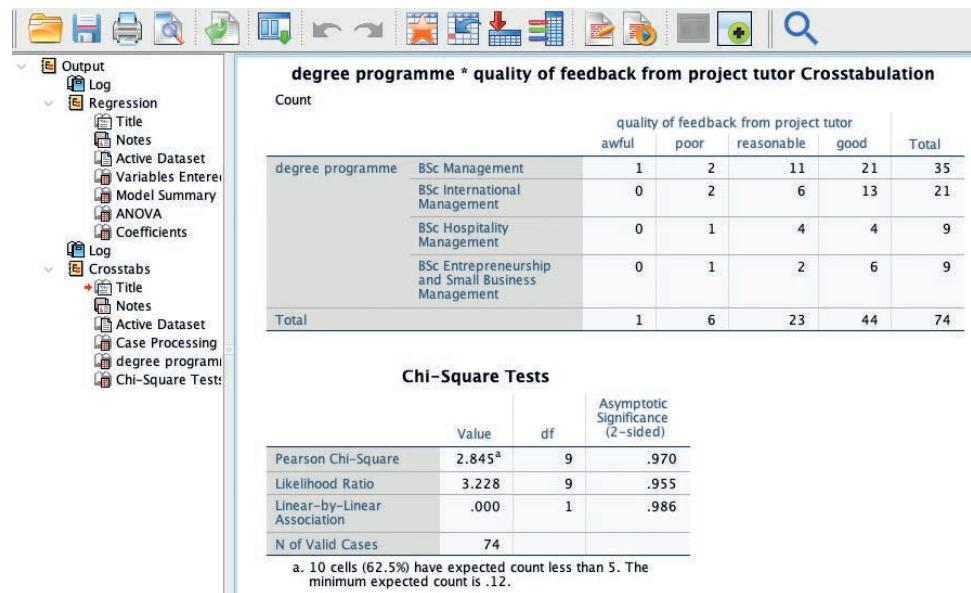
12.4 As part of research into the impact of television advertising on donations by text message to a major disaster appeal, data have been collected on the number of viewers reached by the advertising campaign and the number of donations by text message each day for the past two weeks.

- a Which diagram or diagrams would you use to explore these data?
- b Give reasons for your choice.

12.5 a Which measures of central tendency and dispersion would you choose to describe the Park Trading Company's income (Question 12.1) over the period 2013–2021?

- b Give reasons for your choice.

12.6 a A colleague has collected data from a sample of 74 students. They present you with the following output from the statistical analysis software:
Explain what this tells you about students' opinions about feedback from their project tutor.



12.7 Briefly describe when you would use regression analysis and correlation analysis, using examples to illustrate your answer.

12.8 a Use an appropriate technique to compare the following data on share prices for two financial service companies over the past six months, using the period six months ago as the base period:

| | EJ Investment Holdings | AE Financial Services |
|--------------------|------------------------|-----------------------|
| Price 6 months ago | €10 | €587 |
| Price 4 months ago | €12 | €613 |
| Price 2 months ago | €13 | €658 |
| Current price | €14 | €690 |

b Which company's share prices have increased most in the last six months? (Note: you should quote relevant statistics to justify your answer.)

Review and discussion questions

12.9 Use a search engine to discover coding schemes that already exist in secondary data you are interested in using. This might be ethnic group, family expenditure, industry group, socio-economic class and the like.

a Discuss how credible you think each coding scheme is with a friend. To come to an agreed answer, pay particular attention to:

- the organisation (or person) that is responsible for the coding scheme;
- any explanations regarding the coding scheme's design;
- use of the coding scheme to date.

b Widen your search to include coding schemes that may be of use for your research project. Make a note of the web address of any that are of interest.

- 12.10** With a friend, choose a large company in which you are interested. Obtain a copy of the annual report for this company. Examine the use of tables, graphs and charts in your chosen company's report.
- To what extent does the use of graphs and charts in your chosen report follow the guidance summarised in Box 12.8 and Table 12.2?
 - Why do you think this is?
- 12.11** With a group of friends, each choose a different share price index. Well-known indices you might choose include the Nasdaq Composite Index, France's CAC 40, Germany's Dax, Hong Kong's Hang Seng Index (HSI), Japan's Nikkei Index, the UK's FTSE 100 and the USA's Dow Jones Industrial Average Index.
- For each of the indices, find out how it is calculated and note down its daily values for a one-week period.
 - Compare your findings regarding the calculation of your chosen index with those for the indices chosen by your friends, noting down similarities and differences.
 - To what extent do the indices differ in the changes in share prices they show? Why do you think this is?
- 12.12** Find out whether your university provides you with access to IBM SPSS Statistics. If it does, visit this book's companion website and download the self-teach package and associated data sets. Work through this to explore the features of IBM SPSS Statistics.



Progressing your research project

Analysing your data quantitatively

- Examine the technique(s) you are proposing to use to obtain data to answer your research question. You need to decide whether you will be using any data that could usefully be analysed quantitatively.
- If you decide that your data should be analysed quantitatively, you must ensure that the data collection methods have been designed to make analysis as straightforward as possible. In particular, you need to pay attention to the coding scheme for each variable and the layout of your data matrix.
- Once your data have been prepared for your analysis software, you will need to explore and

present them. Bearing your research question in mind, you should select the most appropriate diagrams and tables after considering the suitability of all possible techniques. Remember to label your diagrams clearly and to keep a copy, as they may form part of your research report.

- Once you are familiar with your data, use appropriate statistics to answer your research questions. This may include one or more of describing data, examining associations and differences, examining relationships, making predictions and examining trends.
- Remember to keep an annotated copy of your analyses, as you will need to quote statistics to justify statements you make in the findings section of your research report.
- Use the questions in Box 1.4 to guide you in your reflective diary entry.

References

Anderson, D.R., Sweeney, D.J., Williams, T.A., Comm, J. Cochran, J., Freeman, J. and Shoesmith E. (2020) *Statistics for Business and Economics* (5th edn). Andover: Cengage Learning.

Anderson, T.W. (2003) *An Introduction to Multivariate Statistical Analysis* (3rd edn). New York: John Wiley.

- Berelson, B. (1952) *Content Analysis in Communication Research*. Glencoe, IL: Free Press.
- Berman Brown, R. and Saunders, M. (2008) *Dealing with Statistics: What You Need to Know*. Maidenhead: McGraw-Hill Open University Press.
- Black, K. (2020) *Business Statistics for Contemporary Decision Making* (10th edn). Hoboken, NJ: Wiley.
- Blumberg, B., Cooper, D.R. and Schindler, D.S. (2014) *Business Research Methods* (4th edn). Maidenhead: McGraw-Hill.
- Christ, K.L., Rao, K.K. and Burritt, R.L. (2019) 'Accounting for modern slavery: an analysis of Australian listed company disclosures', *Accounting, Auditing and Accountability Journal*, Vol. 32, No. 3, pp. 836–65.
- Corder, G.W. and Foreman, D.I. (2014) *Nonparametric Statistics* (2nd edn). Hoboken, NJ: Wiley.
- Dancey, C.P. and Reidy, J. (2020) *Statistics Without Maths for Psychology*: (8th edn). Harlow: Pearson.
- Dawson, J. (2017) *Analysing Quantitative Survey Data for Business and Management Students*. London: Sage.
- Ellis, P.D. (2010) *The Essential Guide to Effect Sizes*. Cambridge: Cambridge University Press.
- Eurostat (2021) Environment and energy statistics – primary production of renewable energies. Available at <https://ec.europa.eu/eurostat/web/products-datasets/-/ten00081> [Accessed 17 November 2017].
- Field, A. (2018) *Discovering Statistics Using SPSS* (5th edn). London: Sage.
- Green, S.B. (1991) 'How many subjects does it take to do a regression analysis?', *Multivariate Behavioural Research*, Vol. 26, No. 3, pp. 499–510.
- Hair, J.F., Black, B., Babin, B., Anderson, R.E. and Tatham, R.L. (2018) *Multivariate Data Analysis* (8th edn). Andover: Cengage.
- Harley-Davidson Inc. (2021) *Harley-Davidson Inc. Investor Relations: Motorcycle Shipments* Available at <http://investor.harley-davidson.com/phoenix.zhtml%3Fc%3D87981%26p%3Dirol-shipments> [Accessed 12 October 2021].
- Hays, W.L. (1994) *Statistics* (4th edn). London: Holt-Saunders.
- Holsti, O.R. (1969) *Content Analysis for the Social Sciences and Humanities*. Reading, MA: Addison-Wesley.
- Keen, K.J. (2018) *Graphics for Statistics and Data Analysis with R*. (2nd edn) Boca Raton, FL: Chapman and Hall.
- Kosslyn, S.M. (2006) *Graph Design for the Eye and Mind*. New York: Oxford University Press.
- Krippendorff, K. (2018) *Content Analysis: An Introduction to Its Methodology*. Los Angeles, CA: Sage.
- Little, R. and Rubin, D. (2019) *Statistical Analysis with Missing Data* (3rd edn). New York: John Wiley.
- Lumley, T., Diehr, P., Emerson, S. and Chen, L. (2002) 'The importance of the normality assumption in large public health data sets', *Annual Review of Public Health*, Vol. 23, pp. 151–69.
- McHugh, M.L. (2012) 'Interrater reliability: the Kappa statistic', *Biochimia Medica*, Vol. 22, No. 3, pp. 276–182.
- Mirbabaie, M. and Marx, J. (2020). 'Breakingnews: Uncovering sense-breaking patterns in social media crisis communication during the 2017 Manchester bombing', *Behaviour & Information Technology*, Vol. 39, No. 3, pp. 252–66.
- National Centre for Social Research (2021) *British Social Attitudes 38*. London: NatCen Social Research. Available at <https://www.bsa.natcen.ac.uk/latest-report/british-social-attitudes-38/introduction.aspx> [Accessed 4 December 2021].
- Office for National Statistics (2021) *SOC2021* Available at <https://www.ons.gov.uk/methodology/classificationsandstandards/standardoccupationalclassificationssoc/soc2020> [Accessed 27 November 2021].
- Office for National Statistics (nd a) *The National Statistics Socio-economic Classification (NS-SEC)*. Available at <https://www.ons.gov.uk/methodology/classificationsandstandards/otherclassifications/thenationalstatisticssocioeconomicclassificationsscrebasedonsoc2010> [Accessed 4 December 2021].

- Office for National Statistics (nd b) *Ethnic Group, National Identity and Religion*. Available at <https://www.ons.gov.uk/methodology/classificationsandstandards/measuringequality/ethnicgroupnationalityandreligion> [Accessed 4 December 2021].
- Pallant, J. (2020) *SPSS Survival Manual: A Step-by-Step Guide to Data Analysis Using IBM SPSS* (7th edn). Maidenhead: Open University Press.
- Prosser, L. (2009) *Office for National Statistics UK Standard Industrial Classification of Activities 2007 (SIC 2007)*. Basingstoke: Palgrave Macmillan. Available at <http://www.ons.gov.uk/guide-method/classifications/current-standard-classifications/standard-industrial-classification/index.html> [Accessed 27 November 2014].
- Qualtrics (2021). *Qualtrics Online Survey Software*. Available at <https://www.qualtrics.com/uk/core-xm/survey-software/> [Accessed 27 Nov. 2021].
- Rose, S., Spinks, N. and Canhoto, I. (2015) *Management Research: Applying the Principles*. London: Routledge.
- SciStatCalc (2013) *Two-sample Kolmogorov-Smirnov Test Calculator*. Available at <http://scistatcalc.blogspot.co.uk/2013/11/kolmogorov-smirnov-test-calculator.html> [Accessed 14 December 2021].
- SurveyMonkey (2021) *SurveyMonkey*. Available at www.surveymonkey.com [Accessed 27 Nov. 2021].
- The Economist (2021) *The rEV Index*. Available at <https://impact.economist.com/sustainability/project-the-rev-index/rev-index/> [Accessed 1 Dec. 2021].
- Tukey, J.W. (2020) *Exploratory Data Analysis*. Hoboken, NJ: Pearson (originally published in 1977).
- Wang, Z., Mao, H., Li, Y.J. and Liu, F. (2017) 'Smile big or not? Effects of smile intensity on perceptions of warmth and competence', *Journal of Consumer Research*, Vol. 43, pp.787–805.
- Wasserstein, R.L. and Lazar, N.A. (2016) 'The ASA's statement on p-values: context, process and purpose', *The American Statistician*, Vol. 70, No. 2, pp. 129–33.
- Winston, W. (2022) *Microsoft Excel Data Analysis and Business Modelling* (7th edn) Redmond, WA: Microsoft Press.

Further reading

- Berman Brown, R. and Saunders, M. (2008) *Dealing with Statistics: What You Need to Know*. Maidenhead: McGraw Hill Open University Press. This is a statistics book that assumes virtually no statistical knowledge, focusing upon which test or graph, when to use it and why. It is written for people who are fearful and anxious about statistics and do not think they can understand numbers!
- De Vaus, D.A. (2014) *Surveys in Social Research* (6th edn). Abingdon: Routledge. Chapters 9 and 10 contain an excellent discussion about coding data and preparing data for analysis. Part IV (Chapters 12–18) provides a detailed discussion of how to analyse survey data.
- Field, A. (2018) *Discovering Statistics Using SPSS* (5th edn). London: Sage. This book offers a clearly explained guide to statistics and using SPSS. It is divided into four levels, the lowest of which assumes no familiarity with the data analysis software and very little with statistics. It covers entering data and how to generate and interpret a wide range of tables, diagrams and statistics using SPSS.
- Hair, J.F., Black, B., Babin, B., Anderson, R.E. and Tatham, R.L. (2018) *Multivariate Data Analysis* (8th edn). Andover: Cengage. This book provides detailed information on statistical concepts and techniques. Issues pertinent to design, assumptions, estimation and interpretation are systematically explained for users of more advanced statistical techniques.
- McCandless, D. (2021) *Beautiful News. Positive Trends, Uplifting Stats, Creative Solutions*. London: William Collins. This book of infographics focuses on positive things happening in the world and is uplifting! Like the author's books *Knowledge is Beautiful* (2014) and *Information is Beautiful* (2012), it is best considered as a visual miscellaneum of facts and ideas to explore.

Collecting Primary Data Using Interviews & Diaries

Saunders et al. (2023; 440-500)

Chapter 10



Collecting primary data using interviews and diaries

Learning outcomes

By the end of this chapter, you should be able to:

- classify interview types in terms of their structure, mode and medium, and understand their purposes;
- discuss the logistical and resource issues that affect the use of different types of interviews;
- identify potential data quality issues related to the use of one-to-one mode interviews using online, telephone and face-to-face media, and evaluate how to overcome these;
- develop your competence to undertake different types of one-to-one mode interviews using online, telephone and face-to-face media;
- highlight the advantages and disadvantages of different types of one-to-one mode interviews using online, telephone and face-to-face interview media;
- highlight the advantages and disadvantages of one-to-many and two-to-many mode group and focus group interviews using both online and face-to-face interview media;
- discuss the nature of visual interviews and ways to incorporate these using online and face-to-face interview media;
- explain how qualitative and quantitative research diaries can be used to collect primary data;
- discuss the advantages and issues of using research diaries;
- progress your research project by collecting primary data using research interviews and research diaries.

10.1 Introduction

The **research interview** is a purposeful conversation between two or more people, during which the interviewer asks concise and unambiguous questions and listens attentively to the interviewee's responses. Usually just referred to as interviews, they rely on establishing rapport between the interviewer and the interviewee, and can help you to gather valid and reliable



data to help answer your research question(s) and meet your objectives. Interviews can also be used to help refine your ideas where you have not yet fully formulated a research question or objectives.

In Section 4.2 we considered how our philosophical assumptions inform our views about the nature of reality, highlighting the diversity of research philosophies in business and management research. This also applies to interview-based research (Reisner and Whittle 2021),

Journalists' interviewing skills

Interviews occur constantly. Every day there is scope to watch and listen to interviews, or to read about them. Interviews of all sorts occur, such as those related to business, jobs, celebrities, the arts, current events, and news stories. Some are face-to-face while others are conducted virtually by telephone or online, using video telephony, video conferencing or online chat. Interviews may be accessed online, in newspapers and on television and radio programmes. Every time an event happens, those who witness it, those who are involved in it and those who have some expertise associated with it will be interviewed. However, despite the seeming ease with which interviews are undertaken, their conduct requires considerable skill.

One profession that relies on good-quality interviewing skills is journalism. The BBC Academy (n.d.) outlines key interviewing skills on its website. Interviewers need to think clearly about the purpose of each interview and to be aware that their first question will set the direction of an interview and establish its style. Think of a 'confrontational' interviewer you have seen or heard and contrast that with the style of a 'friendly, inviting' interviewer! Interviewers also need to be clear in the way they ask questions and not to be obscure or to



LightField Studios/Shutterstock

use jargon. This means finding ways to ask questions about complex issues that are simple and direct. Open questions invite interviewees to describe or explain, or to develop a previous answer. Closed questions seek straightforward answers, like 'yes' or 'no'. In journalism, this type of question can be used to get to the heart of a particular matter and for this reason it is often called the 'killer' question. Where an interviewee wants to avoid directly answering such a question, its use will expose this reluctance to give a straightforward answer. The use of a 'killer' question is unlikely to be appropriate in business and management research interviewing, but the skills outlined on the BBC Academy (n.d.) downloadable guides and associated videos are likely to be helpful to business and management researchers.

philosophical assumptions underlying differing views regarding the use of interviews. If you adopt a critical realist philosophical position you are more likely to emphasise an objective view, your interviews being seen as a method to collect data from interviewees who are witnesses to a reality that exists independently from them. Within these assumptions, the research interview is seen as unproblematic, effective and reasonably structured means to gather data; providing you can gain access to appropriate participants. If you adopt an interpretivist (or social constructionist) philosophical position, you are more likely to emphasise your interviewees' views and cultures; considering them as social actors who interact with, interpret, and create their social world as well as being shaped by it. Your interviews are likely to be less structured, the data being socially constructed and co-produced by the views and interpretations of your participants and you as interviewer. You will be asking questions, responding to the participant's views, and interpreting the resulting data (Denzin 2001; Heyl 2005): having a central role in constructing meanings and being reflexive. These two views of the interview, drawing from their associated research philosophies, emphasise differences in purpose, structure and standardisation, the medium through which the interviews are operationalised, and the mode in which they are undertaken. They emphasise the need for you, as an interviewer, to reflect on and evaluate your philosophical assumptions and their influence on your approach to interviewing.

Interview is a general term for a variety of types. This variation is important since, the standardisation of questioning and amount of structure (Section 10.2), the medium through which it is operationalised and the mode in which it is undertaken (Section 10.3) need to be compatible with your research question(s) and objectives. Our main focus in this chapter is semi-structured and in-depth interviews; structured interviews (based on the use of researcher- or interviewer-completed questionnaires) being discussed in Chapter 11. Section 10.4 outlines the potential for using semi-structured and in-depth interviews. Section 10.5 identifies data quality issues associated with their use and discusses how to evaluate your own interviewing practice. Section 10.6 discusses preparing for semi-structured and in-depth interviews and the development of interview themes. The conduct of one-to-one face to face interviews, including phrasing questions, is considered in Section 10.7. Recognising that much of the associated advice can also be applied to other interview media, the conduct of online interviews and telephone interviews are considered in Sections 10.8 and 10.9 respectively. Section 10.10 considers group interviews and a distinctive form of these, focus groups; whereas Section 10.11 considers another distinctive form of interview, the visual interview.

The chapter also considers the use of both quantitative and qualitative (research) diaries alongside interviews as well as to collect primary data (Section 10.12).

10.2 Standardisation and structure in questioning

One commonly used typology differentiates between standardised interviews and non-standardised interviews. Another commonly used typology differentiates between structured interviews, semi-structured interviews and unstructured interviews. These typologies overlap: standardised interviews and structured interviews refer to the same type; while non-standardised interviews may be divided into semi-structured and unstructured interviews (Figure 10.1). We use this three-fold typology of structured, semi-structured and unstructured to describe the nature of, and differences in questioning in research

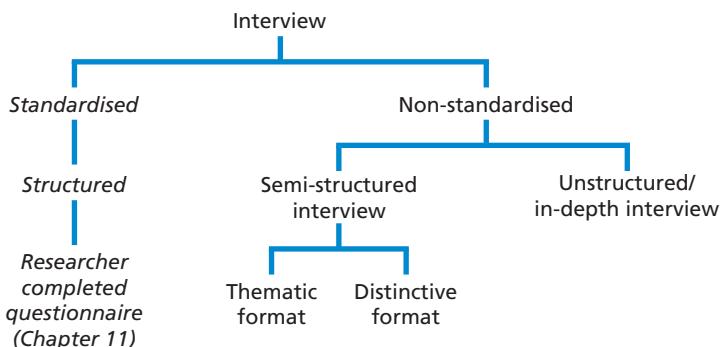


Figure 10.1 Interview structure

interviews, recognising that they refer to distinct but not entirely separate structures, within which there is scope for variation.

Structured interviews

Structured interviews are conducted using researcher-completed questionnaires. As we discuss in Section 11.2, such interviews are ‘standardised’, using a predetermined set of identical questions to collect comparable data from each participant .If you use a structured interview, you ask each participant the questions from the questionnaire or interview schedule exactly as written and in the same tone of voice so that you do not suggest any bias. Responses to each question are recorded on a standardised schedule, usually with pre-coded answers (Sections 11.5 and 12.3). The data are collected either for information, such as collecting attitudes and opinions (Section 11.5), or hypothesis testing (Section 12.9). As structured interviews are used to collect quantifiable data they are also referred to as ‘quantitative research interviews’. In comparison, semi-structured and unstructured interviews are ‘non-standardised’. These are often referred to as ‘qualitative research interviews’.

Semi-structured interviews

Semi-structured interviews vary considerably in the degree of structure, drawing on one or more themes. Some have pre-set questions that the interview follows, whereas others are more open, the interviewer varying the order in which questions are asked and asking new questions dependent upon the research situation. Cassell (2015) distinguishes between two types:

- those that follow a thematic format in their questioning;
- those that have a different but distinctive format.

Thematic format semi-structured interviews

Thematic format semi-structured interviews collect data to answer your research question or address a research topic using a list of pre-determined themes and initial questions to guide the conduct of each interview. The interviews can be exploratory in nature,

collecting data about a particular topic. Alternatively, they can be explanatory in nature, the questioning being more structured around different theoretical aspects to support theory development.

Distinctive format semi-structured interviews

Distinctive format semi-structured interviews comprise a range of different structure formats. For each, the distinctive format usually forms a part of your interview, for example asking each participant to focus on a particular event such as a critical incident or encouraging participants to tell their stories. Other questions can also be asked to establish relevant contextual information. Each Distinctive format can be used for both exploratory and explanatory purposes. It may be used to generate stories about the participant's experiences as in a **narrative interview**, to gain insights through a chronological reflection on their life experiences as in a **biographical interview**, to focus on critical incidents as in a **critical incident** interview, to focus on a group of participants' perceptions or experiences as in a **focus group** or use visual images as in **visual interviews**.

Whatever the format, how you use your questioning will depend on your philosophical assumptions and your research question and objectives. If you adopt a critical realist position, you will believe that there is a truth waiting to be discovered that is external to the interpretations of your participants (Sections 4.2 and 4.4). Here you will use a more consistent approach in which you systematically explore each theme with every participant. This will allow you to compare your participants' responses to each theme to identify the underpinning reality that you seek to reveal.

Alternatively, where you adopt an interpretivist position (Section 4.4), the way in which you deal with your list of predetermined themes, and the associated questioning, is likely to be more flexible and contingent on what each participant says. The order in which you ask each participant questions will vary depending on the flow of the conversation and the data they share with you. You may omit a theme or modify your questions in a particular interview, given the context or some other characteristic that you encounter. New aspects to explore, including themes, may emerge from participants' interpretations or the research setting.

The way you develop your interview will also be likely to affect the nature of the semi-structured interviews you conduct. Themes may emerge from the findings of previously conducted unstructured interviews, and possibly from discussions with others such as your friends or project tutor. Either of these sources would suggest that you have commenced your research inductively (Section 4.5), possibly where you are using Grounded Theory Method (Section 5.5 and Section 13.9) or an inductive (data driven) Thematic Analysis (Section 13.6), Template Analysis (Section 13.7) or Explanation Building and Testing (Section 13.8). Here you will naturally follow an exploratory and emergent course of action and allow your interview themes and associated questions to evolve depending on what emerges.

Where themes used in your semi-structured interviews are derived from existing theory you will be commencing your data collection deductively, and your intention will be to test this theory in the context of your own research (Section 4.5). Here, questions about these theoretically deduced themes will need to be asked consistently across interviews to produce comparable and valid data to test the applicability of this theory in your research context. These semi-structured interviews may be used in conjunction with a theory testing approach in Thematic Analysis (Section 13.6), Template Analysis (Section 13.7) or deductive Explanation Building (Section 13.8).

Where a surprising fact emerges from the data you have already collected and analysed inductively, you will be likely to switch to deduction. The use of both induction

and deduction is associated with an abductive approach (Section 4.5). As you move to deduction, you will need to use associated questions in a consistent way in subsequent semi-structured interviews to test your emergent theory in the context of these interview settings. Apart from containing the list of themes and questions to be covered, your interview guide for this type of interview will also be likely to contain specific questions to initiate discussion, a possible list of prompts to promote further discussion, and comments to close it. These are discussed in Section 10.7. Data from a semi-structured or in-depth interview may be audio or video recorded with the consent of the participant alongside your making notes.

In-depth/unstructured interviews

In-depth interviews are informal and used to explore in depth a general area in which you are interested and are also referred to as **unstructured interviews**. They do not use predetermined themes or questions to structure or guide the course of the interview, although you will need to have at least some idea of the topic, event, experience, or aspect that you wish to explore with participants. You will start with a broad topic or idea that you wish to explore with your participants, giving them the opportunity to talk freely. These interviews are non-directive, the participant, rather than the interviewer, guiding the interview process (Cassell 2015). They have also been labelled an **informant interview** as the interviewee's perceptions guide the conduct of the interview and related discussion. By comparison, in a **respondent interview** the interviewer exercises greater direction, questions allowing the interviewee's opinions to emerge through their responses (Powney and Watts 1987).

Within an in-depth interview there is no list of pre-determined questions, rather questions and prompts are adapted and emerge during the interview from what your participant tells you rather than being determined beforehand. To minimise contaminating their responses, you need to ensure that prompts and questions emerge from what your interviewee tells you, clarifications being used to probe and explore meanings, rather than building on your own pre-conceived ideas.

Despite an in-depth interview being led by the participant, there is still scope for some variation. The conduct of an in-depth interview may be conversational; albeit one in which the participant is still the principal speaker with the interviewer focused on asking spontaneous questions and using prompts that build upon what the participant says, encouraging them to talk further. Alternatively, it may be directed almost entirely by the interviewee talking, with the researcher needing to ask very few questions or provide hardly any prompts. Mark undertook one such telephone interview with the owner of an SME about success. After his initial question, the participant required few prompts or further questions, offering a wide variety of insightful data on what success meant to them and why. In such a **phenomenological interview**, the focus is upon understanding the participant's lived experiences from their perspective by exploring the meanings and explanations they attribute to these (Cassell 2015). Such interviews are usually undertaken from an interpretivist position to gain an individual's in-depth insider account of a phenomenon such as success. With a phenomenological interview, you give the participants the overall topic and ask them to tell you about it, thereby not imposing your own frame of reference upon them. This means you will have very little control over the interview questioning compared to semi-structured interviews.

You may develop a research design that starts with in-depth interviews as an exploratory and emergent stage. These could be followed by semi-structured interviews to examine possible relationships between themes that have emerged from the analysis of data

from your in-depth interviews. Your semi-structured interviews could then be used to build well-grounded theories, or to test extant theories (Section 2.7). Through a process of convergence (Dick 2013), you use increasingly refined probing questions and compare current data with earlier data to move towards an emergent explanation or theory. This is similar to a grounded theory approach, although without the same level of procedural specificity (Section 5.5), participant selection being based on using heterogeneous (maximum variation) sampling (Section 7.9).

As a participant researcher or practitioner researcher you can both listen and engage in informal conversations, as well as pre-arrange semi and unstructured interview interviews. Where you engage in participant observation or use an action research strategy, your immersion in the research setting will help your taking part in natural, authentic conversations. Where permission (consent) is given by participants, these can generate useful research data.

The link between structure and research purpose

Qualitative research and, in particular, semi-structured and unstructured interviews, allow us capture and understand how people make meaning and sense of what is



Box 10.1 Focus on management research

The role of qualitative inquiries in the time of Covid-19

An editorial by Teti and colleagues (2020) in the International Journal of Qualitative Methods highlights how qualitative inquiries are the best method for capturing social responses to the pandemic and, in particular, to understand how people make meaning and sense of health and illness. They note how the open-ended nature and focus on 'how?' questions in individual and group interviews (as well as observations) can explore different viewpoints, meanings and motivations and highlight five essential contributions of qualitative methods during epidemics:

- 'People's health behaviours do not always fit into epidemiological models' (2020: 2) interviews

highlighting that many different factors limit compliance with preventative measures.

- Additional negative impacts of the virus on 'vulnerable populations' (2020: 2), despite or because of preventative measures, are unlikely to be picked up by quantitative measures; for example, closing Schools removing access to free school meals.
- Qualitative measures are well suited to exploring the reasons epidemic solutions and strategies work or fail uncovering 'unexpected consequences or surprising outcomes' (2020: 2); for example, the impact of preventative measures such as social distancing on mental health.
- Qualitative reports of 'medical response experiences' (2020: 3) highlight the needs of medical and other first responders uncovering challenges such as limited testing availability and a lack of personal protective equipment.
- Ethnographic studies have highlighted the importance of 'getting community buy-in' (2020: 3) in achieving affective social responses to manage the outbreak.

Table 10.1 Purpose of different interviews structures

| | Research purpose | | | |
|-----------------|------------------|-------------|-------------|------------|
| | Exploratory | Descriptive | Explanatory | Evaluative |
| Structured | | ✓✓ | ✓ | ✓ |
| Semi-structured | ✓ | | ✓✓ | ✓✓ |
| Unstructured | ✓✓ | | | ✓✓ |

✓✓ = more frequent, ✓ = less frequent

happening (Box 10.1). Each of the interview structures we have outlined has a distinct purpose. Structured interviews are normally used to gather data that can be quantified and analysed quantitatively (Chapter 12), for example as part of a survey strategy. Semi-structured and unstructured (in-depth) interviews are used to gather qualitative data which are normally analysed qualitatively (Chapter 13), for example as part of a case study or Grounded Theory strategy. Such qualitative designs are likely to be used to answer ‘what’ and the ‘how’, and ‘why’, questions.

In Section 5.3 we outlined how the purpose of your research can be classified as either exploratory, descriptive, explanatory or evaluative. Different interview structures may be used to gather data for each kind of study (Table 10.1).

- In an exploratory study, in-depth interviews can be helpful to find out what is happening and to understand the context. Semi-structured interviews may also be used in an exploratory study. Both can be used to collect background or contextual material for your study, and are helpful where you adopt an inductive approach, such as in the development of a grounded theory (Sections 5.5).
- In a descriptive study, structured interviews can be used to identify general patterns. They are helpful where you adopt a deductive approach to test a theory, the standardisation making it easier to collect quantitative data on the same variables and test statistical propositions or hypotheses (Chapter 12).
- In an explanatory study, semi-structured interviews may help to understand relationships between variables, such as those revealed from a descriptive study (Section 5.5). Structured interviews may also be used in relation to an explanatory study, where quantitative data are collected. Research interviews used for an explanatory purpose may be useful in both inductive and deductive approaches because of the intention to explain why relationships exist (Section 2.5).
- In an evaluative study, you may find it useful to use one or more interview structures, depending on the nature of your study. Semi-structured interviews may be used to understand and evaluate the relationships. Research interviews used for an evaluative purpose may be useful in either inductive, deductive or abductive approaches (Section 4.5).

We can therefore see that the various types of interviews have a variety of purposes. The key is to ensure consistency between your research question(s) and objectives, the strategy you will employ and the methods of data collection you will use. Box 10.2 provides a checklist to help you in your deliberations about whether to use in-depth or semi-structured interviews.



Box 10.2 Checklist

To help you decide whether to use in-depth or semi-structured interviews

- ✓ Is your research exploratory or explanatory?
- ✓ Will it help to be able to probe interviewees' responses to build on or seek explanation of their answers and meanings?

- ✓ Will it help to seek personal contact in terms of gaining access to participants and their data?
- ✓ Are your data collection questions large in number, complex or open-ended?
- ✓ Will there be a need to vary the order and logic of questioning?
- ✓ Will the data collection process with each individual involve a relatively lengthy period?
- ✓ Will interviews allow you to reveal and explore social phenomena that you would not be able to observe in action?

10.3 Interview media and modes

We can also differentiate interviews by the medium in which they are conducted – online, face-to-face or telephone; and their mode – one-to-one or group (one-to-many, or two-to-many participants) (Figure 10.2).

Interview media

Online

The Covid-19 pandemic hastened a seismic change in the use of online media for synchronous communication, particularly through the use of cloud-based video communication platforms such as Zoom and Skype. Invariably, there has been a corresponding increase in potential participants' familiarity with the medium and an increased the use of online interviews for research, although they can be more difficult to conduct than face-to-face interviews due to the more limited visual cues you and your participant can sense. Synchronous and asynchronous online interviews (discussed in detail in Section 10.8) are increasingly

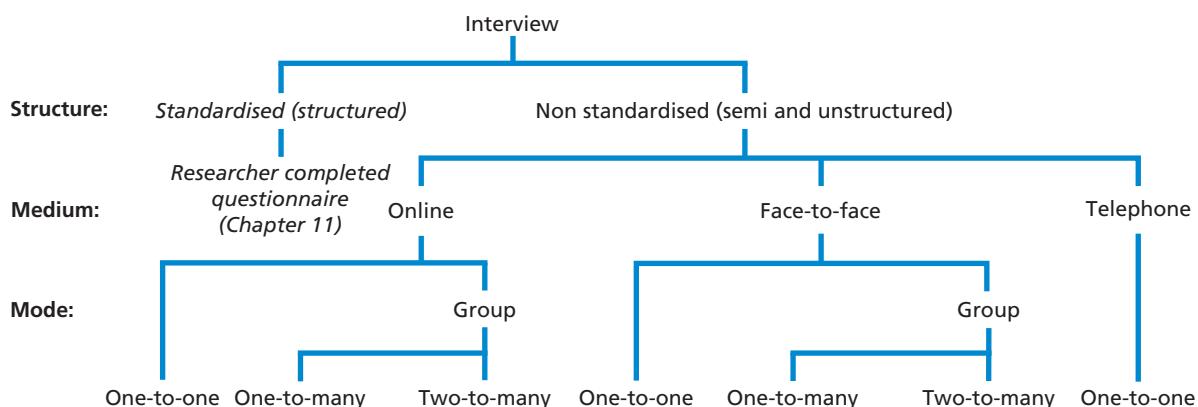


Figure 10.2 Interview structure, medium and mode

seen as more convenient or appropriate as problems of distance between participants, and associated travel costs, are reduced. However, the inevitable reduction in visual cues from the participant, alongside the security of the online platform and logistical issues such as availability of computing equipment and Internet need to be considered (Lobe et al. 2020). Online interviews are associated with both semi-structured and unstructured interviews and can be conducted in on-to-one, one-to-many and two-to-many modes.

Face-to-face

Despite the increase in the use of online interviews, face-to-face interviews, where you meet your participant in person, are still used widely. An advantage of this mode is that you meet each of your research participants, are more aware of visual cues, as well as verbal and paralinguistic (voice tone and non-verbal sounds) signals, and have a greater opportunity to build rapport while allaying any concerns that participants may have about sharing data with you. A face-to-face interview may encourage open discussion, leading to data that are rich and free from bias, given the scope to explore and check understandings during this meeting.

Telephone

Despite the increase in the use of video communication technologies, there may be some situations where you conduct an interview by telephone. Distances and the quality of Internet connections between you and your participants may mean that using a telephone is the only feasible way to conduct interviews, although these interviews are likely to be more difficult, an absence of visual cues meaning you and your participants will have to rely on verbal and paralinguistic signals in your communication. Its use is associated primarily with one-to-one interviews and is discussed in more detail in Section 10.9.

Interview modes

One-to-one interviews

Interviews may be conducted on a one-to-one basis, between you and a single participant. As we go on to outline, such one-to-one interviews may be conducted in person, through a face-to-face interview, over the telephone, or as an Internet-mediated interview (Figure 10.2).

Group interviews

There may be other situations where you conduct a semi-structured or in-depth research interview with a small number of participants to explore an aspect of your research through a group discussion that you facilitate. In some circumstances, two interviewers may conduct an interview, such as in the case of a group interview, where one interviewer leads the discussion and the other acts as principal note taker. Both one-to-many and two-to-many modes are used in group interviews where two or more participants take part in an interviewer managed discussion.

There are two different types of group interview. One type is referred to as a group interview and the other as a focus group (Figure 10.2). These titles are sometimes used interchangeably, although this should be avoided as each have a distinct purpose (Section 10.10). Be warned; a variety of terms are used interchangeably to describe group interviews, which are often wrongly assumed to have equivalent meanings (Boddy 2005).

These include focus group, group interview, group discussion and various combinations of these words!

In this chapter **group interview** is used as a general term for all semi-structured and in-depth interviews conducted with two or more participants at the same time. **Focus group** is used to refer to a specific type of group interview where the topic to be explored is predetermined and precisely defined and the role of the researcher is to facilitate or enable discussion amongst participants rather than lead this, or be the focal point of interaction (Krueger and Casey 2015). In this way, a focus group is a type of group interview, but not all group interviews should be labelled as focus groups.

10.4 The potential of semi-structured and in-depth interviews

Semi-structured and in-depth interviews can, dependent upon your methodological choice (Section 5.4), be used on their own in a quantitative research design, or as part of a multiple or mixed methods design. You may, as part of a survey strategy decide to use in-depth or semi-structured interviews initially to help identify the questions that should be asked in an online questionnaire. The data that you gather from such exploratory interviews will be used in the design of your questionnaire. Alternatively, semi-structured interviews may be used as part of a mixed methods design to explore, explain or validate themes that have emerged from analysing questionnaire data.

The potential for using such interviews is dependent upon four aspects:

- the research purpose;
- the need for personal contact (in some form);
- the questions to be asked;
- the time required from participants.

We examine each of these in turn.

Research purpose

Where you are undertaking an exploratory study, or a study that includes an exploratory element, it is likely that you will include in-depth or semi-structured interviews in your design. Similarly, an explanatory study is also likely to utilise interviews to help infer causal relationships between variables (Sections 2.4 and 11.4), understand reasons for decisions that participants have taken or understand reasons for their attitudes and opinions.

Semi-structured and in-depth interviews also provide the opportunity to ‘probe’ a response, where you want your interviewees to explain, or build on, their previous answers. This is important if, for example, you are adopting an interpretivist philosophy and wish to understand the meanings that participants ascribe to various phenomena (Section 4.4). Interviewees may use words or ideas in a particular way and probing their meanings will add significance and depth to the data. They may also lead the conversation into areas that you had not previously considered but which are significant for your understanding, helping you to address your research question and objectives, or formulate your research question. Interviews also afford each interviewee an opportunity to hear herself or himself ‘thinking aloud’ about things she or he may not have previously thought about. While you should be able to collect a rich and detailed set of data, you need to be

aware that how you interact with your interviewees and phrase questions will impact on the data you collect.

Need for personal contact

We have found that managers and employees are more likely to agree to be interviewed, rather than complete a questionnaire, especially where the interview topic is seen to be interesting and relevant to their current work. An interview provides them with an opportunity to reflect on events without needing to write anything down. It allows you to offer personal assurances about the way their data will be used (Sections 6.2 and 6.5).

In most research designs interviews will be pre-arranged, the interviewer and participant meeting at a specific time for an agreed period either virtually using cloud-based peer-to-peer software or telephone, or in a specified place. However, for some research designs, such as participatory inquiry (Section 4.3), or where as an internal or practitioner researcher you participate in the research setting rather than entering just to conduct a pre-arranged interview (Section 5.9), your scope to conduct interviews more spontaneously increases. Such participatory research designs include using participant observation (Section 9.3) or action research (Section 5.5).

Situations are likely to occur where you will consider the choice between using research interviews and other qualitative methods such as observation (Chapter 9). In this regard, a distinction has been made between contrived and natural data. **Natural** or naturally occurring **data** are those observed from real conversations that take place in everyday, authentic situations. **Contrived data** are those that result from a researcher organising an experiment, interview or survey (Speer 2008). For some research topics there are reasons why it is not possible to collect observed, natural data. These reasons relate to the taken-for-granted assumptions, sensitivity and hidden nature of some social phenomena (such as in personal relationships) that mean it is difficult to gain access to and observe these in action. Using interviews to explore such phenomena means that you can gain access to authentic accounts that you would not be able to observe in action. As a result, the distinction between natural and contrived data may be too rigid (Speer 2008).

Questions to be asked

An in-depth or semi-structured interview is likely to be the most advantageous approach to attempt to obtain data in the following circumstances:

- where there are a large number of questions to be answered;
- where the questions are open ended;
- where the order and logic of questioning may need to be varied (Box 10.3).

Time required from participants

Often the complexity of issues to be covered or their number and variety mean that a semi-structured or in-depth interview is the best or only means of collecting data. In our experience, where expectations have been established clearly about the length of time required and participants understand and agree with the objectives of the research interview, they are willing to be interviewed. We have found that our participants tend to be generous with their time, especially when their interview is arranged at a time convenient



Box 10.3 **Focus on student research**

The need to vary the order and logic of questioning

Val undertook a series of semi-structured interviews into the approaches used to manage public relations (PR) activities in 30 organisations. It soon became evident that it would not be meaningful to ask the same questions in each organisation. For example, some organisations had centralised PR as part of the marketing function, whereas in other organisations it was devolved to individual business units. Another significant theme was associated with the public-relations styles adopted. Some organisations adopted

a 'press agency' approach where the main focus was to get the organisation or product mentioned in the media as often as possible, the nature of the mention being of secondary importance. Others adopted a 'public information' approach where the main aim was to get media exposure for the organisation or product.

The impact of these and other themes meant that it was not sensible to ask the same questions at each interview, even though many questions remained applicable in all cases and the underlying intention was to ensure consistency between interviews. It was not until each interview had started that Val was able to learn how these different themes were addressed within the particular organisation. Fortunately, the flexibility offered by using semi-structured interviews enabled her to do this.

for them. However, for those of you who fancy a free lunch, we do not want to raise your expectations falsely, and the start time for an interview should not be set with this in mind!

Your aim will be to obtain data in relation to each interview theme or the overall particular focus while respecting participants rights not to answer particular questions. Where you conduct the interaction skilfully an interview is more likely to achieve this than the use of a self-completed or interviewer-completed questionnaire. Where your participant does not answer to a particular question or questions, you are likely to have some idea why a response was not provided. This may lead you to revise the question or to compose an alternative. Section 6.7 considers the ethical issues associated with seeking to obtain answers.

10.5 Data quality issues and evaluating interviewing practice

Before discussing how to prepare for and conduct semi-structured or in-depth interviews we consider data quality issues associated with these types of research interview and how to evaluate your interviewing practice. This is because your preparation for and conduct of interviews will be influenced by the need to ensure quality data and, as part of this, your own interviewing practice. We also recommend you evaluate and reflect upon your own practice by undertaking a pilot interview before collecting data for your research.

Data quality issues

Reliability/dependability

The lack of standardisation in semi-structured and in-depth interviews can lead to concerns about reliability/dependability (Section 5.8, Table 5.6) and whether alternative researchers would reveal similar information. The concern about reliability/dependability in these types of interview is also related to issues about three types of potential bias. The first of these

is **interviewer bias** where the comments, tone or non-verbal behaviour of the interviewer creates bias in the way participants respond to questions. This may be because you impose your own beliefs and frame of reference through the questions that you ask. You may also demonstrate bias in the way you interpret responses. Where you are unable to gain interviewees' trust, or perhaps where your personal credibility is seen to be lacking, the value of the data given may also be limited, raising doubts about its validity and reliability.

One response to helping minimise interviewer bias is paying careful attention to phrasing your questions, so they do not lead the participant to a particular response (summarised later in Table 10.2). In addition, it is worth reiterating that findings are not necessarily intended to be repeatable, reflecting reality at the time they were collected, in a situation which may be subject to change. The assumption is that the circumstances being explored are complex and dynamic and the value of using in-depth or semi-structured interviews derives from their flexibility to explore the complexity of the topic. Consequently, ensuring such research could be replicated by other researchers is neither realistic nor feasible without undermining the strength of this type of research.

You still, however, need to explain your research design, the reasons underpinning your choice of strategy and methods, and how your data were obtained. This will allow others to understand the processes you used to reach your research findings. The use of in-depth or semi-structured interviews should not lead to a lack of rigour in relation to the research process – rather there is a need to use a rigorous design and ensure your explanation of how the data were obtained and analysed provides sufficient detail to show your findings are dependable.

Interviewee and participant bias

Interviewee or response bias can be caused by participants' perceptions about the interviewer, or perceived interviewer bias. However, the cause of this type of bias may not be linked to perceptions of the interviewer. Taking part in an interview is an intrusive process. This is especially true in the case of in-depth or semi-structured interviews, where your aim will be to explore events or to seek explanations. Although a participant may, in principle, be willing to participate, they may still be sensitive to the unstructured exploration of certain themes. Participants may decide not to reveal and discuss an aspect because this would intrude on sensitive information they do not wish, or are not empowered, to discuss. They may therefore provide only a partial 'picture' of the situation, casting themselves in a 'socially desirable' role, or their organisation in a positive (or even negative) fashion.

Participant bias results from the nature of the individuals or organisational participants who agree to be interviewed (Box 10.4). The amount of time required for an interview may result in unwillingness to take part by some biasing your sample (Section 7.9). This is an issue that you will need to consider carefully and attempt to overcome through the approach taken to sampling.

Cultural differences

Further concerns may arise from cultural differences between the interviewer and intended interviewees. Gobo (2011) sees the research interview as the product of individualistic societies, which may not be so well suited to societies and participants with a different cultural orientation. He argues that the research interview makes certain assumptions:

- it is acceptable to discuss issues with outsiders;
- issues may be considered public and able to be discussed rather than being kept private and restricted;
- it is permissible for a person to hold independent views and to speak as an individual.



Box 10.4 Focus on student research

Willingness (or otherwise) to be interviewed

Saffron's research project involved her interviewing people about their perceptions of the benefits of different hair products. She decided that the best way to

conduct these interviews was, with the permission of the owner, to interview customers at her local hairdresser. Saffron discovered that although some of the customers were willing to be interviewed, others were not. A minority of customers, often smartly dressed in business suits, refused outright, saying that they had insufficient time. In contrast, others, particularly pensioners, were happy to answer her questions in considerable detail and appeared to wish to prolong the interview.

Gobo also refers to societies where there may be a tendency to respond to an interviewer's questions only by being positive or by agreeing.

Cultural differences that you as an interviewer may have to cope with may be more subtle. They may affect rapport and what the participant is willing to say, how the researcher interprets their words and meanings, or fails to understand these, and influence the questions that the interviewer asks. Where your research involves interviewing participants from a different culture, whether in a cross-national or multicultural setting, you will need to ensure that you minimise any form of bias or threat to reliability. Cultural reflexivity may well help here. This will involve you reflecting on the nature of the relationship between you and your intended participants and how differing and similar cultural customs may affect your interactions (Court and Abbas 2013). Prior to interviewing, you could visit a workplace and observe, listen or participate in informal conversations so that you become more familiar with the research setting. Such understandings will help you develop rapport with potential participants and gain their acceptance.

Cultural reflexivity will also involve you considering how to engage your participants and involve them. This is likely to include evaluating how best to conduct interviews: whether to conduct these in one-to-one or group mode; choosing the most appropriate level of structure and formality to use; and whether to attempt to gather data in a single interview or in more than one to develop rapport and understanding. It may also be appropriate to use a distinct format such as a narrative interview or a thematic format or engage in participatory forms of research such as action research or ethnography (Section 5.5). Adopting a culturally reflexive approach may help to overcome cultural differences that affect what is discussed and not discussed, clarify what is important and what is not, and reveal what should be followed up and explored. Box 10.5 offers a series of questions you can ask yourself to better understand and anticipate the cultural context in which an interview takes place.

Generalisability/transferability

An issue is often raised about the generalisability/transferability of findings from qualitative research interviews, although the validity/credibility of the data they produce is generally seen to be less of an issue (Section 5.8). Generalisability/transferability refers to the extent to which the findings of a research study are applicable to other settings. This may be questioned in relation to the statistical generalisability of qualitative research studies where these are based on a small sample. However, this should not be interpreted as meaning that a qualitative study is intrinsically less valuable than a quantitative study as a well-planned rigorous case study can produce valuable findings. As we noted in Section 10.2, such



Box 10.5 Checklist

To better understand an interview's cultural context

- ✓ Where is the interview location?
- ✓ Who chose the interview location?
- ✓ What are the interview location's characteristics?
- ✓ What meanings does the interview location have for the participant/interviewer?

- ✓ What is the cultural relationship between the participant and the interviewer?
- ✓ Are there any cultural norms that might influence what is said/not said?
- ✓ Are there any issues of status/gender/age that might influence what is said/not said?
- ✓ How can you as interviewer know what is not being said?
- ✓ (Where there is more than one researcher) What cultural/status/gender/age/other factors may affect relationships between the researchers?

Source: Developed from Court and Abbas (2013).

studies are more likely to be used to explore and explain and provide insights that can be used to develop theory, rather than to provide statistical generalisations about a population.

Interviews can also be used to test existing theory or for an emergent theory to be subsequently discussed in relation to a pre-existing theory. Where you can relate your research project to existing theory you will be able to demonstrate that your findings have a broader theoretical significance than the case or cases that form the basis of your work. It will be up to you to establish how the findings from your particular case or cases are related to existing theory to demonstrate their transferability or generalisability. In addition, your full description of the research questions, design, context, findings and resulting interpretations in your project report will allow other researchers to design similar research projects to obtain comparable data.

Validity/credibility

Validity/credibility refers to the extent to which the researcher has gained access to a participant's knowledge and experience, and can infer their intended meanings from the language used by that person. The scope to explore meanings during a semi-structured or in-depth interview may help to enhance the validity/credibility of the data collected, although forms of bias and cultural differences may impair this outcome.

Semi-structured and in-depth interviews can achieve a high level of validity/credibility (Section 5.9) where conducted carefully using clarifying questions, probing meanings and by exploring responses from a variety of angles or perspectives (Section 10.7). These can help you to build trust and rapport, collect sufficient data and provide you with the opportunity to ask participants to check these data. Credibility may also be achieved by accounting for negative cases (those that are counter to other cases) during analysis, in the explanations you develop and being reflective and reflexive about your research.

Evaluating your interviewing practice

Our discussion has already highlighted how semi-structured and in-depth interviews are complicated interactions in which you as interview face challenges of phrasing questions that are unbiased and unlikely to be misunderstood. Within each interview, you need to try and ensure that you have not dominated the verbal interactions either by directing what is said or in the amount you have spoken. You need to avoid leading the participant to answer in a particular way by offering them your own experiences ideas or views, or

potentially biasing their responses through the way you have phrased questions. You also need to minimise the likelihood of your participant being unwilling to answer or elaborate on the questions you ask.

We consider it important that you engage critically with your own interviewing practice and identify how it might be developed and improved early on in your research by undertaking at least one pilot interview. Reissner (2018) highlights four widely used practical strategies which can be used for this:

- Detailed examination of your interview interactions focussing on the conversation and associated contextual and cultural factors as knowledge is created through the interview.
- Engagement by immersing yourself in the data you have collected such as watching video recordings or listening to audio recordings or reading interview transcripts to learn what has worked well and identify interactional challenges.
- Reflection using reflective diaries or journals to capture your own insights into your interviewing practice through introspection and dialogue.
- Enlisting participants in reflexive dialogue to jointly examine how they respond to different aspects of the interviews, including the questions. However, this depends upon participants being willing to invest time beyond participating in the interview.

All but the last of these can, in particular, be aided by mapping the conversational space, and then immersing yourself in the data to look at your own interviewing practice including the cleanliness of the language you have used to ask questions. We outline each of these next. The use of reflection is also outlined in Section 1.5, while further details regarding enlisting participants in reflexive dialogue can be found in Hibbert et al. (2014).

Mapping conversational space

Reissner (2018) argues that mapping and analysing the conversations that make up interviews can help you engage reflexively with your own interviewing practice, in particular, your questioning and listening and relationship with the participants. She proposes using a conversational space map (Figure 10.3) to represent the pattern of interactions between

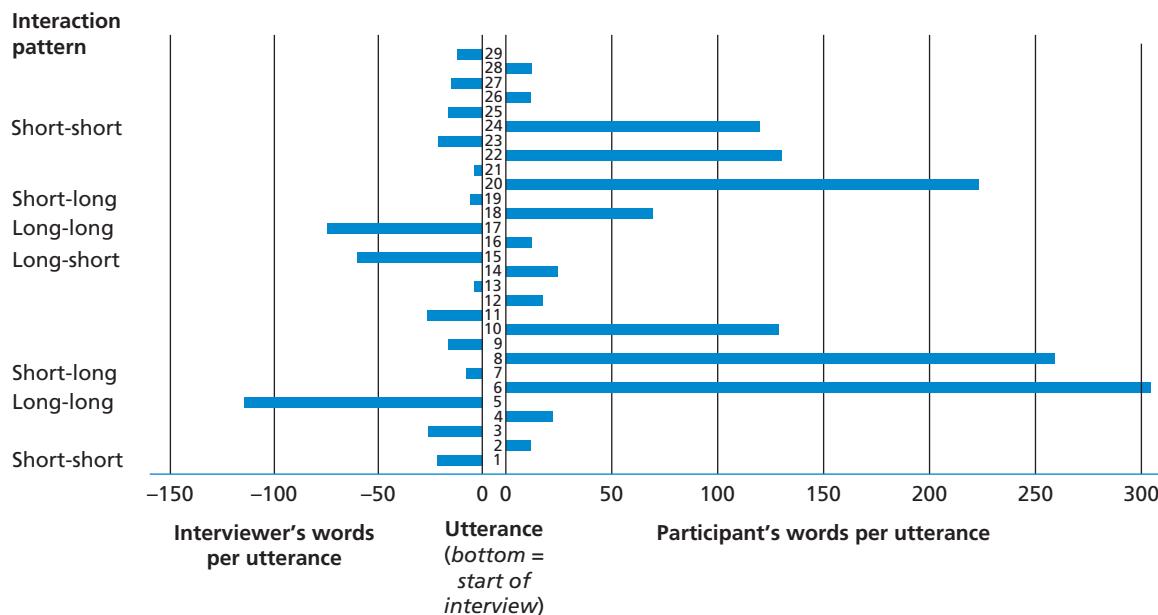


Figure 10.3 Conversational space map

the interviewer and participant, providing an overview of the interview dynamics within which more detailed analysis of your practice can be undertaken. Subsequent analysis of the interview transcript and the audio-visual or audio recording can help identify the actual reasons for these patterns and highlight potential issues. (Box 10.6)

Each **conversational space map** represents visually the number of words in each utterance made by the interviewer and interviewee for an entire interview. It can be used to highlight the different interaction patterns between you as interviewer and the participant.



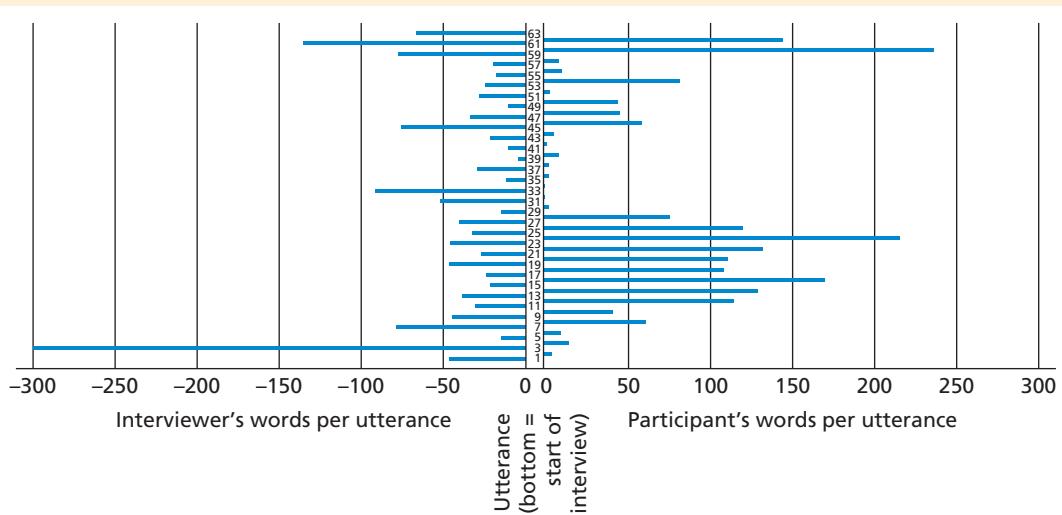
Box 10.6 Focus on student research

Using a conversational space map

Sophia had decided to use semi-structured thematic interviews. Prior to collecting data, she undertook, and video recorded a pilot interview using Zoom cloud-based video conferencing. Having reflected upon what she had said to the participant and her use of paralinguistic signals and visual cues she decided to examine the pattern of interaction with her participant, creating a Conversational Space Map of their interactional practice. This revealed that, although the participant spoke more than Sophia did, there were parts of the interview where she dominated the conversation. Looking at the first half of her interview, she felt her first few lengthy utterances were easily justified as she was explaining her study, reading the confidentiality statement, and asking demographic

questions, which did not require lengthy answers. For the remainder of the interview, although there were places where her participant gave a longer response, Sophia felt she was talking too much. In contrast her participant was offering very short answers.

Sophia decided to re-examine the interview recording and transcript to try and understand why this might be. She noticed some of the questions she asked were vague and she had needed to clarify what she wished to know to the participant by asking supplementary questions. She also noted that although she had asked several closed questions that the participant had answered very briefly, her participant often seemed unwilling to elaborate on the answers given. She decided she needed to review the questions outlined on her interview checklist and think more carefully about her use of closed and open questions as well as how each could be phrased more precisely so her participants could better understand them and be encouraged to give fuller answers.



In Figure 10.3 the interview starts (at the bottom of the ‘map’) with a short-short pattern where the interviewer is collecting demographic data. Where this pattern continues, it can suggest the interviewer is finding it difficult establish flow or the participant is refusing to engage. This is followed by a long-long interaction where the interviewer puts their question in context. Such patterns where they continue can indicate that the interview is an exchange of ideas which may not be the purpose of the interview. The series of short-long interactions occur as the interviewer elicits answers through follow-up questions; often considered the ideal for interviews. The next interaction is a long-short interaction where the participant was reluctant to answer. This pattern, where it continues, can also indicate the interviewer is telling the participant too much about their own views and dominating the interview. After further interactions long-long and short-long interactions, the latter comprising follow-up questions, the interview closes with a series of short-short interactions.

Assessing language cleanliness

We have already noted the importance of reviewing the impact of your questioning on responses and, in particular, the potential introduction of bias. Fortunately, Cairns et al. (2021: 9) offer a method for reviewing this called the **cleanliness rating**. This method is derived from **clean language practice**, where the language used to ask questions minimises the potential for contaminating the participant’s experiences. It assesses the extent each question constrains or enables the participant to respond from within their own world view. Within this, interview questions are placed a continuum from ‘clean’ to ‘leading’, their likely impact in terms of bias varying (Table 10.2).

Table 10.2 Language cleanliness rating categories

| Cleanliness rating | Description | Extent of presupposition/evaluation |
|---|---|-------------------------------------|
| Classically clean | Question only uses universal constructs and participant content. | none |
| Clean repeat | Recap of participant’s words with no new content, presupposition, or evaluation. | none |
| Contextually clean – topic | Question introduces interview topic/theme with minimum superfluous content or supposition. | minimal |
| Contextually clean – logic | Question, while not classically clean, remains within logic of participant’s description, with no new content, presupposition, or evaluation. | minimal |
| Mildly leading | Question/statement suggests or implies an answer or way of answering but participant’s responses indicate no reason to doubt their responses. | present, but no reason to doubt |
| Strongly leading | Question/statement suggests or implies an answer and participant’s responses raise doubts of authenticity. | present and gives reason to doubt |
| Other: Non leading comments or gestures | Comment/gesture/non-verbal utterance encourages participant to continue and indicates interviewer is attentive. Statement/question /response about the process. | none |

Source: Developed from Cairns-Lee, H., Lawley, J. and Tosey, P. (2021, online first)

To use the cleanliness rating, following an interview you:

- 1** Obtain a full transcript of the interview, and;
- 2** (Re)familiarise yourself with the context of the interview including the information provided to participants (e.g. information sheet, interview structure, mode, medium and type).
- 3** Examine each interviewer question or prompt in the context of the interviewees prior and subsequent responses.
- 4** Assign one of the six cleanliness rating categories (Table 10.2) to the question asked.
- 5** If possible, ask a friend to also assess the interview and compare your ratings.

Undertaking a cleanliness rating for your own interview questioning is especially helpful in developing your own interviewing skills. It is also helpful when writing about your method and providing a reflexive account of how you undertook your research and showing the extent your findings are meaningful.

10.6 Preparing for semi-structured or in-depth interviewing

Like all research methods, the key to a successful interview is careful preparation. We believe it is crucial that you plan precisely how you are going to demonstrate your competence and credibility to obtain the confidence of your interviewees and collect quality data.

To help ensure data quality, we now consider some key aspects that your preparation needs to include whatever the modes of interviewing or medium used. These are:

- contextual knowledge;
- explaining the interview's thematic or distinctive format prior to the interview;
- appropriateness of the intended interview medium;
- logistics and time management.

Contextual knowledge

You will already have some background knowledge from reviewing the literature (Chapter 3). You now need to become knowledgeable about the context in which each interview will take place. There is likely to be information online in organisational websites, national, local or specialist news sites and relevant trade association's websites. Organisational websites usually allow access company annual reports, other organisation-related and product information, market and financial data and press releases. Research databases providing access to relevant organisational information will be accessible through your university library or professional association. Such background information will help you to develop a good level of contextual knowledge which, in the interview, will help to demonstrate your credibility. It will also allow you to assess the accuracy of responses and encourage participants to offer a more detailed account of the topic under discussion. As you undertake later interviews, you will also be able to draw on the initial analysis that you made of data previously collected.

Successfully interviewing participants from different cultures requires some knowledge about those cultures. An in-depth interview offers the opportunity to explore meanings, including those that may be culturally specific, but you will need to be aware of cultural differences and their implications (Section 10.5). Brinkmann and Kvale (2015) highlight some of the verbal and paralinguistic signals and visual cues that may have contrary or

different meanings between cultures. For example, answering ‘yes’ to a question may indicate agreement in some cultures, but in others may be a way of indicating the question has been understood, or its importance recognised. A nod of the head indicates agreement in some cultures but in others it may mean something else. Brinkmann and Kvale (2015) note the importance of being aware of social conventions in a culture to understand how answers are constructed and not to cause offence. Cultural differences exist not only between countries but between groups, social classes and organisations and some prior knowledge about those you wish to interview will invariably be helpful.

Explaining the thematic or distinctive format

Your credibility as interviewer can be promoted through the supply of relevant information to participants before the interview. Providing participants with a list of the interview themes or outlining the distinctive format before the event as part of the information sheet (Section 6.7), where this is appropriate, should help this.

Themes

This list of themes (Box 10.7) can help to promote validity and reliability because it informs the interviewee about what you are interested in and provides them with the opportunity to prepare for the interview by assembling supporting organisational documentation from their files. Access to organisational documentation also allows for triangulation of the data provided (Sections 8.2 and 8.3). Our experience is that participants are generally willing to supply a photocopy or a PDF file of such material, subject to confidentiality concerns and the need to exclude personal details in the research report.

Interview themes may be derived from the literature that you read, the theories that you consider, your experience of a particular topic, common sense and discussions with co-workers, fellow students, tutors and research participants, or a combination of these approaches. You will need to have some idea of the broad theme (or themes) that you wish to discuss with your participants even if you intend to commence with exploratory, in-depth interviews as part of a Grounded Theory strategy to your research project (Section 5.5).



Box 10.7 Focus on student research

Developing interview themes

Karl was interested in understanding why some employees in his organisation used the online IT Help Desk while others did not. This subject was felt to be important in relation to perceptions about service-level agreements, service relationships and service quality. He decided to provide his interviewees with a list of themes that he wished to explore during interviews.

After some deliberation and reading of the academic literature he came up with the following list of themes:

- when and how employees use the online IT Help Desk;
- the support employees feel they are receiving;
- the services employees feel the online IT Help Desk should be providing;
- employees' knowledge of service-level agreements;
- the extent to which employees consider the online IT Help Desk is meets their needs.

He subsequently used this list of themes to develop his interview guide (Box 10.8).



Box 10.8 Focus on student research

Extract from an interview guide

Karl was interested in understanding why some employees in his organisation used the online IT Help Desk while others did not. Using his interview themes (Box 10.7), he began to develop his guide:

Help Desk support

- 1 Have you used the online IT Help Desk?

- a *Probe:* In what ways? [asking for actual examples]
- b *Probe [using example(s) given]:* Can you tell me more about what happened when you used . . . [use example given]?
- c *Probe [if not used]:* Can you tell me why you have not used the online IT Help Desk?

- 2 What do you consider an online IT Help Desk should do?

- a *Probe:* In what way do you think they should . . . [use phrase from participant's response]?
- b *Probe:* why do you think they should . . . [use phrase from participant's response]?

Without at least some focus, your interview will lack a sense of direction and purpose. You should therefore start with a set of themes that reflect the variables being studied, or at least one or more general questions related to your research topic that you could use to start your interview. These can be incorporated into your interview guide (Box 10.8) and included as an appendix in your project report.

This lists topics that you intend to cover in the interview along with initial questions and probes that may be used to follow up initial responses and obtain greater detail from your participants. When creating your guide, you need to try to ensure that the order of questions is likely to be logical to your participants and that the language you use will be comprehensible. Using your guide, you will be able to develop and/or explore research themes through the in-depth or semi-structured interviews that you conduct to see whether you can identify and test relationships between them (Chapter 13).

Distinctive format

For distinctive format interviews, you will need to describe briefly the format of the interviews in your information sheet. For example:

Narrative interview: in the interview you will be asked to tell your story about your experiences at [organisation name].

Biographical interview: in the interview you will be asked talk about your career from your first job until now.

Critical incidents interview: in the interview you will be asked to describe actual incidents or events that relate to particularly good or particularly bad customer service.

Visual interview: in the interview you will be given a series of advertisements for household products and asked questions about the advertisement.

You will also need to prepare an interview guide and any additional materials, such as photographs for a visual interview.

Appropriateness of the intended interview medium

The medium through which you choose to conduct your interviews will influence the data you collect and whom you collect it from. As we discussed in Section 6.6, you should

Table 10.3 Selected attributes of interview media

| Attribute | Online (video communication) | Telephone | Face to face |
|---|--|---|---|
| Interviewer's control over interview environment | Minimal regarding suitability although can request. Not possible to know who else is present if out of camera view. | Minimal regarding suitability although can request. Not possible to know who else is present. | High |
| Participants' privacy^a | Need to remove all possible identifiers from software, video background, alongside other confidentiality and anonymity aspects | Confidentiality and anonymity aspects | Confidentiality and anonymity aspects |
| Technology requirements | Web enabled, video communication | Telephone and audio recorder | Minimal other than audio/video recorder |
| Visual cues, verbal and paralinguistic signals | Visual and audio although limited by what camera shows, time lag between speech and video | Audio only | Visual and audio |
| Interview recording | Through the technology | Needs additional equipment | Needs additional equipment |
| Interview transcription | Automatic, but needs cleaning | Additional cost – voice recognition software can help | Additional cost – voice recognition software can help |

^a Discussed in Sections 6.7 and 6.8

Sources: Authors' experiences, Carter et al. (2021); Cassell (2015); Lobo et al. (2020);

choose the environment for your interviews with regard to your own personal safety. You should also think about the impact that the medium may have upon your participants and the way they respond during the interview (Table 10.3, Box 10.9).

The environment should be convenient for your participants, where they will feel comfortable and where the interview is unlikely to be disturbed. For face-to-face interviews, you will be able to decide, in some cases with the participant, where the interview will be conducted. However, for telephone and online interviews you have far less influence, while you will be able to discuss these requirements when arranging the interview and request the participant is in a space that is quiet with a good Wi-Fi/telephone connection, where they will not be overheard when talking normally, and where they are unlikely to be interrupted; unlike in Box 10.9, you will not be physically present.

You also need to try and ensure a location that is quiet so that outside noise will not reduce the quality of your recording of the interview. Each of us has experienced situations when conducting interviews where noise from outside the building or even from within it has been disruptive. In particular, Mark recalls an interview in a room where noise from building work outside meant that although he was able to hear the participant's responses clearly while the interview was taking place, much of the audio recording of this interview was unintelligible due to the sound of a very loud pneumatic drill! Similarly, we have all experienced those online conversations where conversations are interrupted by someone entering the room uninvited.

In many cases, the interview location will be arranged by those whom you interview. When you interview organisational participants such as managers in their offices, this has the advantage that they can find documents which support points they are making. Where you are undertaking an online interview, you will need to try and impress on the participant the importance of not being disturbed.

The different interview media also offer different advantages in terms of being able to see or hear visual cues and verbal and paralinguistic signals relating to the participant's feelings and responding accordingly. These are most obvious in face-to-face interviews, online interviews reducing their utility due in part to the camera angle and time lags between video and audio, with only verbal signals being available for telephone interviews.

Technology requirements and the ability to easily record the interview also differ between media (Table 10.3). The availability of technology may in some cases limit participation from certain groups, particularly where a web-enabled device and Wi-Fi is needed and, in some cases, the authority to install the software (Carter et al. 2021).

Logistics and time management

Logistics

Before conducting interviews, you need to consider their scheduling. Thought needs to be given to the number of interviews it is feasible to undertake in a given period, including time required to compose notes and/or transcribe audio recordings of each one and undertake your initial analysis of the data collected (Section 13.4). Additional time and consideration will be required if you also need to translate your transcription into another language.

Conducting face-to-face interviews may become costly, particularly where participants are dispersed due to the need to travel, although this can be kept to a minimum by cluster sampling (Section 7.6) or using online (Section 10.8) or telephone (Section 10.9) modes. Choice of mode should be determined primarily by the nature of the research question and objectives, and the suitability of the mode for the intended participants rather than cost considerations. This highlights the need to examine the feasibility of the proposed question and research strategy in relation to resource constraints, including time available and expense, before proceeding to the collection of data.



Box 10.9 Focus on student research

Choosing an appropriate environment

Writing her reflective diary after her first interview Emily commented:

I had agreed a date, place and time for the interview with my participant, endeavouring to ensure this was convenient for them. Our interview lasted approximately 35 minutes and was conducted in a campus coffee shop. This was a convenient, quiet, but also

somewhere where I felt we would be comfortable/relaxed.

An hours before the interview, I found out the café closed at 4 pm, exactly the time I had arranged the interview. I sent text messages, emails to try and arrange an alternative place for the interview but got no response. When I got there 10 minutes before the interview, only the bit selling coffees was closed! Phew! The rest of the coffee shop remained open, and my participant turned up having not seen my messages. The interview went ahead as planned and I apologised for not being able to buy them a cup of coffee. They thought it was quite funny, but next time I will check more carefully.

Interviewing is time-consuming. Where the purpose of the interview is to explore themes or to explain findings, you may need a fairly lengthy discussion. In such cases the time required is unlikely to be less than one hour and could easily exceed this, perhaps taking two hours or longer. This may have an adverse impact on the number and representativeness of those who are willing to be interview participants. Where managers or other potential participants receive frequent requests to participate in research projects, they are likely to have considered how much of their time they are willing to devote to such activities. It is therefore important for you to establish credibility with, and to engender the interest of, potential participants.

Time management

You need to consider very carefully the amount of time that will be required to conduct an interview. In our experience, this is usually underestimated. The likely time required should be highlighted in any initial contact, and it may be better to suggest that interviews are envisaged to last up to, say, one, one and a half, or two hours, so that a willing participant sets aside sufficient time. Some negotiation may be possible with an interested participant who feels unable to agree to a request for, say, two hours but who is prepared to agree to a briefer meeting. The interview can also be arranged at a time when the interviewee will be under least pressure.

Another possible strategy is to arrange two or more shorter interviews to explore a topic thoroughly. This allows participants to reflect on the themes raised and questions being asked, and provide a fuller account. To establish this option, it may be beneficial to arrange an initial meeting with a potential participant to discuss this request. A series of exploratory interviews may then be agreed. Consideration also needs to be given to the number of interviews that may be undertaken in a given period; it being easy to overestimate what is practically possible (Box 10.10). Where undertaking interviews at one establishment, it may be more practical to undertake several interviews in one day, although there is still a need to maintain concentration, make notes and write up information, and to conduct your initial analysis. Even in this situation, conducting more than three interviews per day is likely to be challenging.

It is likely that, having assessed the advantages and disadvantages (Table 10.5), you will have audio or video recorded your interview. You will therefore need to decide whether to work directly from the recording or to produce a transcription of all or parts of the recording. This decision will depend on your research strategy and the way in which you intend



Box 10.10 **Focus on student research**

Calculating the number of in-depth interviews to be undertaken in one day

Firoz arranged two interviews in a capital city during the course of a day, which involved travelling some miles across the city during the lunch hour. Two interviews appeared to be a reasonable target.

However, he massively underestimated the time required forgetting to allow for: the total travelling time to and from the city; the time to find the appropriate buildings; the transfer time during a busy period; the time to conduct the interviews; the need to maintain concentration, to probe responses, to make initial notes and then to write these up without too much time elapsing. Because of his experience, Firoz decided not to conduct more than one interview per day where significant travel was involved, even though this necessitated more journeys and greater expense.



Box 10.11 Checklist

To prepare for your semi-structured or in-depth interview

- ✓ What level of knowledge about your research topic will be required to demonstrate your competence and credibility to gain the confidence of your participants?
- ✓ What level of contextual knowledge will be required to demonstrate your competence and credibility to gain the confidence of your participants?
- ✓ What level of knowledge about the culture of your participants will be required to gain their confidence before they are willing to share data?
- ✓ What will be thematic focus or distinctive format you wish to use during your interview?
- ✓ What type of information will it be useful to send to each participant prior to the interview?
- ✓ What did you agree to supply to your participant when you arranged the interview? Has this been supplied?
- ✓ Have you considered the impact that your interview location may have on participants' responses and for your own personal safety?
- ✓ Have you considered the appropriateness of the interview medium and the impact it may have on participants' ability to take part and their responses?
- ✓ Have you considered logistical issues, particularly the time needed to operationalise your interviewing?

to analyse your qualitative data (Chapter 13). For example, using a Grounded Theory strategy (Section 13.9) is likely to mean that you will need to transcribe fully each interview. Each hour of audio recording is likely to take at least seven hours to transcribe or to process ready for entry into computer-assisted qualitative data analysis software, unless you are a very competent audio-typist, or you know one who will undertake this task for you! Use of software to assist the transcription of audio recordings may also be helpful.

Where it is necessary to translate your interview, care will be needed to ensure that the meanings contained in the original or source language are reproduced authentically in the translated language. However, translation may be more problematic than just technically producing language equivalence. Chidlow et al. (2014) discuss the need to go beyond translational equivalence and to use a contextualised approach to promote understanding. We consider potential problems associated with translations in Section 11.5 and outline different translation techniques in Table 11.4, together with their respective advantages and disadvantages.

Box 10.11 provides a checklist of the key points considered in this section to help you to prepare for semi-structured or in-depth interviews.

10.7 Conducting one-to-one face-to-face interviews

We first consider conducting one-to-one face-to-face semi-structured or in-depth interviews before using this as a comparator for interview forms using other modes and media. The aspects we discuss here are intended to help maximise the reliability/dependability and validity/credibility of the data produced. They are, in many instances are also applicable to one-to-many and two-to-many modes as well as different forms of online and telephone interviews. These aspects relate to the:

- interviewer appearance;
- opening the interview;

- approach to questioning;
- use of different types of questions;
- behaviour during the interview;
- demonstrating attentive listening skills;
- summarising and testing understanding;
- dealing with difficult participants;
- recording data;
- evaluating your interview practice.

We discuss these in turn. Key points are summarised as a checklist at the end of this section (Box 10.16).

Interviewer appearance

Your appearance may affect the perception of the interviewee. Where this has an adverse effect on your credibility in the view of interviewees, or results in a failure to gain their confidence, it can affect the reliability of the information provided. Where appropriate you should consider wearing a similar style of clothing to those to be interviewed. For example, your interviewees would not expect you to wear the same workwear that they need to put on to work on the production line. Rather, you will need to wear clothing that will be generally acceptable for the setting within which the interview will occur (Box 10.12).

Opening the interview

Where the interviewee has not met you before, the first few minutes of conversation will have a significant impact on the outcome of the interview – again related to the issue of your credibility and the level of the interviewee's confidence. Often such interviews occur in a setting that is unfamiliar to you. Despite this, it is your responsibility to manage the conversation. You will need to explain your research to the participant and, hopefully, gain consent (Section 6.7). As part of this you will need to establish your credibility and gain the interviewee's confidence. During initial discussions the interviewee is often uncertain about sharing information, and about how their data will be used. They need clarification about the exact nature of the data that you wish to obtain. We have found the participant information sheet (Section 6.7, Box 6.17) and consent form (Box 6.18) are both extremely



Box 10.12 Focus on student research

Checking out the dress code

Mal arranged to visit the administration centre of a large insurance company on a Friday to conduct one-to-one interviews with staff drawn from its telephone sales division. He felt that it was appropriate to wear fairly 'formal' clothes to match what he thought would be the dress code of the organisation. Indeed, for four

days of the working week this assumption would have been appropriate. However, the organisation had recently introduced the practice of not wearing such formal work clothes on Fridays. Mal found himself the only one dressed formally in the organisation on the day of his visit. Taking lunch proved to be a memorable experience, as he mingled with everyone else dressed in jeans and tee shirts, etc. His 'mistake' proved to be an amusing opening at the start of each interview rather than a barrier to gaining access to participants' data. Indeed, it might not have been appropriate for him to match the 'dress-down' style of participants too closely.



Box 10.13 Focus on student research

Opening a semi-structured interview

As part of her research project, Beth undertook a series of face-to-face semi-structured interviews with freelance consultants working for a range of organisations. She covered the following points at the start of each interview:

- The participant was thanked for considering the request for access and for agreeing to the meeting.
- The purpose of the research and its progress to date were outlined briefly. As part of this, the participant was given an information sheet to keep.
- The previously agreed right to confidentiality and anonymity was reiterated by stating that nothing said by the participant would be attributed to him/her without first seeking and obtaining permission.

- The participant's right not to answer any question was emphasised and that the interview would be stopped if the participant wished.
- The participant was told about the nature of the outputs to which the research was intended to lead and what would happen to the data collected during and after the project.
- The offer to provide a summary of the research findings to the interviewee was also restated and the participant was told when this would happen.
- The request to audio record the interview was restated and, where agreed, this was used subsequently.
- Before the substantive discussion started, Beth again requested permission to undertake the interview, summarised the themes to be covered, confirmed the amount of time available and requested that the participant read and signed the informed consent form.

All these points were dealt with within the first few minutes of the interview.

helpful in reducing such anxieties. There may also be a degree of curiosity on the part of the interviewee about why they were asked to participate. This can offer an opening for you to start a conversation, probably before the main interview commences. It may be relevant to ask participant about their role within the host organisation. However, you need to make sure that these opening moves to demonstrate credibility and friendliness, are not overstated, do not introduce bias, or that too much time is used, and the interviewee starts to become bored or restive.

The start of the interview needs to be shaped by you. It is your opportunity to minimise, wherever possible, the participant's uncertainties about providing data, establish their rights, and based upon this, hopefully, obtain informed consent. Box 10.13 provides a structure that you can adapt for starting your interviews.

Your assurance that confidential information is not being sought should make participants more relaxed and open about the data that they are willing to provide. Combined with assurances about anonymity, this should increase the level of confidence in your trustworthiness and reduce the possibility of interviewee or response bias. You can also demonstrate your commitment to confidentiality by not naming other individuals or organisations that have participated in your research, or by talking about data you have already obtained.

Approach to questioning

Your approach to questioning should reduce the scope for bias during the interview and increase the reliability of the information obtained. Your questions need to be phrased clearly, so the participant can understand them, and asked in a neutral voice tone. They

need to direct the participant to the research focus to elicit useful and interesting responses. These can be followed up using appropriately worded probing questions.

Questions and prompts that lead the participant, or which indicate bias on your part, therefore need to be minimised. Cairns-Lee et al. (2021) offer a typology of question features which may result in bias by potentially leading participants' responses:

- 1** Introduced content, where the interviewer uses terms that have not previously been used by the interviewee, resulting in the interviewee adopting the interviewer's terms;
- 2**
 - a** Presupposition of structure and context, where the structure or syntax of the question presupposes a situation exists which the interviewee has not already stated;
 - b** Presupposition of a logical relationship, where the interviewer's question or prompt presupposes a cause-and-effect relationship that the interviewee has not already stated;
- 3** Evaluation where the interviewer expresses an opinion or raises a doubt or objection to something already stated by the interviewee.

These, along with example-biased questions and non-leading alternatives are outlined in Table 10.4.

Interviewing without introducing bias, however unintentional, is exceptionally difficult, even for experienced interviewers. Even though you may be able to spend time crafting your initial opening questions and potential prompts, there is no guarantee you will be able to follow these during your interview. The spontaneous nature of semi-structured and unstructured interview questioning will almost certainly result in unintentional leading as you will have little chance to craft your questions. It is therefore important to assess the cleanliness of the language used in your questions and the impact in relation to the

Table 10.4 Leading interview questions

| Leading feature | Example-biased question (nature of leading) | Example non-leading alternative |
|---|---|---|
| <i>Introduced content</i> | Can you outline the freedom you have in how you work on a project? [introduces the metaphor of 'freedom', and presupposes this is related to how work on a project] | How do you work on a project? |
| <i>Presupposition of a situation</i> | Is it hard to decide which project to work? [presupposes the participant can decide which project they work on] | Can you tell me anything else about project working? [followed by prompting if the participant raises the issue of deciding which project to work on] |
| <i>Presupposition of a relationship</i> | How did the change impact on your workload? [presupposes the change did impact on participant's workload] | And when the change happened, what happened to you? [followed by prompting if the participant says there was impact] |
| <i>Evaluation</i> | Presumably this job requires only basic skills? [evaluates the skills required as being only basic, undermining participant's own opinion of self] | And what kind of skills are needed for this job? |

Source: Developed from Cairns-Lee, H., Lawley, J. and Tosey, P. (2021, online first)



Box 10.14 Focus on student research

(Mis)understanding terminology

Sven was conducting an interview with the European sales manager of a large multinational corporation.

Throughout the interview the sales manager referred to the European Division. Sven assumed that the sales manager meant continental Europe. However, by chance, later questions revealed that, for this organisation, Europe extended into parts of Asia, including Turkey, the United Arab Emirates, Saudi Arabia, Kuwait and Israel. Until this point in the interview, Sven had assumed that these countries were the responsibility of another sales manager!

responses given (Table 10.2), and recognise explicitly where bias may have been introduced in your findings.

Long questions or those that are really made up of two or more questions (known as double-barrel questions) should also be avoided if you are to obtain a response to each aspect that you are interested to explore. Questions should also avoid overtly theoretical concepts or jargon since your understanding of such terms may vary from that of your participants. If theoretical concepts or specific terminology need to be used, you will have to check that both you and the participants have the same understanding (Box 10.14).

When asking questions, it is important that, wherever possible, these are grounded in the real-life experiences of your participants and are discussed using their terminology and words. One approach to questioning (and distinctive interview format) which makes use of key participant experiences is the **critical incident technique**. In this, participants are asked to describe in detail a critical incident or number of incidents that are relevant to the research question. A **critical incident** is defined as an activity or event where the consequences were so clear that the participant has a definite idea regarding the effects (Keaveney 1995).

It is also important to consider when to ask sensitive questions. Leaving these until near the end of an interview will provide your participant with some time to develop trust and confidence in you and to allay any doubts about your intentions, as Box 10.15 illustrates. This also affects the nature of questions you may ask early in an interview, as you attempt to build trust and gain your participant's confidence. When asking potentially sensitive questions, their wording needs to avoid any negative inferences related to, for example, responsibility for failure or error.



Box 10.15 Focus on student research

Establishing trust and asking sensitive questions

Sam recalls an occasion when her treatment by her participants altered as her group interview progressed. For the first hour of a two-hour interview, it appeared

to her that the participants were convinced that she was really there to sell them a consultancy service. When they accepted that she was not going to try to sell them something, the mood of the interview changed, and they became much more relaxed and responsive to the questions that Sam wished to ask. It was at this point that she was able to ask and pursue more sensitive questions that could have led to the interview being terminated during the period when the participants mistrusted her motives.

Use of different types of questions

Formulating appropriate questions to explore areas in which you are interested is critical to achieving success in semi-structured or in-depth interviews. It is also important to word your questions factually and avoid introducing new content, presupposing a situation or relationship or offering a prior valuation (Table 10.4). We now discuss the types of question that you can use during semi-structured and in-depth interviews.

Open questions

The use of **open questions** allows participants to define and describe a situation or event. An open question is designed to encourage the interviewee to provide an extensive and developmental answer and can be used to reveal attitudes or obtain facts. It encourages the interviewee to reply as they wish. An open question is likely to start with, or include, one of the following words: ‘what’, ‘how’ or ‘why’:

- ‘What is your current role in the organisation?’
- ‘How does your organisation recruit new employees?’
- ‘Why did your organisation open an office in Singapore?’

Probing questions

Probing questions are used to explore previous responses that are of significance to the research topic. Although often worded like open questions they request a particular focus or direction, often repeating some of the participant’s own words, in the case of the next question ‘successful’:

- ‘How do you know if this new marketing strategy is successful?’
- ‘What happened next?’
- ‘And why is that?’

Questions may also be left unfinished, for example:

- ‘That’s interesting . . .’
- ‘Tell me more about . . .’

Probing questions can also be used to seek an explanation where the response does not reveal the reasoning involved or where you do not understand the interviewee’s meaning:

- ‘How do you plan to develop your use of digital marketing during the next two years?’
- ‘That’s interesting: do you mind telling me more about this relationship between the shift to the partnered product development model and the expansion of the established products division?’

The use of reflection may also help you to probe a theme. This is where you will ‘reflect’ a statement made by the interviewee by re-using some of their words, in the next question ‘did not understand the need for advertising’:

- ‘Why do you think that these employees did not understand the need for advertising?’

The intention will be to encourage exploration of the point made without offering a view or judgement on your part. Where an open question does not reveal a relevant response,

you may also probe further using a supplementary question that rephrases the original question.

Specific and closed questions

These types of question may be used as introductory questions when you commence questioning about a particular interview theme:

‘Could you tell me about the change to the pricing policy?’
‘Can you outline the production process?’

They can also be used to obtain specific information or to confirm a fact or opinion (Section 11.3):

‘How old are you?’
‘How many people responded to the customer survey?’
‘Has the old Central Region been merged with the Southern Region?’
‘What is your opinion of the new training programme?’

Other means to prompt responses

There are several ways of prompting further answers to a question you have asked. These include:

- follow-up expressions, such as: ‘Ah’, ‘Oh’ or ‘Um’;
- short follow-up statements, such as: ‘That’s interesting’;
- short follow-up questions, such as: ‘Will you please tell me more?’, ‘When did that happen?’ or ‘What happened before?’;
- short reflective questions where you repeat what you have just been told to reflect it back, such as: ‘So that was when . . . ?’ or ‘So, they felt the investment had been worthwhile?’;
- interpretation and extension questions, where you seek to explore the implications of an answer, such as: ‘So what do you see as the implications of the move to online retailing?’;
- silence, where the participant is effectively invited to fill this by offering more information;
- using these devices in combination to explore a theme, but you will need to be careful if you use this approach as it may be interpreted as being overbearing, stressful and confrontational. It will be more productive and ethical to maintain an even pace and respectful stance when asking questions.

Questions to avoid

In phrasing questions, remember that you should avoid using leading questions or proposing answers (Section 11.4). A classic example of a leading question is:

‘So, tell me, is this the first time you have cheated in a test?’

Behaviour during the interview

Appropriate behaviour by the researcher should also reduce the scope for bias during the interview. Comments or non-verbal behaviour, such as gestures, which suggest

your own opinions, should be avoided. Rather, a neutral (but not an uninterested) response to the interviewee's answers should be used to ensure your own opinions do not bias responses. You should enjoy the interview opportunity, or at least appear to do so; any appearance of boredom on your part is hardly likely to encourage your interviewee!

Your posture and tone of voice may also encourage or inhibit the flow of the discussion. You should sit slightly inclined towards the interviewee and adopt an open posture, avoiding folded arms. This should provide a signal of attentiveness to your interviewee. Tone of voice can also provide a signal to the interviewee. You need to project interest and enthusiasm through your voice, avoiding any impression of anxiety, disbelief, astonishment or other negative signal.

Demonstrating attentive listening skills

The purpose of a semi-structured or in-depth interview is to find out and understand your participant's views, explanations and meanings. This type of interaction will not be typical of many of the conversations that you normally engage in, where those involved often compete to speak rather than concentrate on listening. You therefore need to listen attentively, attending to and being sensitive to your participants by spending the time needed to listen to them to build your understanding. However, you will need to keep your own opinions to yourself so as not to introduce bias.

It will be necessary for you to explore and probe explanations and meanings, but you must also provide the interviewee with reasonable time to develop their responses, and you must avoid projecting your own views.

Summarising and testing understanding

You can test your understanding by summarising responses provided by the interviewee using their own words and phrases as much as possible. This will allow your participant to confirm whether your summary is adequate, add further points, and correct your understanding where appropriate. This can be a powerful tool for avoiding a biased or incomplete interpretation. It may also offer a means to explore and probe the interviewee's responses further.

In addition, you may also ask the interviewee to read through the factual account that you produce of the interview (Section 5.11). Where the interviewee is prepared to undertake this, it will provide a further opportunity for you to test your understanding and for the interviewee to add any further points of relevance that may not previously have been apparent. Beware, participants may wish to amend their grammar and expressions to reflect written language rather than their spoken words.

Dealing with difficult participants

Inevitably, during your interviews you will meet some participants who are difficult to interview. It is imperative that you remain polite and do not show any irritation. Although it is impossible for us to highlight all possible difficulties, the most common difficulties are summarised in Table 10.5, along with suggestions about how to deal with them. However, while reading Table 10.5, the best advice we can give is to undertake practice interviews in which a colleague introduces one or more of these 'difficulties' and you have to deal with them!

Table 10.5 Difficult interview participants and suggestions on how to address them

| Recognised difficulty | Suggestion |
|---|--|
| Participant gives only monosyllabic answers, these being little more than 'yes' or 'no' | Reasons for this are varied. If it is due to limited time, or worries about anonymity, then this can be minimised by careful opening of the interview (Box 10.12). If the participant gives these answers despite such precautions, try phrasing your questions in as open a way as possible; also use long pauses to signify you want to hear more. |
| Participant repeatedly provides long answers that digress from the focus of your interview | Although some digression should be tolerated, as it can lead to aspects in which you are interested, you will need to impose more direction. This must be done subtly so as not to cause offence, such as by referring to an earlier relevant point made and asking them to tell you more, or requesting that they pause so you can note down what they have just said. |
| Participant starts interviewing you | This can suggest that you have created rapport. However, you need to stress that you are interested in their opinions and that, if they wish, they can ask you questions at the end. |
| Participant is proud of their status relative to you and wants to show off their knowledge, criticising what you do | This is extremely difficult ,and you will have to listen attentively and be respectful. Remember you are also likely to be knowledgeable about the research topic, so be confident and prepared to justify your research and the research design you have chosen. |
| Participant becomes noticeably upset during the interview and, perhaps, starts to cry | Another difficult one for you. You need to give your participant time to answer your question and, in particular, not do anything to suggest that you are feeling impatient. If your participant starts crying or is obviously very distressed, it is probably a good idea to explain that the question does not have to be answered. Do not end the interview straight away as this is likely to make the participant even more upset. |

Source: King (2004); authors' experiences

Recording data

Where possible we believe it is beneficial to audio record an interview, and also make notes as it progresses. Using both methods to record interview data has several advantages. Notes provide a backup if the audio recording does not work. Making notes can help you to maintain your concentration, formulate points to summarise back to the participant to test your understanding and devise follow-up probing questions. Note taking demonstrates to your participant that their responses are important to you. It also allows you to record your own thoughts and any events that would not be evident from the audio recording. For example, if you think there may be a relationship between themes that you wish to explore later, if your interviewer uses a facial expression or provides another non-verbal cue, or if someone enters the room, you can make a note about each of these. Most people have their own means of making notes, which may range from an attempt to create a verbatim account to a diagrammatic style that records key words and phrases, perhaps using mind mapping (Section 2.3).

The task of note making in this situation will be a demanding one. As you test your understanding of what your participant has told you, this will allow some time to complete your notes concurrently in relation to the aspect being discussed. Most participants recognise the demands of the task and act accordingly. For example, Adrian recalls one participant who paused at the end of the main part of each of his answers to allow notes to be completed before adding some supplementary data that could also be noted down. However, the actual interview is not the occasion to perfect your interviewing skills, and we advise you to practise in a simulated situation: for example, by watching an interview on television and attempting to produce a set of notes.

To optimise the value from the interview you should create a full record, including contextual data as this will help you interpret what your participants say and when you write about your methods in your project report (Box 10.16). If you cannot do this immediately after the interview, this should be done as soon as possible. There is the possibility that



Box 10.16 Focus on research in the news

The last mogul: an interview with Universal Music's Lucian Grainge

By Ann Nicolaou and Andrew Edgecliffe-Johnson

Within this piece drawing on an interview with Lucian Grainge, the authors provide a considerable amount of contextual detail:

Somehow, he [Lucian Grainge] emerged with no significant lung damage. He looks healthy and upbeat as he tells the story publicly for the first time, perched at the head of a conference table that could easily fit two-dozen staff. Today it is just us and some Purell sanitiser bottles, in the type of work-from-home paradise that wealthy CEOs have enjoyed for the past year and a half. The outdoor patio where we sit is flanked by orange and lemon trees, wafting citrus through the air. It's sunny, but not too sunny. The temperature hovers around 24C. A light breeze blows in from the Pacific Ocean. Peak California.

Grainge sizes up the day ahead of him, peppering an assistant with questions. 'We've got the team meeting at 11. Then I have a Zoom. And apparently, we're doing photos at 1, can that be quick? I'm not doing an album art cover. And Abel is coming at 2, right? But that is late for lunch, so we will eat before.' He drifts into a story about a famous pop star showing up at 11 pm to a 7 pm dinner at Nobu. 'The reality is in California, kitchens close at 9.40. They only keep the kitchen open because they know that you're coming with someone.'

There is a vulnerability and scrappiness about Grainge, even as he perches at the top of the business. 'I love to be underestimated' is a phrase he repeats on several occasions. Asked whether falling gravely ill from coronavirus has changed him, he says 'No' definitively before continuing. 'I've been through quite a lot. You don't get from nowhere, to where I am, without always trying to prove yourself. What happened yesterday, happened yesterday. I love the next move and I love winning. I don't romanticise the past.'



Source: Extract from 'The last mogul: an interview with Universal Music's Lucian Grainge' Financial Times, 16 September 2021. Copyright © 2021 The Financial Times Ltd

you may mix up data from different interviews, where you carry out several of these within a short period of time and you do not complete a record of each one at the time it takes place, compromising the trustworthiness of your data. In addition to your notes from the actual interview, you should also record the following **contextual data**:

- the location of the interview (e.g. the organisation, the place);
- the date and time;
- the setting of the interview (e.g. was the room quiet or noisy, could you be overheard, were you interrupted?);
- background information about the participant (e.g. role, post title, gender);
- your immediate impression of how well (or badly) the interview went (e.g. was the participant reticent, were there aspects about which you felt you did not obtain answers in sufficient depth?).

You may be wondering how, if you are recording these data, you can still ensure the anonymity of your participants where this has been promised. As we outlined in Section 6.7, the best course of action is to ensure that your data are completely and genuinely anonymised. To help achieve this you should store the contextual data separately from your interview transcripts. We suggest that you should only be able to link these two sets of data by using a ‘key’, such as an impersonal code number. Where it is absolutely necessary to retain a ‘key’ that allows participants to be linked to their data using their real name, this ‘key’ should be kept securely and separately, not by those who control the data.

Audio recording the interview where permission is given, making notes, compiling a full record of the interview immediately or soon after it has occurred and producing a set of contextual data and related memos (Chapter 13) are all means to control bias and produce reliable data. Most interviewers audio record their interviews, where permission is given. Audio recording interviews has both advantages and disadvantages, and these are summarised in Table 10.6.

Permission must be granted before starting to audio record an interview. You should explain why you believe this is beneficial and to provide guarantees about your participant’s rights over its use. Where it is likely to have a detrimental effect, it is better not to use a recorder. However, most participants adapt quickly to the use of the recorder. It is more ethical to allow them to control the recorder so, if you ask a question that they are prepared to respond to, but only if their words are not audio recorded, they have the option to switch it off (Section 6.7). It will inevitably be necessary to make notes in this situation.

Table 10.6 Advantages and disadvantages of audio recording the interview

| Advantages | Disadvantages |
|---|---|
| Allows the interviewer to concentrate on questioning and listening | May adversely affect the relationship between interviewee and interviewer (possibility of ‘focusing’ on the audio recorder rather than the interview process) |
| Allows questions formulated at an interview to be accurately recorded for use in later interviews where appropriate | May inhibit some interviewee responses and reduce reliability |
| Can re-listen to the interview, especially during data analysis | Possibility of a technical problem |
| Accurate and unbiased record provided | Time required to transcribe the audio recording (Section 13.4) |
| Allows direct quotes to be used | |
| Permanent record for others to use | |

Source: authors' experience.

Evaluating your interview practice

After each interview it is worth making brief notes regarding how you felt it went and any specific issues. Subsequently you can listen to your recording and evaluate your interviewing practice (Section 10.5). The checklist in Box 10.17 is designed primarily to help you conduct one-to-one face-to-face interviews. However, many of the questions asked are also applicable to other interview modes and media.



Box 10.17 Checklist

To help you conduct one-to-one face-to-face interviews

Appearance at the interview

- ✓ How will your appearance at the interview affect the willingness of the participant to share data?

Opening the interview

- ✓ How will you open the interview to gain the confidence of your participant?
- ✓ What will you tell your participant about yourself, the purpose of your research, its funding and your progress?
- ✓ What concerns, or need for clarification, may your participant have?
- ✓ How will you seek to overcome these concerns or provide this clarification?
- ✓ In particular, how do you intend to use the data to which you are given access, ensuring, where appropriate, its confidentiality and your participant's anonymity?
- ✓ What will you tell your participant about their right not to answer particular questions and to end the interview should they wish?
- ✓ How will you explain the structure of the interview?

Asking questions and behaviour during the interview

- ✓ How will you use appropriate language and tone of voice, and avoid jargon when asking questions or discussing themes?
- ✓ How will you word open questions cleanly to obtain relevant data?
- ✓ How will you word probing questions cleanly to build on, clarify or explain your participant's responses?
- ✓ How will you avoid asking leading questions that may introduce forms of bias?

- ✓ Have you devised an appropriate order for your questions to avoid asking sensitive questions too early where this may introduce participant bias?
- ✓ How will you maintain a check on the interview themes that you intend to cover and to steer the discussion where appropriate to raise and explore these aspects?
- ✓ How will you avoid overzealously asking questions and pressing your participant for a response where it is clear that they do not wish to provide one?
- ✓ How will you avoid projecting your own views or feelings through your actions or comments?
- ✓ How might you identify actions and comments made by your participant that indicate an aspect of the discussion that should be explored to reveal the reason for the response?
- ✓ How will you listen attentively and demonstrate this to your participant?
- ✓ How will you summarise and test your understanding of the data that are shared with you to ensure accuracy in your interpretation?
- ✓ Where appropriate, how will you deal with a difficult participant while remaining polite?

Recording data during the interview

- ✓ How will you record the interview data? Where this involves using an audio recorder, have you requested this and provided a reason why it would help you to use this technique?
- ✓ How will you allow your participant to maintain control over the use of an audio recorder, where used, if they wish to do this?
- ✓ Have you practised interviewing to ensure you can carry out all tasks, including listening, note taking and identifying where you need to probe further?

Closing the interview

- ✓ How will you draw the interview to a close within the agreed time limit and thank the participant for their time and the data they have shared with you?

10.8 Conducting one-to-one online interviews

Interviews may also be conducted electronically via the Internet using video communication technologies, email or texts. These are collectively referred to as either **online interviews** or **electronic interviews**. A distinction is made between electronic interviews conducted in real time (**synchronous**) and those not conducted in real time (**asynchronous**). An **asynchronous online interview** will be conducted through exchanges of text. This will use email or text messaging but will involve gaps in time or delays between the interviewer asking a question and the participant providing an answer (Figure 10.4). In this way it is sometimes partly undertaken offline. A **synchronous online interview** will be conducted in real time using email, instant messaging or video communication technologies (Figure 10.4). In this section we briefly discuss asynchronous email and text interviews, and synchronous email and video-based interviews and the advantages and disadvantages associated with each.

Asynchronous email interviews

An email interview is generally described as asynchronous because it is not necessary to ask questions and answer these sequentially without any time gaps. However, it may be possible to conduct an email interview in one period, where the interviewee responds immediately to each question and emails continue to be exchanged until the interviewer draws it to a close and thanks the interviewee for her/his participation. This may be after a pre-arranged period has been reached.

Using email interviews

Each **email interview** consists of a series of emails each containing one or small number of questions and the replies to these. Although you can send one email containing a series of questions, this would really be an online questionnaire (Sections 11.2 and 11.6). After making contact and obtaining agreement to participate, you initially email a question or small number of questions, or introduce a topic to which the participant will (hopefully) reply. You then respond to each reply by asking further questions, raising points of clarification and pursuing ideas that are of further interest. Email interviews may last for some time where there is a delay between each question being asked and an answer being received. This can be advantageous in terms of allowing time for reflection on the part of the interviewer, in forming appropriate questions, and the participant, in terms of providing a considered response, but it may also mean that the interviewee may lose focus and interest so that the email interview ends without all questions being answered. Like all forms of text-based interview data are recorded as they are typed, removing problems

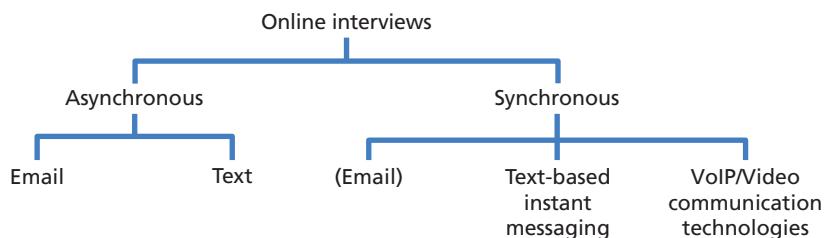


Figure 10.4 Forms of online interview

associated with other forms of recording and transcription such as cost, accuracy and participants' apprehension.

Synchronous text-based interviews

An electronic interview conducted by text-based instant messaging is synchronous because the technology uses real-time transmission. However, there still may be time gaps between the interviewer asking a question and the interviewee providing a response, so it may extend over several periods when both are online. For research topics where the researcher wishes to ask personal or sensitive questions, the anonymity offered by the interviewer and interviewee typing synchronously to each other is likely to produce reliable and useful data. For these types of research topic, the lack of face-to-face contact may prove to be an advantage rather than a disadvantage (Pearce et al. 2014).

Video-based interviews

Video-based interviews using video conferencing or telephony have become far more widely used since the Covid-19 pandemic. They may be conducted using a Voice over Internet Protocol (VoIP) or web conferencing service between two or more people. Many of the associated platforms can also be used to produce an audio-visual recording of the interview and an initial transcript (which will need checking) providing that the research participant consents to this. However, as platforms differ in their functionality, it is important to ensure that the one you intend to use meets your and your intended participants' needs (Lobe et al. 2020).

Using video-based interviews

Using this technology has significant advantages where the population you wish to interview are geographically dispersed, although you may need to be cognisant of you and your participant being in different time zones. However, it is limited to those with access to the necessary technology and, in some instances, permission to install the software. In addition, responses are likely to be shorter and less detailed than for face-to-face interviews (Carter et al. 2021). Alongside concerns about the security of data on these platforms (Lobe et al. 2020), there are several aspects associated with using video interviews that need to be considered. These include (Carter et al. 2021):

- ensuring that the proposed technology is accessible by both yourself and the participant and will allow the desired level of interaction;
- deciding how you wish the chat and mute functions to be set at the start of the interview and subsequently whether you wish participants to be able to access or control them;
- (for group interviews) deciding whether to use break-out rooms for small sub-group discussions;
- deciding whether to allow screen sharing;
- ensuring participants' usernames are switched off to preserve and they can use virtual or blurred backgrounds to help preserve anonymity;
- considering including questions about the use of the online platform to refine your methodology.

In addition to the practical advice offered in Box 10.18 online interviews using video conferencing have their own set of ethical issues that you will need to consider (Sections 6.6 and 6.7).



Box 10.18 Checklist

Advice for interviewing undertaking interviews video conferencing

- ✓ Follow the advice in Box 10.17 as appropriate.
- ✓ Practise by recording yourself and evaluate your questions for cleanliness of language.
- ✓ Make sure the participant knows the weblink (and password) and date and time for the video conference prior to the interview.
- ✓ Do not show meeting identifiers or passwords in the email subject line.
- ✓ Ensure participant(s) need a password to join the meeting.

- ✓ Ensure the location you are using has strong and consistent Internet access.
- ✓ Test your equipment prior to the interview and, if possible, have a back-up in place.
- ✓ Manage participants to prevent others to joining the 'meeting' without an invitation.
- ✓ Use the waiting room to screen participants.
- ✓ Only allow participants to join the 'meeting' after the host (interviewer) has arrived.
- ✓ Subject to permission being granted, video record and automatically transcribe the 'meeting'.
- ✓ Ensure data are not stored within the video conferencing app by turning off participant screen-sharing and saving and storing all recordings directly (with password protection) in accordance with your university's data management protocol.

10.9 Conducting one-to-one telephone interviews

Most semi-structured or in-depth one-to-one interviews occur either face-to-face or online. However, they may also be conducted by telephone, using either a voice/listening-only mode or a video calling service, highlighting the blurred boundary between telephone and online media. In this section we focus on voice/listening only mode, contrasting such telephone interviews with face-to-face interviews.

Telephone versus face-to-face

The purpose of your interview will be broadly the same regardless of whether it is conducted by telephone or face-to-face. However, the way in which an interview is conducted is likely to affect its outcomes. Irvine et al. (2012) compared the nature of spoken interactions in two sets of semi-structured interviews: six conducted by telephone and five carried out face-to-face. The aim of their study was to evaluate the impact of the interview medium on the nature of spoken interactions, based on actual data that had been transcribed systematically. Their analysis revealed five areas of interactional differences between telephone and face-to-face interviews.

Telephone interviews were generally shorter than face-to-face ones although there was a great deal of variation between these interviews. Possible explanations relate to less rapport developed and greater effort required, with the implication that telephone interviews may be less suitable for research studies that are designed to rely on richly detailed and in-depth accounts. In telephone interviews, participants spoke for less time and generally gave shorter answers, being more likely to ask if their responses were adequate. This may be related to lack of visual cues, reduced scope to discuss the purpose of the research at the start of interviews, less rapport developed and greater task orientation in this interview mode. Conversely, while telephone participants spoke for less time, the researcher spoke for slightly more, markedly altering the balance between the two.

Telephone participants were slightly more likely to ask the researcher to clarify or repeat her questions. This did not mean they experienced difficulty in their understanding; instead, this may be explained by the quality of the phone connection, the effort and concentration required in a listening-only mode of interview and resulting fatigue, and the need for interview questions to be phrased clearly and succinctly.

In face-to-face interviews, the researcher was more likely to interact with the participant during an answer, such as by summarising the answer to show understanding. This may be due to greater rapport being developed during face-to-face interviews and because during telephone interviews the interviewer needed to concentrate more on listening given the absence of non-verbal prompts or signals. Researchers using face-to-face interviews used verbal acknowledgements (e.g. by saying ‘Yeah’, ‘Ah’, ‘Oh’, ‘Um’, etc.) to the interviewee more frequently than in telephone interviews. This appears surprising as use of verbal acknowledgements may be expected to be more frequently used in telephone interviews to compensate for the lack of visual contact. Possible explanations suggested may be because of the need to concentrate on listening during telephone interviews and also because the researcher used the lack of visual contact to concentrate on taking notes.

Telephone interviews are associated with several disadvantages. Vogl (2013) places these disadvantages into two broad categories: the limited scope for personal contact, and the reliance on verbal and paralinguistic signals during a telephone interview. We discuss each of these briefly.

Personal contact when using voice/listening during a telephone interview is limited. A telephone interview may be perceived as impersonal and relatively anonymous, and it may be more difficult to establish rapport and trust as a result. Conducting a telephone interview will also be difficult if your participants are uncomfortable using this medium. This may lead to issues of (reduced) reliability where your participants are reluctant to engage in an exploratory discussion by telephone, or even a refusal to take part.

Telephone interviews place reliance on verbal and paralinguistic signals. Such interviews conducted using voice/listening-only exclude visual cues between participant and interviewer aiding understanding. Verbal signals refer to what the interviewer or participant utters, such as:

‘that’s interesting, may I ask you to say more about . . . ’;

I am not sure I understand your question, please can you rephrase it?’

Paralinguistic signals refer to any vocal effects used by a speaker that affect the way words are spoken or sounds are used, and which often convey meaning in their own right. Such effects include voice quality, tone or pitch of voice, rhythm or rate of speech, stress placed on individual words, syllables or sounds and on groups of words known as prosodic or sentence stress, and sounds made using the breath such as a sigh or use of ‘hmm’ or ‘mm’. The focus here is on how things are said as opposed to just what is being said.

Using telephone interviews

Using the telephone to conduct an interview means in the absence of visual cues, you as interviewer will need to concentrate more on how something is said as well as what is said to be sensitive to any nuances used by the interviewee. Listening only may help to provide focus in this interaction but without the scope to recognise, explore and understand visual signalling this will be demanding.

There are invariably practical issues that need to be managed when using telephone interviews. These include your ability to control the pace of a telephone interview and to record data. Conducting an interview by telephone and taking notes is difficult and

demanding. The normal visual cues that allow your participant to control the flow of data shared are absent. As the interviewer, you will be unable to witness the non-verbal behaviour of your participant, which may adversely affect your interpretation of how far to pursue a particular line of questioning. You may also encounter difficulties in developing more complex questions in comparison with a face-to-face interview situation.

Conducting one-to-one interviews by telephone can offer advantages associated with access, speed and lower cost. You may be able to interview participants with whom it would otherwise be otherwise impractical due to the distance and prohibitive costs involved and time required. Even where ‘long-distance’ access is not an issue, conducting interviews by telephone can still offer advantages associated with speed of data collection and lower cost. It may also be safer for the researcher to conduct interviews by telephone in some circumstances.

For some (Holt 2010; Trier-Bieniek 2012), the limited scope for personal contact and reliance on verbal and paralinguistic signals are not considered disadvantageous. Not meeting participants can support the production of open and full accounts due to the anonymity of a voice/listening-only mode reducing inhibitions to discuss very personal matters. The use of the telephone can also allow participants to choose a suitable time of day to be interviewed, to stop an interview in progress when this became unavoidable and to rearrange a time for it to continue, and to move around their environment when necessary to avoid being overheard.

As with all interviews, establishing rapport is important to achieve in-depth answers in telephone interviews. However, there may be less small talk to build such rapport in the initial stage of a telephone interview (Irvine 2011). This more task-oriented approach can set the mood for what follows, leading to a quicker pace and less depth. In those telephone interviews where greater rapport is established, more in-depth answers are likely and greater exploration of these is possible. Irvine (2011: 215) suggests that telephone interviewers ‘consider ways of establishing a more relaxed conversational style prior to asking specific interview questions’.

Audio recording of telephone interviews is usually relatively straightforward with a mobile phone, if you have your participant’s consent. However, for some makes of phone you may need to download a separate app. After your interview we recommend you evaluate your interview practice (Section 10.5). Your reflections and notes should help you to improve your approach to your next telephone interview, rapport and trust.

10.10 Conducting group interviews and focus groups

Semi-structured and in-depth interviews may also be conducted as group interviews both online and face-to-face, one or more interviewers asking questions and records responses with a group of participants. Typically group interviews (and focus groups) involve between 4 and 12 participants, the precise number depending upon the nature of the participants, the topic matter and the skill of the interviewer. Some suggest a narrower range of participants of between 6 and 8, especially if conducted online. Inevitably, the more complex the subject matter the smaller the number of interviewees. Participants are normally chosen using non-probability sampling, often with a specific purpose in mind (Section 7.9), such as they are typical of the group being researched or they represent those who are critical to a particular operation. For many group interviews the underlying reason is that you believe you will learn a great deal from these specific individuals, the group’s dynamics adding depth to the data as shared meanings emerge (Dodds and Hess 2020). Krueger and Casey (2015: 43) refer to such participants as being ‘information rich’.

If using group interviews, or specifically focus groups, consideration of the following issues is helpful.

- Where your research project (or part of it) occurs within an organisation, the request to participate in a group interview may be received as an instruction rather than allowing them choice about whether to take part. This often occurs where the request is sent in the name of a manager or because of your own position in the organisation, and is likely to lead to some non-attendance or to unreliable data. In our experience, participants often welcome the chance to ‘have their say’. However, you need to exercise care in wording the request that is sent to them to take part. You will also need to exercise similar care in your introduction to the group when the interview occurs, providing a clear assurance about confidentiality.
- Once your sample has been selected, participants should be grouped so as not to inhibit each individual’s possible contribution. This may be due to perceptions about status differences or because of the dominance of certain individuals. We would advise using a series of horizontal slices through an organisation so that, within each group, participants have a similar status and similar work experiences. (Using a vertical slice would introduce perceptions about status differences and variations in work experience.) In this way, group interviews can be conducted at several levels within an organisation. A reference may be made about the nature of the group to provide reassurance, and you may consider asking people to introduce themselves only by their first name without referring to their exact job.
- To realise the benefits of a group interview, it is important to encourage every person to participate. This commences when you ask each person to introduce himself or herself. You may also need to encourage contributions by drawing group members into the discussion, particularly where some appear reluctant to take part. This needs to be managed sensitively and participation is likely to increase naturally as group members become more familiar with each other. Occasions may occur during a group interview when participants talk over one another, and you will need to manage the flow of contributions while ensuring that each participant has an opportunity to offer her or his contribution. Where one or two people dominate the discussion, you should seek to reduce their contributions by encouraging others:

‘What do you think, Yuksel?’

‘How does Emma’s point relate to the one that you raised, Kristie?’

A question posed more generally to other group members should also have the effect of inhibiting the contribution of a dominant member:

‘What do other people think about this?’

‘What are your views of Johan’s suggestion?’

As interviewer you can manage the discussion.

- You can try to reduce the contribution of a dominant member by temporarily minimising eye contact with them and draw others into the discussion by looking or gesturing in their direction.
- You will need to remain attentive throughout the interview, appearing friendly and relaxed in your approach but also purposeful and interested, encouraging each member to take part and providing opportunities to listen to and discuss contributions.
- You will need to ensure that participants understand each other’s contributions and that you develop an accurate understanding of the points being made. Asking a participant to clarify the meaning of a particular contribution, where it has not been understood,

and testing understanding through summarising should help to ensure this: ‘Can you explain what you mean by . . . ? [repeat phrase just said by participant]

- You should choose a neutral location rather than, say, in a manager’s office, where participants may not feel relaxed. There should be no likelihood of interruption or being overheard.
- You should, where possible, arrange the seating in a circular fashion so that everyone will be facing inward and so that they will be an equal distance from the central point of this circle.
- Finally, students often ask, ‘When will I know that I have undertaken sufficient group interviews or focus groups?’ Writing about focus groups, Krueger and Casey (2015) suggest that you should plan to undertake three or four group interviews with any one type of participant. If after the third or fourth group interview you are no longer receiving new information, you will have reached **saturation**, in which case you will have heard the full range of ideas.

The demands of conducting all types of group interview, including focus groups, and the potential wealth of ideas that may flow from them mean that it is likely to be difficult to manage the process and note key points at the same time (Box 10.19). We have managed to overcome this in two ways: by audio recording group interviews or using two interviewers. To audio record a group interview you will need the consent of each participant. Where two interviewers are used, one person facilitates the discussion, and the other person makes notes. We would recommend that you use two interviewers, even if you are audio recording the group interview, as it will allow one interviewer to concentrate fully on managing the process while the other ensures the data are recorded. As with one-to-one interviews, your research will benefit from the making of notes about the nature of the interactions that occur in the group interviews that you conduct. We would not advise you to undertake more than one group interview in a day on your own because of the need to write up your notes immediately and the danger of forgetting or confusing data.



Box 10.19 Focus on management research

Using group interviews

Dodds and Hess (2020) undertook nine group interviews with vulnerable people and their support group during the Covid-19 pandemic. The interviews explored a sensitive topic, youth alcohol consumption and family communication, and were moved online during the pandemic. Initially, group interviews were conducted face-to-face in the participants’ own homes, but to comply with social distancing policies they were undertaken using an online video conferencing platform. For both face-to-face and online group interviewers there

were two researchers, one acting as interviewer and the other as a notetaker.

The researchers found online group interviewers were easier to manage than face-to-face group interviews in terms of group dynamics. Shy participants were able to hide partially out of camera giving them a perceived safety barrier and enabling them to speak without having to look directly at the interviewer. In comparison to face-to-face interviews, noticeably more sensitive information was shared around drinking behaviour and personal experiences with alcohol when interviewed online. Online groups interviews were considered engaging and convenient and easy to set up. However, while communication was clear with both groups, the researchers felt conversations flowed better in face-to-face group interviews. They also noted participants were worried about privacy issues online.

Using group interviews

In a group interview your role is to ensure that all participants have the opportunity to state their points of view in answer to your questions and to record the resulting data. The interview tends to be relatively unstructured and fairly free flowing in terms of both breadth and depth of topics. The onus is on you to explain the interview's purpose, to encourage participants to relax, and to initiate, encourage and direct the discussion. You will need to both encourage participants to provide answers to questions, and allow them to range more freely in discussion where this may reveal data that provide you with important insights. Thus, once you have opened the interview (Box 10.12), and the discussion is established, you will need to manage it carefully.

Group interactions can lead to a highly productive discussion as participants respond to your questions and evaluate points made by the group. However, as the opportunity to develop an individual level of rapport with each participant is less (compared with a one-to-one interview), a group effect where certain participants effectively try to dominate the interview may emerge. Some participants may agree publicly with the views of others, while privately disagreeing, any reported consensus, in reality, being a view that nobody wholly endorses and nobody disagrees with (Stokes and Bergin 2006). You therefore need to test the validity of emergent views by trying to encourage involvement of all group members and pursuing the interview's exploratory purpose using open and probing questions. A high level of skill will be required for you to be able to conduct this type of discussion successfully, as well as to try to record its outcomes.

Despite this reference to the potential difficulties of using group interviews, there are distinct advantages arising from their use (Box 10.19). The presence of several participants allows a breadth of points of view to emerge and for the group to respond to these views. A dynamic group can generate or respond to a number of ideas and evaluate them, thus helping you to explore or explain concepts. You are also likely to benefit from allowing your participants to consider points raised by other group members and to challenge one another's views. However, while group interviews, and in particular focus groups, are able to identify principal issues accurately, they are not able to provide the depth and detail in relation to specific issues that can be obtained from individual interviews (Stokes and Bergin 2006).

Group interviews can also help to identify key themes that will be used to develop items that are included in a subsequent questionnaire. For example, the initial use of group interviews can lead to a 'bottom-up' generation of concerns and issues, which subsequently inform a questionnaire's content.

Using focus groups

Focus groups are well known because of the way they have been used by political parties to test voter reactions to particular policies and election strategies, and in market research to test reactions to products, as well as being used in academic research (McNaughton and Myers 2007). A **focus group**, sometimes called a 'focus group interview', is a group interview that focuses upon a particular issue, product, service or topic by encouraging discussion among participants and the sharing of perceptions in an open and tolerant environment (Krueger and Casey 2015). Participant interaction is a key feature of focus group design, although this focus on enabling interactive discussion is used for two distinct purposes.

Positivist or critical realist researchers use the focus group to encourage interactions between participants as an effective means to articulate pre-held views about a particular issue or topic. The aim of using focus groups in this way is to reveal these pre-held views. Interpretivist researchers use focus groups to construct meanings through social

interactions and sense making about a topic. The aim of using focus groups for this purpose relates to the ability to analyse how participant interactions and group dynamics lead to the construction of shared meanings (Belzile and Oberg 2012).

If you are running a focus group, you will probably be referred to as the **moderator** or ‘facilitator’. These two labels emphasise the dual purpose involved in running a focus group, namely to:

- keep the group within the boundaries of the topic being discussed;
- generate interest in the topic and encourage discussion, while at the same time not leading the group towards any particular opinion.

In some focus groups, the moderator’s role may be less evident in comparison to the researcher’s role in other group interviews. This is because the moderator’s role is to facilitate and encourage group interaction. However, while some parts of a focus group may be largely non-directive, other parts may require greater direction from the moderator.

The purpose of a focus group is likely to affect the level of interviewer-led structure and intervention that is required. Focus groups used to reveal participants’ views are likely to be associated with greater structure; those used to study how participants interact are likely to be associated with less structure. Exploratory focus groups are likely to be more unstructured, while those with a theoretical, impression-gathering, diagnostic or explanatory purpose are likely to be semi-structured.

Focus group participants are selected because they have certain characteristics in common that are relevant to the topic being discussed. Group discussions may be conducted several times, with similar participants, to enable trends and patterns to be identified. The size of a focus group may vary according to the nature of the topic. A focus group designed to obtain views about a product is likely to be larger than one that explores a topic related to a more emotionally involved or sensitive construct, such as attitudes to performance-related pay or the way in which employees rate their treatment by management. You may also choose to use smaller groups as you seek to develop your competence in this interviewing technique.

10.11 Conducting visual interviews

Most interviews are based on people talking and listening, even though face-to-face interviews also contain a visual dimension, where visual cues are used to guide their conduct and aid understanding (Section 10.7). The subordinate status of this visual dimension in conventional interviews is altered when the visual aspects are included to create a distinctive format. In a **visual interview**, visual images are used to elicit interviewee accounts and interpretations and stimulate dialogue, the visual becoming fully integrated in the production of participant meanings.

One approach used in visual interviews is **photo-elicitation**. In this technique a participant is given one or more photographic or digital images to interpret. The participant will be asked to focus on and explain the objects or activity in the image from their perspective, constructing meanings related to the image (Box 10.20). At its simplest, this means that you provide a participant with an image to elicit a story. Settings shown in many images will be familiar to the participants and can be used to elicit an insider’s verbal account from them. These images may either be found from those that already exist or created by the researcher or created by research participants (Box 10.20).

Visual interviews can use other types of images to stimulate dialogue and elicit participant accounts and interpretations. We discuss different types of visual image in Section 9.7.



Box 10.20 **Focus on** **management** **research**

Use of photo-elicitation based on participant photography

In a study published in the *International Journal of Consumer Studies*, Vermaak and de Klerk (2017) used a photo-elicitation technique in visual interviews based on participant photographs. The focus of their study was to explore millennial consumers' experiences of using retail dressing rooms. The researchers conducted two sets of research interviews with a purposive sample

of female consumers aged between 18 and 25 years, who were given the task of shopping for clothes.

The first set of interviews explored consumer expectations about using retail dressing rooms. At the end of this first interview, each participant was provided with a digital camera to take photographs of any feature of a dressing room that positively or negatively impressed her in a retail store she visited.

Each participant's photographs were then used by the researchers during a second interview to elicit responses about this experience. Vermaak and de Klerk (2017: 13) report that they used questions to elicit responses from participants, like: 'Why did you take this photo?' or 'What did you think about this dressing room?' Responses to these questions were explored by using probing questions such as, 'Why do you say so?'

These may also be found or created by the researcher, or created by the participant, before being used in a visual interview. We now discuss visual interviews based on researcher-found or created images and visual interviews based on participant-created images.

Using researcher-found or created images

In this format of visual interview, you introduce visual images during the interview and asks the participant to interpret what they see. This type of visual elicitation may also be used during one-to-one interviews and group interviews both online and face-to-face. One digital image may be introduced at a time or a few images may be presented simultaneously. An individual image may show a particular situation or aspect related to the research topic. It may be an image of the research setting. Each image presented individually will be intended to elicit interpretation and stimulate discussion. Several images shown simultaneously will represent different attributes of a topic. For example, in consumer research participants could be asked to discuss the relative merits of the attributes shown, possibly being requested to choose from amongst these or to rank them.

You may also use visual images to gather contextual details or background information about the research setting. These will allow contextual and mundane details to be seen, discussed and evaluated, which would often be missed in a conventional talking and listening interview. You may also introduce other types of visual image such as an organisation chart to be able to understand the broader context within which the research is being conducted.

Visual interviews can use images of the research participants taken by the researcher. These images are explored using a particular photo-elicitation technique known as **autodriving**. This technique was developed and refined by Heisley and Levy (1991). The term autodriving refers to an informant interview that is 'self-driven' by the participant talking about visual images of themselves. You will initially enter the setting where your research participant lives, works or conducts their daily activities and observe them taking visual images that capture aspects of the activity being observed. Images you selected to represent key aspects of this activity will then be used in an interview with the participant who features in them, to elicit their interpretations of what is shown. Autodriving captures the participant's actions and interactions in the setting placing them in the role of an

outsider looking in on a scene in which she or he takes a part. The interviewer uses these visual images to elicit the participant's insider perspective of what is shown, producing insights that would be unlikely to be revealed using any other method.

Using participant-created images

A more participatory approach will range from a researcher encouraging participants to create their own images, which can then be explored in visual interviews (Box 10.20), through collaborative forms of visual research in which participants are involved in different aspects of the research process, including taking, selecting, analysing and interpreting images, to the use of a participant-led approach known as photovoice (Section 9.7). The level of participation increases over this range of possibilities from passive participation to active and fully engaged participation.

We discussed participatory approaches in Section 9.7, including the use of participatory video, participatory audio and participant photography. These approaches involve research participants using their mobile phone or being provided with a camera and given the freedom to choose what to record related to the focus of the research. A further participatory approach exists that may be used before or during the conduct of a visual interview. This is **participant drawing** where a participant is asked to create a drawing using paper and pencil to represent her or his feelings about an issue, or some aspect of his or her experience (Box 10.21). Techniques of participant photography or drawing are



Box 10.21 Focus on student research

Using participant drawing in an interview

As part of her research Heather wanted her participants to reflect on what they saw as the essence of leadership. She decided to ask each of them to 'sketch' what they considered to be the essence of leadership at the start of their interview. Although some participants were initially reluctant to draw, protesting that their drawings would not be 'any good', offering colour pens and paper to each participant resulted in 30 drawings. Most were relatively quickly sketched in 5 to 10 minutes. Despite the initial hesitation from some participants, most were pleased with their sketches. Heather used each participant's drawing as the basis for their subsequent interview. Drawings such as the one included in this box highlight aspects of leadership such as taking followers on a journey



(the path), through troubled times (the clouds and rain) and areas where it might be difficult to see what was going to happen next (the trees), to a successful finish (the chequered flag) where things were brighter (the sun shining).

Source: Developed from Cairns-Lee, H. (2017)

sometimes used to facilitate interaction with children or participants who are less articulate but may be used successfully with many types of participants.

In an approach where the researcher encourages participants to create their own images, a visual interview is still likely to be based on the interviewer eliciting interpretations from each participant. However, you will also need to explore the participant's reason for creating each image, whether literal or abstract (Box 10.21). This will include asking why they chose to take the image, the significance of the objects, activity or interaction shown in the image, and how the image represents their experience or viewpoint. Exploring images may encourage a participant to 'relive' an experience or reflect on its personal significance. The emphasis in this process will be to facilitate participant interpretation, to understand subjective perspectives using a non-judgemental approach.

A participatory approach may involve participants collaborating in various aspects of a research project, possibly through all stages from design to presentation. In this approach the role of the participant in a visual interview is more likely to be that of a collaborative discussant. One technique associated with greater participation is **reflexive photography**. This involves participants engaging in participant photography and reflective interviews to explore how their experiences are situated within social structures and attitudes related to class, gender, race, role and other social categories. This technique will involve participants in several stages of participant photography and reflection, often involving group interviews in which participants meet to present, discuss and analyse their photographic or digital images (Ozanne et al. 2013).

Related to reflexive photography and often led by participants, **photovoice** involves participants using participant photography centred on a research focus of social concern and meeting with other participants in group discussions to present, discuss and analyse images that they have created. Images are then selected by the participants to represent this issue of social concern. These are presented in a public exhibition with the aim of generating wider public support and action, or to change public policy.

10.12 Using diaries and diary studies

We already mentioned diaries as a distinctive format for collecting data when we considered the use of video diaries in observation (Section 9.6). Used in research interviews, a **diary** is a systematic, participant-centred data collection procedure that participants complete over time to record their data prior to being interviewed. The diary entries record data longitudinally as the data are collected consecutively rather than at a single point in time, the time horizon ranging from a few days to months (Section 5.6) (Box 10.22). The period over which the diary will be completed or created will be agreed between the researcher and the participants, although extended use of diaries is likely to lead to participant fatigue and attrition. Diaries are designed to record data about participants' experiences, activities, social interactions, behaviours, attitudes, emotions or sense of well-being. More broadly, diaries provide an unobtrusive means to understand complex processes in organisations. A research project based on the use of research diaries is often called a **diary study**. Research participants in a diary study may be referred to as diarists or diary-keepers.

Diaries allow data to be collected at multiple points in time. This may involve you asking your participants to complete or create a diary entry on a daily basis, or at some other interval depending on the frequency of the activity or aspect that your research is designed to focus upon. This activity or aspect may occur more than once a day, or less often than daily. Box 10.22 provides examples of research participants being asked to complete timed diary entries more than once a day, for a pre-arranged number of days or weeks.



Box 10.22 Focus on management research

Quantitative diary studies

In a study published in the *Journal of Management*, Vogel and Mitchell (2017) undertook research on the ways employees respond to diminished self-esteem after suffering from abusive supervision. Based on a theoretical model that examined the effects of abusive supervision on affected employees' workplace behaviours, which was mediated by sense of self-esteem and moderated by intention to leave, they undertook three field studies. The third of these was based on a quantitative diary study, referred to as a 'daily diary design'. For this study, they used 85 participants. These participants were given a link to an initial questionnaire that was used to measure demographic data about each participant, traits related to their self-esteem, their intentions to remain in or leave their employment and scope for alternative employment. The 83 participants who completed the initial questionnaire within one week were then sent daily questionnaires to complete during the last hour of each working day for 21 days. The measures in this daily questionnaire included those to assess, 'daily abusive supervision', 'daily self-esteem', 'turnover intentions' and 'daily workplace deviance'.

In a quantitative diary study published in the *Journal of Organizational Behaviour*, Prem et al. (2017) focused on the within-individual effects of time pressure and learning demands on knowledge workers' scope to thrive at work. In total, 124 participants took part in this 5-day diary study, in which each participant completed questionnaires three times each day: during the morning, afternoon and at the end of the working day. Prior to commencing this five-day diary study, participants completed a general questionnaire. The daily questionnaires were based on abbreviated scales and measured time pressure and learning demands in the morning, appraisal of the challenges and hindrances related to work in the afternoon, and learning and sense of vitality in the final daily questionnaire.

In a quantitative diary study published in the *Academy of Management Journal*, Uy et al. (2017) examined the relationship between surface acting at work, resulting emotional exhaustion and next-day work engagement. They also examined the moderating effects of giving and receiving help at work. Their analytical focus in examining these relationships was at the within-individual level. After completing an online initial, background questionnaire, the 102 participants took part in this 5-day diary study, with each participant completing a questionnaire before commencing work, a second questionnaire at the end of the working day while still in the workplace and a third questionnaire at home before going to bed.

Diarists are encouraged to record activities or events soon after they occur, so their perceptions of their experiences are foremost in their minds. This allows more accurate recall of what they experienced, when and where it occurred, how they were involved and how they feel about it. This is likely to be advantageous when compared to interviews that ask participants to recollect a past activity or event when their perceptions and feelings about it have been lessened by time. In this way, data from diaries should have high internal validity/credibility/authenticity (Section 5.8). Repeatedly collecting data through the completion or creation of multiple diary entries is also likely to produce rich data sets. This should allow frequencies, patterns and themes to be recognised in these data, and discussed in an interview. Diary data are collected unobtrusively since diarists complete or create their own entries without the researcher being present.

Diary entries may be recorded in writing, where participants complete paper-based questionnaires or create hand-written entries, they may be completed online where participants complete questionnaires or word process entries, and in qualitative studies they

may be recorded as audio diaries or video diaries (Whiting et al. 2018; Crozier and Cassell 2016). These means to record diary entries will each have implications for participants. Your choice of method to record diary data will need to consider its suitability for the context of the research and for the circumstances of individual participants. We first look at types of diaries in research, and then explore advantages, issues and strategies associated with their use.

Types of diaries

Like interviews, diaries may be highly structured and formalised, using pre-specified questions and responses, or less structured and informal, where participants are asked to write, type or audio record their responses to pre-determined open questions. The first type is used in a quantitative diary study and the second type is used in a qualitative diary study.

Quantitative diary study

A **quantitative diary study** will be composed of a series of identical, reasonably short questionnaires that are designed to be self-completed by a participant (Section 11.2). This will enable repeated measurements to be obtained from each participant at regular intervals through the course of the study. Participants may be asked to complete a questionnaire every day and to date this, while in some studies they will be asked to complete more than one questionnaire per day, which they will be asked to date and state the time at which it was completed. A quantitative diary study may last from a few days to several weeks, depending on the nature of the research objectives and the purpose of the research (Box 10.22).

Completing a quantitative research diary will demand both time and dedication from those who agree to participate, especially where they need to complete more than one entry (questionnaire) each day. The different media for delivery and return of questionnaires in Section 11.2 (for example, online or by hand) are also relevant for quantitative diary entries and need to facilitate ease of completion and return. Uy et al. (2017) facilitated this by asking their participants to complete paper diary entries, which were issued every day in person by a researcher and once completed, were deposited in a ‘drop box’ in the workplace. In some circumstances this approach will not be appropriate. Biron and Van Veldhoven (2016) produced booklets that contained an initial, longer questionnaire, the diary entries composed of shorter questionnaires and a set of instructions, which were either given out to participants, or posted to them. At the end of the study, these booklets were either collected in person or returned using the postage paid self-addressed envelope provided. In some circumstances, you will be able to ask your participants to access a daily diary online, which they can then complete and return electronically.

The completion of daily questionnaires in a quantitative diary study allows analysis to be conducted on variations between participants’ responses (referred to as a **between persons analysis**), as well as providing the basis for subsequent interviews. Where participants complete more than one daily questionnaire daily, differences such as levels of stress or exhaustion of individuals at different times of each working day can be considered.

Qualitative diary study

A **qualitative diary study** comprises participants writing or typing diary entries, or audio or videorecording their spoken thoughts alongside a date and time. Qualitative diary entries will either be created at regular intervals, usually daily, or alternatively, related to the occurrence of a particular event or activity. These options are referred to as an interval-contingent approach and event-contingent approach, depending on the focus of

the research question and research objectives. Qualitative research diaries are used in longitudinal studies, ranging from a few days to three months, although extended use of this method may be associated with participant fatigue and attrition. A qualitative diary study is also likely to be conducted with a small number of participants, ranging from 4 to 12 (Box 10.23).

Qualitative diary studies are designed to produce accounts of participants' experiences, and their thoughts and feelings related to these (Box 10.23). Their production can be structured by providing participants with a diary template or prompt sheet containing a number of open questions to which you ask them to respond. You may ask diarists to describe:

- what was experienced?
- why did it occur?
- when and where did it occur?
- how were you involved?
- what were the outcomes?
- how do you feel about it?

However, this semi-structured approach may be seen as making assumptions about the nature of participants' experiences and imposing a rational and structured approach to their entries.

Alternatively, you may wish participants to reflect more on their subjective interpretations of their experiences, in which case you may wish to provide them with a less structured template that offers guidance about their purpose in creating diary entries but without a framework of questions to which to respond. In either case, the support that you offer to your diarists will be important in terms of the relevance, quality and quantity of the data that they produce in their research diaries (Day and Thatcher 2009). This is likely to include at least some guidance about creating diary entries to prevent participants dropping out of a diary study because of uncertainty about what to record (Crozier and Cassell 2016) (Box 10.23).

Crozier and Cassell (2016) also recognise that qualitative research diaries are suitable to encourage participants to produce discursive and narrative accounts, which can be used to understand how individuals react to particular events and cope with these (Box 10.23). Qualitative research diary entries can also be used to generate reflective accounts, where a diarist uses an entry to evaluate an earlier experience or set of experiences retrospectively.

While qualitative diary entries may be created unobtrusively by participants, some potential diarists may nevertheless find this process to be difficult or intrusive. Keeping a diary will be unfamiliar to many participants, who may not feel that they have the skills and struggle to express their thoughts and feelings. This sense of struggle may be related to the feeling that what they are being asked to do is intrusive; thoughts and feelings that they normally keep to themselves or only share with significant others will now be recorded and shared with people who they do not know or trust. Some participants may be uncertain about what will happen to the sensitive data which they create. While some will welcome and relish the opportunity to create a qualitative research diary, others may be reticent and self-conscious. You will therefore need to consider how you communicate with intended participants about issues related to informed consent, participant guidance and support, privacy, the avoidance of risk and harm, the voluntary nature of participation, anonymity and confidentiality, responsibility in the way data are handled, analysed and reported, and subsequently managed, if you wish to use this research method (Sections 6.6 and 6.7).



Box 10.23 **Focus on** **management** **research**

Examples of qualitative diary studies

In a study published in *Qualitative Research in Psychology*, Day and Thatcher (2009) asked eight participants to keep hand-written qualitative research diaries. The pages in each diary were headed with a printed date to encourage completion of entries. Other support provided to participants included sending text message reminders to complete entries and the opportunity to discuss any issues about this task with one of the researchers. This study was conducted over a period of three months, and while participants were not expected to complete an entry every day because of the event-based nature of the research, some did by explaining why a particular day's activities were not relevant to the research focus.

In an audio diary study published in the *Journal of Occupational and Organizational Psychology*, Crozier and Cassell (2016) selected six participants to represent a small but diverse sample with regard to age and gender. The choice of this relatively small sample was related to the in-depth, individual level of the analysis, which was designed to focus on within-individual variability. Participants were asked to audio record entries twice a week for four weeks. A message was sent to each participant by mobile phone on the day an audio diary entry was scheduled to be recorded. Participants were provided with a set of instructions for audio recording and a prompt sheet to provide some structure and guidance about what to comment and reflect on in their audio recordings. This list of 10 prompts indicates the semi-structured nature of this approach, although in the spirit of a semi-structured approach (Section 10.2) participants were advised to be flexible in the way they used these prompts depending on their relevance to the situation. The individual recordings that participants produced varied from less than 2 minutes to 12 minutes each. In total, the six participants produced 287 minutes of audio recordings.

Using diaries

Your successful use of a diary study will depend on planning your study very carefully and attempting to anticipate all possible issues. Depending on the quantitative or qualitative nature of your diary study, this will include pilot testing your proposed instructions, guidance notes, questionnaires, template or prompt sheet in a suitable context to evaluate these, and making changes where necessary. You will need to consider your participants and discuss the nature of their participation carefully with them, gain informed consent and provide assurances and information related to ethical, participatory and logistical issues. As a diary study will demand both time and dedication from participants, you will need to discuss the requirements of participation. Establishing informed and realistic expectations before commencement may help to reduce participant attrition. This will include clear expectations about what to include in a qualitative diary entry, how to complete a quantitative diary, the frequency of diary entries, the likely time required to complete or create each entry, the way entries will be recorded and any logistical issues related to this, and the overall duration of the diary study.

Achieving positive outcomes when you conduct a diary study will depend on the instructions, research diary and support provided to participants. The instructions should include a short, clear statement that informs participants about what they need to do, when or how often and how to contact the researcher to ask for advice. In a qualitative diary study the instructions will also include a template or prompt sheet to guide participants as they compose diary entries. Providing participants with a suitable means to complete regular questionnaires or create diary entries will be vital. You will need to

ensure that the means you use, be it online, paper-based or audio/audio-visual recording, is appropriate for your participants and to the setting of your research. An inappropriate or difficult to use diary technique is likely to lead to a poor outcome.

The support you offer your participants in a diary study will also be very important. Contact during the early days of the study will enable you to find out whether participants are experiencing problems in completing diary entries and allow an opportunity to resolve these and deal with any concerns or doubts. Assurances given at this stage may help to avoid participant attrition. You may also send each participant a message by mobile phone on the day a diary entry is scheduled to be recorded. After participants have become used to completing or creating diary entries, you may feel more confident about the conduct of the study. However, there is still a risk that participants may stop completing entries at agreed intervals, or even stop participating. Keeping in contact may help to avoid these possibilities. In a longer diary study lasting several weeks you should consider contacting your participants regularly to check if any issues have arisen. Some participants may relish the task of completing or creating their diary and some of these may not welcome regular checking; you will therefore need to be sensitive to this type of participant, as well as to others who welcome reminders or need reassurance to keep participating.

Depending on the means you are using to conduct your diary study you will also need to consider the return of diary entries or complete diaries at the end of the study. You will need to recognise that participants will have invested a great deal of time and dedication to completing or creating their diaries. They will have become involved in the research project. You will need to consider offering them a debriefing at the time they complete their diaries and later when you have analysed the data and produced your report. An interview with each participant at the time you collect or receive the completed diary, or final diary entry, may be very helpful to you in terms of making sense of the data you have gathered. We summarise these as a checklist for preparing to use diaries (Box 10.24).



Box 10.24 Checklist

Preparing to use diaries

- ✓ develop a short clear statement about the purpose of your diary study and what participants need to do;
- ✓ state how often and when diary entries should be created or completed, related to the use of an interval-contingent or event-contingent approach;
- ✓ state how many diary entries each participant should produce in total and over what period of time;
- ✓ recognise the scope for participant fatigue and attrition, leading to non-completion;
- ✓ consider the logistical difficulties participants may face in accessing, completing and returning individual diary entries or completed diaries;

- ✓ provide a template or prompt sheet to guide participants, taking into account that:
 - too little guidance or support in an unstructured qualitative diary study may lead to uncertainty and adverse consequences for participation, and for the relevance, quality and quantity of data that are produced;
 - provision of a diary template or prompt sheet in a semi-structured qualitative diary study should be sufficient to facilitate relevant, high quality responses without being restrictive;
- ✓ provide an appropriate means for participants to complete and return their diary entries;
- ✓ state how to contact the researcher to ask for advice;
- ✓ decide whether you will interview participants to help understand and interpret their data.

10.13 Summary

- The use of semi-structured and in-depth interviews allows you to collect rich and detailed data, although you will need to develop a sufficient level of competence to conduct these and to be able to gain access to the type of data associated with their use.
- Interviews can be differentiated according to the standardisation and structure of questioning, the mode adopted to conduct them, and medium through which they are operationalised:
 - Standardisation and structure: structured (standardised) or unstructured (semi-structured or in-depth);
 - Mode: one-to-one or group (one-to-many or two-to-many);
 - Medium: online, telephone or face-to-face.
- Semi-structured and in-depth research interviews can be used for exploratory, explanatory and evaluative research purposes.
- The potential for semi-structured and in-depth interviews is dependent upon the purpose of your research, need to establish personal contact, the nature of the questions you wish to ask and the length of time required from participants.
- Your research design may incorporate more than one type of interview.
- It is important to consider data quality issues and evaluate your interviewing practice, including the cleanliness of your questions to assess the credibility of your data.
- Data quality issues related to reliability/dependability, forms of bias, cultural differences and generalisability/transferability may be overcome by considering why you have chosen to use interviews, and careful preparation.
- In preparing for semi-structured and in-depth interviews you need to consider the interview context, its structure, medium, and logistical and time management issues.
- In conducting semi-structured and in-depth interviews you need to consider the appropriateness of your appearance, opening comments when the interview commences, approach to questioning, appropriate use of different types of question, nature of your behaviour during the interview, demonstration of attentive listening skills, scope to summarise and test understanding, ability to deal with difficult participants and how you will record data accurately and fully.
- Apart from one-to-one interviews conducted on a face-to-face basis, you should consider conducting such interviews online or by telephone. In using these media, you need to consider additional aspects that are medium specific alongside many of those considered for face-to-face interviews.
- Group interviews or focus group interviews have advantages associated with participants interacting and discussing your questions. However, they are considerably more difficult to manage than one-to-one interviews.
- Visual images or diaries may be used as a distinctive format interview depending on the purpose of your research.
- Data may also be collected using a quantitative or qualitative diary study.

Self-check questions

Help with these questions is available at the end of the chapter.

10.1 What type of interview would you use in each of the following situations:

- a a market research project?

- b** a research project seeking to understand whether attitudes to working from home have changed?
- c** following the analysis of a questionnaire?
- 10.2** What are the advantages of using semi-structured and in-depth interviews?
- 10.3** During a presentation of your proposal to undertake a research project, which will be based on semi-structured or in-depth interviews, you feel that you have dealt well with the relationship between the purpose of the research and the proposed methodology, when one of the panel leans forward and asks you to discuss the trustworthiness and usefulness of your work for other researchers. This is clearly a challenge to see whether you can defend such an approach. How do you respond?
- 10.4** Having quizzed you about the trustworthiness and usefulness of your work for other researchers, the panel member decides that one more testing question is in order. He explains that interviews are not an easy option. 'It is not an easier alternative for those who want to avoid statistics', he says. 'How can we be sure that you're competent to get involved in interview work, especially where the external credibility of this organisation may be affected by the impression that you create in the field?' How will you respond to this concern?
- 10.5** What are the key issues to consider when planning to use semi-structured or in-depth interviews?
- 10.6** Mark is just about to undertake a semi-structured interview using Zoom. Looking at the screenshot below, what advice would you give him before he admits his participant to the interview?



Source: © 2022 Mark NK Saunders

- 10.7** Which circumstances will suggest the use of visual interviews based on researcher-created images, even where the researcher favours using visual interviews based on participant-created images wherever possible?
- 10.8** You are designing a qualitative diary study but are not sure whether to ask your participants to record their diary entries on paper, word process them or to create an audio diary. You decide to brainstorm the merits of each approach. What points might be included in this consideration?

Review and discussion questions

- 10.9** Watch and, if possible, record a television interview such as one that is part of a chat show or a documentary. It does not matter if you record an interview of only 10 to 15 minutes' duration.
- a** As you watch the interview, make notes about what the participant is telling the interviewer. After the interview, review your notes. How much of what was being said did you manage to record?
 - b** If you were able to record the television interview, watch it again and compare your notes with what was actually said. What other information would you like to add to your notes?
 - c** Either watch the interview again or another television interview that is part of a chat show or a documentary. This time pay careful attention to the questioning techniques used by the interviewer. How many of the different types of question discussed in Section 10.7 can you identify?
 - d** How important are the visual cues and verbal and paralinguistic signals given by the interviewer and the interviewee in understanding the meaning of what is being said?
- 10.10** With a friend, each decide on a topic about which you think it would be interesting to interview the other person. Separately develop your interview themes and prepare an interview guide for a semi-structured interview. At the same time, decide which one of the 'difficult' participants in Table 10.5 you would like to role-play when being interviewed.
- a** If possible, conduct and audio record both interviews using a cloud-based video-conferencing platform. If this is not possible either audio record or ensure the interviewer takes notes.
 - b** Watch each of the recordings – what aspects of your interviewing technique do you each need to improve?
 - c** If you were not able to record the interview, how good a record of each interview do you consider the notes to be? How could you improve your interviewing technique further?
 - d** As an interviewer, ask your friend an open question about the topic. As your friend answers the question, note down their answer. Summarise this answer back to your friend. Then ask your friend to assess whether you have summarised their answer accurately and understood what s/he meant.
- 10.11** Obtain a transcript of an interview that has already been undertaken. If your university subscribes to online newspapers such as ft.com, these are a good source of business-related transcripts. Alternatively, typing 'interview transcript' into a search engine such as Google or Bing will generate numerous possibilities on a vast range of topics!
- a** Examine the transcript, paying careful attention to the questioning techniques used by the interviewer. To what extent do you think that certain questions have led the interviewee to certain answers?
 - b** Now look at the responses given by the interviewer. To what extent do you think these are the actual verbatim responses given by the interviewee? Why do you think this?



Progressing your research project

Using research interviews

- Assess whether interviews will help you to answer your research question and address your objectives. Where you do not think that they will be helpful, justify your decision. Where you think that they will be helpful, respond to the following points.
- Which structure(s), mode(s) and medium(s) of interview will be appropriate to use? Explain how you intend to use them and how they will fit into your chosen research strategy.
- Draft a topic focus to explore during in-depth interviews or list of themes to use in the conduct of semi-structured or in-depth interviews and use your research question(s) and objectives to assess this or these.
- What threats to the trustworthiness of the interview data you collect are you likely to encounter? How will you seek to overcome these?
- What practical problems do you foresee in conducting your interviews? How will you attempt to overcome these practical problems?
- Ask your project tutor to comment on your judgement about using interviews, the relationship between these and your proposed research

strategy, the fit between your topic focus or interview themes and your research question(s) and objectives, the issues and threats that you have identified, and your suggestions to overcome these.

- Use the questions in Box 1.4 to guide your reflective diary entry.

Using research diaries

- Assess whether the use of a research diary study will help you to answer your research question and address your objectives. Where you do not think that this will be helpful, justify your decision. Where you think that this will be helpful, respond to the following points.
- Which research strategy or strategies do you propose to use? What will be the implications of this strategy or strategies for the type of diary study that you use and the way in which you will analyse your data?
- Which issues are likely to arise in relation to using this diary study? Which strategies will you use to anticipate and seek to overcome these?
- Ask your project tutor to comment on your judgement about using a diary study and its relationship to your proposed research strategy, the issues you have identified that may affect its conduct and your strategies to anticipate and seek to overcome these.
- Use the questions in Box 1.4 to guide your reflective diary entry.

References

- BBC Academy (n.d.) *Interviewing: Asking the Right Questions*. Available at <https://downloads.bbc.co.uk/academy/academyfiles/TVNewsInterviewing/Interviewing%20-%20Asking%20the%20right%20questions.pdf> [Accessed 18 Dec. 2021].
- Belzile, J.A. and Oberg, G. (2012) 'Where to begin? Grappling with how to use participant interaction in focus group design', *Qualitative Research*, Vol. 12, No. 4, pp. 459–72.
- Biron, M. and Van Veldhoven, M. (2016) 'When control becomes a liability rather than an asset: Comparing home and office days among part-time teleworkers', *Journal of Organizational Behaviour*, Vol. 37, pp. 1317–37.
- Body, C. (2005) 'A rose by any other name may smell as sweet but "group discussion" is not another name for "focus group" nor should it be', *Qualitative Market Research*, Vol. 8, No. 3, pp. 248–55.

- Brinkmann, S. and Kvale, S. (2015) *InterViews: Learning the Craft of Qualitative Research Interviewing* (3rd edn). London: Sage.
- Cairns-Lee, H. (2017) *An exploration of leadership and its development through the inner worlds of leaders using metaphor*. Unpublished PhD thesis, Guildford: University of Surrey
- Cairns-Lee, H., Lawley, J. and Tosey, P. (2021, online first) 'Enhancing researcher reflexivity about the influence of leading questions in interviews', *The Journal of Applied Behavioral Science*.
- Carter, S.M., Shih, O.P., Williams, J., Degeling, C. and Mooney-Somers, J. (2021) 'Conducting qualitative research online: challenges and solutions', *The Patient*, Vol. 14, pp. 711–18.
- Cassell, C., (2015). *Conducting research interviews for business and management students*. London: Sage.
- Chidlow, A., Plakoyiannaki, E. and Welch, C. (2014) 'Translation in cross-language international business research: Beyond equivalence', *Journal of International Business Studies*, Vol. 45, pp. 562–82.
- Court, D. and Abbas, R. (2013) 'Whose interview is it, anyway? Methodological and ethical challenges of insider-outsider research, multiple languages, and dual-researcher cooperation', *Qualitative Inquiry*, Vol. 19, No. 6, pp. 480–8.
- Crozier, S.E. and Cassell, C.M. (2016) 'Methodological considerations in the use of audio diaries in work psychology: Adding to the qualitative toolkit', *Journal of Occupational and Organizational Psychology*, Vol. 89, No. 2, pp. 396–419.
- Day, M and Thatcher, J. (2009) ‘“I’m really embarrassed that you’re going to read this . . . ”: Reflections on using diaries in qualitative research’, *Qualitative Research in Psychology*, Vol. 6, No. 4, pp. 249–59.
- Denzin, N.K. (2001) 'The reflexive interview and a performative social science', *Qualitative Research*, Vol. 1, No. 1, pp. 23–46.
- Dick, B. (2013) 'Convergent interviewing [Online]'. Available at <http://www.aral.com.au/resources/coin.pdf> [Accessed 5 February 2022].
- Dodds, S. and Hess, A.C. (2020, online first) 'Adapting research methodology during COVID-19: lessons for transformative service research', *Journal of Service Management*.
- Gobo, G. (2011) 'Glocalizing methodology? The encounter between local methodologies', *International Journal of Social Research Methodology*, Vol. 14, No. 6, pp. 417–37.
- Heisley, D.D. and Levy, S.J. (1991) 'Autodriving: A photoelicitation technique', *Journal of Consumer Research*, Vol. 18, No. 4, pp. 257–72.
- Heyl, B.S. (2005) 'Ethnographic interviewing', in P. Atkinson, A. Coffey, S. Delamont, J. Lofland and L. Lofland (eds) *Handbook of Ethnography*. Thousand Oaks, CA: Sage, pp. 369–83.
- Hibbert, P., Sillince, J., Diefenbach, T. and Cunliffe, A. (2014) 'Relationally reflexive practice: A generative approach to theory development in qualitative research', *Organizational Research Methods*, Vol. 17, pp. 278–98
- Holt, A. (2010) 'Using the telephone for narrative interviewing: A research note', *Qualitative Research*, Vol. 10, No. 1, pp. 113–21.
- Irvine, A. (2011) 'Duration, dominance and depth in telephone and face-to-face interviews: A comparative exploration', *International Journal of Qualitative Methods*, Vol. 10, No. 3, pp. 202–20.
- Irvine, A., Drew, P. and Sainsbury, R. (2012) ‘“Am I not answering your questions properly?” Clarification, adequacy and responsiveness in semi-structured telephone and face-to-face interviews’, *Qualitative Research*, Vol. 13, No. 1, pp. 87–106.
- Keaveney, S.M. (1995) 'Customer switching behaviour in service industries: An exploratory study', *Journal of Marketing*, Vol. 59, No. 2, pp. 71–82.

- King, N. (2004) 'Using interviews in qualitative research', in C. Cassell and G. Symon (eds) *Essential Guide to Qualitative Methods in Organizational Research*. London: Sage, pp. 11–22.
- Krueger, R.A. and Casey, M.A. (2015) *Focus Groups: A Practical Guide for Applied Research* (5th edn). London: Sage.
- Lobe, B., Morgan, D. and Hoffman (2020) 'Qualitative data collection in an era of social distancing', *International Journal of Qualitative Methods*, Vol. 19, No. 1, pp. 1–8.
- McNaughton, P. and Myers, G. (2007) 'Focus groups', in C. Seale, G. Gobo, J.F. Gubrium and D. Silverman (eds) *Qualitative Research Practice*. London: Sage, pp. 65–79.
- Ozanne, J.L., Moscato, E.M., Kunkel, D.R. (2013) 'Transformative photography: Evaluation and best practices for eliciting social and policy changes', *Journal of Public Policy and Marketing*, Vol. 32, No. 1, pp. 45–65.
- Pearce, G., Thøgersen-Ntoumani, C. and Duda, J.L. (2014) 'The development of synchronous text-based instant messaging as an online interviewing tool', *International Journal of Social Research Methodology*, Vol. 17, No. 6, pp. 677–92.
- Powley, J. and Watts, M. (1987) *Interviewing in Educational Research*. London: Routledge & Kegan Paul.
- Prem, R., Ohly, S., Kubicek, B. and Korunka, C. (2017) 'Thriving on challenge stressors? Exploring time pressure and learning demands as antecedents of thriving at work', *Journal of Organizational Behaviour*, Vol. 38, No. 1, pp. 108–23.
- Reissner, S.C. (2018) 'Interactional challenges and researcher reflexivity: Mapping and analysing conversational space', *European Management Review*, Vol. 15, No. 2, pp. 205–19.
- Reissner, S. and Whittle, A. (2021, online first) 'Interview-based research in management and organisation studies: Making sense of the plurality of methodological practices and presentational styles', *Qualitative Research in Organizations and Management: An International Journal*.
- Speer, S.A. (2008) 'Natural and contrived data', in P. Alasuutari, L. Bickman, and J. Brannen (eds) *The Sage Handbook of Social Research Methods*. London: Sage, pp. 290–312.
- Stokes, D. and Bergin, R. (2006) 'Methodology or "methodolatry"? An evaluation of focus groups and depth interviews', *Qualitative Market Research*, Vol. 9, No. 1, pp. 26–37.
- Teti, M., Schatz, E. and Liebenberg, L. (2020) 'Methods in the time of COVID-19: the vital role of qualitative inquiries', *International Journal of Qualitative Methods*, Vol. 19, pp. 1–5.
- Trier-Bieniek, A. (2012) 'Framing the telephone interview as a participant-centred tool for qualitative research: A methodological discussion', *Qualitative Research*, Vol. 12, No. 6, pp. 630–44.
- Uy, M.A., Lin, K.J. and Ilies, R. (2017) 'Is it better to give or receive? The role of help in buffering the depleting effects of surface acting', *Academy of Management Journal*, Vol. 60, No. 4, pp. 1442–61.
- Vermaak, M. and de Klerk, H.M. (2017) 'Fitting room or selling room? Millennial female consumers' dressing room experiences', *International Journal of Consumer Studies*, Vol. 41, pp. 11–18.
- Vogel, R.M. and Mitchell, M.S. (2017) 'The motivational effects of diminished self-esteem for employees who experience abusive supervision', *Journal of Management*, Vol. 43, No. 7, pp. 2218–51.
- Vogl, S. (2013) 'Telephone versus face-to-face interviews: Mode effect on semi-structured interviews with children', *Sociological Methodology*, Vol. 43, No. 1, pp. 133–77.
- Whiting, R., Symon, G., Roby, H. and Chamakiotis, P. (2018) 'Who's behind the lens?: A reflexive analysis of roles in participatory video research', *Organizational Research Methods*, 21 (2): 316–40.

Further reading

- BBC Academy (n.d.) *Interviewing: Preparing for the interview*. Available at <https://downloads.bbc.co.uk/academy/academyfiles/TVNewsInterviewing/Interviewing%20-%20Preparing%20for%20an%20interview.pdf> [Accessed 18 Dec. 2021]. This brief guide is part of a set which also includes *Asking the right questions*. Both guides have hyperlinks to informative videos.
- Brinkmann, S. and Kvale, S. (2015) *InterViews: Learning the Craft of Qualitative Research Interviewing* (3rd edn). London: Sage. This provides a useful general guide to interviewing skills.
- Court, D. and Abbas, R. (2013) 'Whose interview is it, anyway? Methodological and ethical challenges of insider-outsider research, multiple languages, and dual-researcher cooperation', *Qualitative Inquiry*, Vol. 19, No. 6, pp. 480–8. This is a helpful account to understand how cultural differences may impact on the scope to collect data and the implications of operating as either a cultural insider or outsider.
- Krueger, R.A. and Casey, M.A. (2014) *Focus Groups: A Practical Guide for Applied Research* (5th edn). Thousand Oaks, CA: Sage. This provides a useful source for those considering the use of this method of group interviewing.
- Symon, G. and Cassell, C. (eds) (2012) *Qualitative Organizational Research: Core Methods and Current Challenges*. London: Sage. This edited work contains a helpful range of contributions related to qualitative data collection including interviews and focus groups.

Analyzing Data Qualitatively

Saunders et al. (2023; 650-712)

Chapter 13



Analysing data qualitatively

Learning outcomes

By the end of this chapter, you should be able to:

- discuss the diversity of qualitative data and the interactive nature of qualitative analysis;
- identify the key aspects to consider when deciding how to analyse data qualitatively;
- describe the main issues when preparing your qualitative data for analysis including using Computer-Aided Qualitative Data Analysis Software (CAQDAS);
- transcribe a recorded interview or notes of an interview or observation and create a data file for analysis by computer;
- choose from different analytical aids to help you to analyse your qualitative data;
- outline Thematic Analysis and Template Analysis as approaches to analysing qualitative data and differentiate between them;
- differentiate between Analytic Induction, deductive explanation building and pattern matching processes;
- explain how Grounded Theory Method can be used;
- discuss different approaches to Narrative Analysis and discourse analysis;
- recall the processes used in visual data analysis;
- identify the common functions of CAQDAS and the issues associated with its use;
- progress your research project by analysing data qualitatively.



13.1 Introduction

This chapter is about analysing qualitative data, that is both primary and secondary data that are derived from spoken, written or printed words and still or moving visual images that have not been quantified. The diversity of such qualitative data and their implications for interactive analyses are discussed in Section 13.2. As you read this chapter you will rediscover the interconnected and interactive nature of qualitative data collection and analysis, and the need to begin to analyse and interpret your data during the collection process.

In Section 13.3 we discuss key aspects of different qualitative analysis techniques to help you to choose an appropriate way of analysing your data. In Section 13.4 we discuss the preparation of your data for analysis, and in Section 13.5 we outline several aids that will help you analyse these data and record your ideas about how to progress your research.

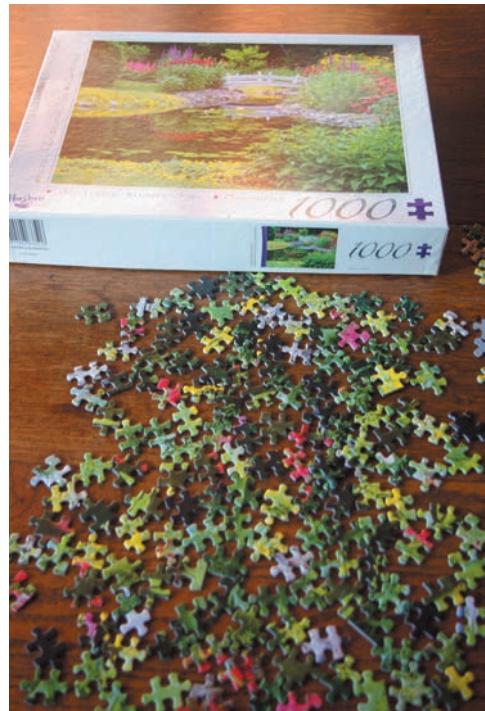
Sections 13.6 to 13.13 outline a range of qualitative analysis techniques, methods, approaches and processes. These are grouped under Thematic Analysis (Section 13.6), Template Analysis (Section 13.7), explanation building and testing (Section 13.8), Grounded Theory Method

Qualitative data analysis and completing a jigsaw puzzle

Nearly all of us have, at some time in our lives, completed a jigsaw puzzle. As children we may have played with jigsaw puzzles and, as we grew older, those we were able to complete became more complex. Qualitative data analysis can be likened to the process of completing a jigsaw puzzle in which the pieces represent data. These pieces of data and the relationships between them help us as researchers to create a picture of what we think the data are telling us!

When trying to complete a jigsaw puzzle, most of us begin by looking at the picture on the lid of our puzzle's box. A puzzle for which there is no picture is usually more challenging as we have no idea how the pieces fit together or what the picture will be! Similarly, we may not be clear about how, or even if, the data we have collected can form a clear picture.

Perhaps you haven't tried to complete a jigsaw puzzle for many years, but you might find the following useful as well as entertaining! Get a friend to give you the contents of a jigsaw in a bag without the box



Source: © Mark Saunders 2018



(since this normally shows the picture of what it is!). Turn all the pieces picture side up. Think about how you will categorise the pieces. What do they mean? You will be likely to group pieces with similar features such as those of a particular colour together. Some pieces can be placed in more than one category such as blue and also green. These pieces are crucial as they give clues to the relationship between pieces. Normally

you might then try to fit these similar pieces together to begin to reveal the picture.

Perhaps completing jigsaws reinforces a sense of there being an external reality ‘out there’, so all we need to do is reveal it! However, for many qualitative researchers, the picture that our pieces of data reveal will depend on our research question and the concepts we use to make sense of what we see!

(Section 13.9), Narrative Analysis (Section 13.10), discourse analysis (Section 13.11), visual analysis (Section 13.12) and data display and analysis (Section 13.13). The use of quotations when reporting the findings of your analysis in your research report is considered in Section 14.3.

A variety of CAQDAS is available, which we consider briefly in Section 13.14. Consequently, although you need to understand which analytical technique, or combination of techniques, is suitable for analysing your data, it may no longer be necessary for you to undertake tasks such as coding your data and developing analytical categories manually. Within this chapter we do not assume you will be using CAQDAS. Although a variety of software packages are available in most universities, and these are helpful for managing and organising data, to support your analysis, they are not analysis programs and make no analytical contribution to your research (Woolf and Silver 2018). Rather the actual process remains thoughtful, reflective and reflexive rather than mechanical. Consequently, using CAQDAS isn’t a quick fix, and you still need to perform the associated processes. We therefore focus on these analysis processes, referring to and including screenshots of the NVivo software in some worked examples. These illustrate generic issues associated with analysis rather than implying that you must use such software.

13.2 Analysing qualitative data, diversity and interactive processes

Diversity and analytical implications

Unlike quantitative research where analysis occurs after data collection, qualitative research often involves the concurrent collection, analysis and interpretation of data. These data are diverse comprising spoken words (verbal data), written, typed or printed words (textual data) and still or moving visual images (visual data). **Verbal data** are collected in the form of extended speech, which are passages of spoken words. These may be audio recorded or derived from existing audio or audio-visual sources. These data are likely to be transcribed and turned into text but may still be classified as verbal data if they maintain their structural integrity as a verbatim account (Chapters 8, 9 and 10). **Text data** are collected as notes from interviews or observations, as written diaries and participant accounts (Chapters 9 and 10) or derived from documents including reports, tweets, emails, blogs and the like (Chapter 8). **Visual data** may be created or found in many forms including drawings, digital images and video (Chapters 9 and 10). These data are associated with particular analytical implications, which we consider in Section 13.12.

This diversity of qualitative data arises from the variety of procedures used to obtain it. Many of these involve the collection of data in natural settings, where the researcher observes or interview participants, or asks participants to collect data themselves through audio or video recording, photography, or keeping a research diary. Such naturalistically collected data are often contrasted with contrived data collected through laboratory-based experiments, or questionnaires that do not consider the context within which these are used.

Qualitative data collected in natural settings are likely to be rich in contextual detail. The opportunity to explore issues in interviews, record mundane details during observations or read through participants' detailed accounts in research diaries is likely to produce descriptive and explanatory data that help to facilitate analysis and interpretation. Much of these data will come directly from participants, by recording their words during interviews, detailing their actions during observations, using their written words from diaries, transcribing the audio recordings they create, watching and making notes about the video recordings they make, or looking at the visual images they provide. Through these, you can give participants a 'voice' through which to talk about and record their experiences and perceptions.

This production of highly contextualised data, emphasis on recording participants' interpretations and practice of using participants to collect data each have implications for qualitative analysis. Qualitative data, characterised by their fullness and richness, provide an opportunity for in-depth analysis, where context can be related to the themes that emerge from analysis, to produce well-grounded and contextualised explanations. In this way, a contrast is drawn between the 'thin' abstraction or description that principally results from quantitative data and the 'thick' or 'thorough' abstraction or description associated with qualitative data (Dey 1993).

The philosophical assumptions underpinning a research project will affect its design and conduct, including data collection and analysis. An interpretivist philosophy often informs qualitative research projects (Section 5.3), emphasising participants' interpretations of their social world, reality being seen as being socially constructed (Sections 4.2 and 4.4). As an interpretivist researcher you would, typically, undertake research inductively (Section 4.5), allowing the conduct of the research to follow the flow of the data collected. These data will reflect variations in participants' experiences and perspectives and your analysis will need to recognise and report the breadth of these experiences and perspectives, rather than attempting to reconcile differences and ignore this diversity of viewpoints. Analysis of data collected through an interpretivist approach therefore needs to be sensitive to their variability and complexity to be meaningful. It will involve deriving research-specific concepts from which a conceptual framework may be developed. This framework will be developed initially during data collection and then refined as your analysis progresses. Various techniques to develop this analytical approach are discussed in Sections 13.6 to 13.13.

A realist philosophy can also inform qualitative research projects (Section 5.3). A realist researcher believes reality exists independently from participants' interpretations of it. In other words, an external 'reality' helps to shape participants' interpretations rather than being constructed by them (Sections 4.2 and 4.4). As a realist, critical realist or pragmatist you would use your perspective to inform the design and conduct of your qualitative research. You would, typically, undertake research deductively (Section 4.5), your themes being derived from existing theory. These themes would inform the questions you asked participants, which would need to be asked or applied consistently to produce comparable and valid data to test the applicability of the theory in this research context. Various techniques to develop this analytical approach are discussed in Sections 13.6 to 13.13 excluding 13.9.

Irrespective of your philosophy, the meanings from your data will principally be derived from words and images, not numbers. As these may have multiple as well as unclear meanings, it is necessary to explore and clarify them with great care. Consequently, the quality of your qualitative research depends partly on the interaction between your data collection and data analysis and allowing meanings to be explored and clarified.

Qualitative data collected by participants will also have analytical implications. These data will reflect the experiences and perspectives of those who collect them. They will be characterised by specific meanings that you will need to understand in order to interpret. This may mean using multiple methods during data collection and analysis, such as using a multi-method qualitative study (Section 5.4), where you conduct interviews (Chapter 10) in conjunction with some form of observation (Chapter 9) to explore meanings and produce participant-focused interpretations.

The non-standardised and complex nature of qualitative data has further implications for their analysis. You will be confronted by either a mass of paper, still images, audio and visual recordings or electronic files that you will need to explore, analyse, synthesise and transform to address your research objectives and answer your research question. Most of the analytical techniques discussed later in this chapter will involve you using processes where you summarise some parts of your data to condense them; code and categorise data to group them according to themes that begin to make sense; and then to link these categories and themes in ways that provide you with a structure or structures to answer your research question. Without using such techniques, the most that may result is an impressionistic view of what these qualitative data mean.

Interactive nature

Qualitative analysis is undertaken during and after data collection; the analysis process helping to shape the direction of data collection, especially when following an inductive or grounded approach. Research propositions that emerge from your data in an inductive approach or those you commenced with at the start of your data collection in a deductive approach will be tested as you compare them with the data in your study. The key point here is the relative flexibility that this type of process permits.

The interactive nature of data collection and analysis allows you to recognise important themes, patterns and relationships during data collection: in other words, for these to emerge from the process of data collection and analysis. This means you are likely to need to re-categorise and re-code your existing data to see whether emergent themes, patterns and relationships are present in the cases where you have already collected data. You will also be able to adjust your future data collection to establish whether related data exists in those cases where you intend to conduct your research.

This concurrent process of data collection and analysis has implications for the way in which you manage your time and organise your data and related documentation. It will be necessary to arrange interviews or observations with enough space between them to allow sufficient time to write up or word process a transcript or set of notes, and to undertake a preliminary analysis before proceeding to the next (Section 10.6). You may also be able to find a little time between interviews to carry out a cursory level of analysis. As part of this we have found it extremely helpful to listen to audio recordings of interviews while travelling to and from the university.

There is a clear limit to the value of continuing to undertake interviews or observations without analysing these. There is also a danger of data overload where you just continue to collect data. This will be associated with a lost opportunity to understand what your data reveal in relation to your research question and the directions that might be worth pursuing subsequently as your research progresses. Important ideas that occur to you as

you undertake an interview, conduct an observation, read a document, listen to an audio recording, or view a set of images or a visual-recording may be lost if you do not record these because you are focused only on collecting data.

13.3 Choosing a qualitative analysis technique

Choosing a qualitative analysis technique can be confusing. Choice in qualitative analysis is different to choice in quantitative analysis. Quantitative analysis necessitates specified statistical techniques dependent on the data type and what you are trying to illustrate, describe, examine or predict (Sections 12.5 to 12.13). Choice in qualitative analysis is not necessarily between a ‘right’ and ‘wrong’ technique. Some forms of qualitative analysis are not exclusive; in other words, you may have to choose between alternative ways to analyse your qualitative data. This may also mean using two or more complementary techniques to analyse your data, thereby gaining more insights than would be possible from a single technique. You therefore need to understand the nature of different techniques to be able to choose those which offer the possibility of complementary insights.

Aspects that are particularly important in your choice of qualitative analysis techniques relate to:

- the methodological and philosophical basis;
- the approach to theory development;
- the analytical techniques used.

We discuss each of these in turn.

Methodological and philosophical basis

Some qualitative research strategies are associated with a specific or prescriptive methodology. In these the research philosophy, approach to theory development and research practices including analytical techniques are closely defined. Of the research strategies we consider (Section 5.5), Grounded Theory has a specific methodology. To use the Grounded Theory Method you would need to follow each of the elements associated with this approach (Section 13.9). While the specific or prescriptive nature of this type of methodological approach might be considered rigid, it provides you with a clear set of guidelines for your entire research project including the analytical technique.

Other qualitative research strategies are neither so closely specified nor prescriptive. This means you will need to choose an analytical technique that is appropriate for your research philosophy, strategy (Sections 4.4 and 5.8) and the nature of the data you collect. Where, for example, you use an interpretivist philosophy, it will be important to ensure that your choice of analytical technique(s) is compatible with this research philosophy. Here you would need to allow the voices of your participants to emerge through your analysis, and probably include participant quotations in your findings. An analytical technique concentrating on condensing participants’ data to display them in a highly reduced and summarised form would be unlikely to be suitable for interpretivist research.

Approach to theory development

Theory is developed using either deductive, inductive or abductive reasoning (Section 4.5). Where you commence your research project using a deductive approach you will use existing theory to inform your data collection and analysis. Where you commence your research

project using an inductive approach you will seek to build a theory that is grounded in your data. Subsequently if, based on a surprising fact, you collected additional data, and to potentially revise or modify an existing theory you would be using an abductive approach. Some qualitative analysis techniques we discuss in Sections 13.6 to 13.13 are specifically associated with either a deductive or inductive approach, while others can be used regardless of your approach to theory development.

Analytical techniques used

Analytical techniques can be differentiated by their:

- use of data fragmentation and reduction versus maintaining data integrity;
- analytical focus.

We discuss each of these in turn.

Data fragmentation and reduction versus maintaining data integrity

In qualitative data analysis, it is generally accepted that to analyse large amounts of non-standardised data it is necessary to fragment these data by coding and reorganising them into analytical categories. This process often involves simplifying or reducing qualitative data by summarising their meanings to be able to comprehend them and undertake further analysis. Such techniques include Thematic Analysis (Section 13.6), Template Analysis (Section 13.7), Grounded Theory Method (Section 13.9) and data display and analysis (Section 13.13).

Where it is important to maintain the integrity of the data by analysing them without using fragmentation and reduction, other alternative techniques can be used. These include Narrative Analysis, where the sequential and chronological nature of storied data are essential to and maintained during analysis (Section 13.10), and discourse analysis (Section 13.11), where analysis relies on the wholeness of data.

Analytical focus

The focus of qualitative analysis techniques can be:

- thematic;
- actions or processes;
- the use of language.

Thematic analysis is seen by some as a generic approach rather than as a specific technique as it is used in various analytical techniques. In practice, there are a number of variants, which can be used as standalone analytical techniques. We refer to one of these as Thematic Analysis, deliberately using capital letters to distinguish it from other variants (Section 13.6). Even if you use another variant we advise you to read this section on Thematic Analysis carefully as it contains helpful insights whichever approach you use. Further standalone variants of thematic analysis include Template Analysis (outlined in Section 13.7) and data display and analysis (considered in Section 13.13). Thematic analysis is also used in some approaches to the Grounded Theory Method (Section 13.9), in Narrative Analysis (Section 13.10), and may be used in analytical induction (Section 13.8), deductive explanation building (Section 13.8) and visual analysis (Section 13.12).

While some approaches in the Grounded Theory Method code data thematically, Charmaz (2014) advocates coding data for actions in order to stay close to meanings in the data and to understand these through the actions or interactions that take place (Section 13.9).



Box 13.1 Checklist

To help you to choose a qualitative analysis technique, or combination of techniques

- ✓ Is your analytical technique linked to a specific or prescriptive methodology? (In this chapter the Grounded Theory Method is linked to such a methodology; all other analytical techniques discussed are not.)
- ✓ Is your analytical technique appropriate for your research strategy and the underpinning research philosophy? (Your research philosophy has implications for both these.)
- ✓ Is your analytical technique(s) appropriate for your approach to theory development? (Some qualitative analysis techniques are associated specifically with either a deductive or inductive approach, while others may be used more flexibly.)
- ✓ Will it be beneficial to fragment your data during analysis or alternatively to maintain the integrity of your data items? (Most qualitative analysis techniques involve fragmentation and reorganisation of data and sometimes their reduction, but some maintain the original form of the data during analysis.)
- ✓ What is the most appropriate analytical focus for foci for your data? (The focus of analysis varies between techniques, most focusing on analysing themes, some focussing on actions or processes and others the use of language.)

Other qualitative analysis techniques consider the use of language in the data. These focus on structural elements to understand the implications of how language is used or how narratives are constructed. We introduce two analytical approaches that focus on the use of language: structural narrative analysis (Section 13.10) and discourse analysis (Section 13.11).

We summarise these aspects as a checklist to help you to choose an appropriate technique, or combination of techniques (Box 13.1).

13.4 Preparing data for analysis

Despite the varied forms of qualitative data, most analyses focus on verbal or text data being converted into word-processed text. Even when we are analysing visual data, our analyses and associated interpretations make considerable use of words. Consequently, to undertake most analyses, you will need to convert your data to word-processed text, check it for accuracy and typographical errors, and plan ahead for your analysis, especially if using CAQDAS software.

Within this, it is important to emphasise the need to make back-up copies of your recordings, transcriptions of both these and your notes to ensure your data are not lost. In this section we focus upon the conversion of qualitative data from oral or handwritten form to word-processed text, as this is the way that you are most likely to use these in your analysis. As part of this, we discuss the general requirements of CAQDAS packages (see Section 13.14).

Transcribing qualitative data

In Chapter 10 we emphasised that, in qualitative research interviews, the interview is often audio recorded and subsequently **transcribed**, that is, reproduced verbatim as a

Table 13.1 Reducing the time needed to transcribe audio recordings

| Way | Potential problems |
|---|--|
| Pay a touch-typist to transcribe your audio recordings | <ul style="list-style-type: none"> Expense of paying someone else Important data such as pauses, coughs, sighs and the like may not be included Lack of familiarity with the data as you are not transcribing them yourself Transcription will still require careful checking as errors can creep in |
| Use a transcription machine with a foot-operated play–pause–rewind–fast forward mechanism and software to control the audio speed | <ul style="list-style-type: none"> Although this will allow you to control the audio recorder more easily, the speed of transcription will still be dependent upon your typing ability Transcription will still require careful checking May not be able to gain access to a transcription machine |
| 'Dictate' your audio recordings using voice-recognition software | <ul style="list-style-type: none"> Need to teach voice-recognition software to recognise your voice Need to listen to and dictate the entire audio recording Transcription will still require careful checking as the software is not entirely accurate |
| Only transcribe sections of each audio recording that are pertinent to your research (data sampling) | <ul style="list-style-type: none"> Still need to listen to the entire recording carefully first, at least twice May miss certain things, meaning you will have to go back to the audio recording later Sections you transcribe will still require careful checking |

word-processed account. We also emphasised that, as an interviewer, you would be interested not only in what participants said, but in the way they said it as well. This means that the task of transcribing audio recorded interviews is likely to be time-consuming as you will need not only to record exactly what was said and by whom, but also try to give an indication of the tone in which it was said and the participants' non-verbal communication using transcription notation (Table 13.2). Without this additional contextual information, important incidents that affect the conduct of your interview or observation may be missed (Boxes 10.12 and 13.4). You also need to ensure it can be linked to the contextual information that locates the interview (Section 10.7).

Even if you are a touch-typist, you will find transcribing an audio recording extremely time-consuming. Most research methods texts suggest that it takes a touch-typist between 6 and 10 hours to transcribe every hour of audio recording. Consequently, it is helpful if your interviews are transcribed as soon as possible after they are undertaken to avoid a build-up of audio recordings and associated transcription work. Fortunately, there are several possible ways of reducing the vast amount of personal time needed to transcribe interviews verbatim. These are summarised in Table 13.1 along with some of the potential problems. One aspect, however you choose to transcribe the data, is making sure that the transcription is accurate by correcting any transcription errors. This process is known as **data cleaning**. Once this has been done, some researchers send a copy of the transcript to the participant for final checking. While this can be helpful for ensuring factual accuracy, we have found that interviewees often want to correct their own grammar and use of language as well! This is because spoken and written language are very different. You therefore need to think carefully before offering to provide a copy of a complete transcript to an interviewee.

Table 13.2 Commonly used Jefferson transcription symbols

| Symbol | Meaning |
|----------|--|
| (.) | Notable pause, but not a significant length |
| (0.3) | Timed pause, number denotes time in seconds |
| [] | Onset and off set of overlapping utterances |
| > < | Pace of speech quicker within brackets |
| < > | Pace of utterance slower within brackets |
| () | Something is being said, but words spoken too unclear to transcribe |
| (word) | Uncertainty of what was said, but a likely possibility |
| (()) | Transcriber's description of what is happening, e.g. ((coughs)) |
| _____ | Underlining of word or words denotes an increase in volume or emphasis |
| CAPITALS | Using capitals denotes shouting |
| ↑ | Rise in intonation |
| ↓ | Fall in intonation |
| :: | Prolongation of the previous sound |
| (h) | Laughter (humour) within the utterance |

Sources: Developed from Silverman (2013), University of Leicester (n.d.)

When transcribing interviews and group interviews, you need to be able to distinguish between the interviewer and the participant or participants. This means you need to have clear speaker identifiers such as '17FA' for the 17th interviewee who is a female administrator. This tends to be more visible in the transcript if they are in capitals (Box 13.2). Similarly, you need to be able to distinguish between topic headings you use, questions and responses. One way of doing this, dependent upon the precise requirements of your CAQDAS, is to put topic headings in CAPITALS, questions in *italics* and responses in normal font. The most important thing is to be consistent within and across all your transcriptions. Some authors also recommend the use of specific transcription symbols to record intakes of breath, overlapping talk and changes in intonation, a widely used form being the Jefferson system that captures both what was said, and the way in which it is said (Table 13.2). This helps you derive meanings from the words and the way they are uttered when placed within an interactional sequence (Reissner and Whittle 2021).

In a transcription of a more structured interview, you also need to include the question number and the question in your transcription. For example, by including the question number 'Q27' at the start of the question you will be able to search for and find question 27 quickly. In addition, by having the full question in your transcript you will be far less likely to misinterpret the question your respondent is answering.



Box 13.2 **Focus on student research**

Extract from an interview transcript

Michael had decided to use the code IV to represent himself in the transcripts of his in-depth interviews and 01FS to represent his first interviewee, a female student. By using capital letters to identify both himself and the interviewee Michael could identify clearly where questions and responses started. In addition, it reduced the chance of a mistype in the transcription as identifiers were always a combination of capital letters and numbers. Michael used

transcription symbols such as '(.)' to represent a brief pause and (h) to represent a laugh. He also included brief comments relating to a respondent's actions in the interview transcript. These he enclosed with double parentheses (()). A brief extract from a transcript follows:

IV: So tell me, why do you use the Student Union Bar?

01FS: Well,((in-breath)) (..), a lot of my friends go there for the final drink of the evening (0.3) there is an atmosphere and the drinks are cheap. I don't feel embarrassed to walk in on my own and there's always someone to talk to and scrounge a fag off (h).

Naming datafiles

Each interview you transcribe should be saved as a separate word-processed file. As part of this we recommend using a filename that maintains confidentiality and preserves anonymity, can easily be recognised, and which codifies important information. When doing this Mark always starts his transcription filenames with the interview number and saves the word-processed transcripts for each research project in a password protected separate subdirectory. Subsequent parts of the filename provide more detail. Thus the file '26MPOrg1.docx' is the transcript of the **26th** interview, **Male**, **Professional**, undertaken at **Organisation1**. As some CAQDAS programs require filenames of eight or fewer characters, you may need to limit your filenames to this length.

Using electronic text data including scanned documents and automatic transcription

For some forms of text data such as, for example, email interviews or electronic versions of documents, including organisational emails, blogs and web-based reports, your data will already be in electronic format. Although these data have already been captured electronically, you are still likely to need to spend some time preparing them for analysis. This is likely to involve you in ensuring that these data are:

- suitably anonymised, such as by using separate codes for yourself and each participant;
- appropriately stored for analysis, for example one file for each interview, each meeting's minutes or each organisational policy;
- free of typographical errors and, where these have occurred, they have been 'cleaned';
- free of other errors such as parts of speech being incorrectly recognised by automatic transcription.



Box 13.3 Checklist

Transcribing interviews

- ✓ Have you thought about how you intend to analyse your data and made sure that your transcription will facilitate this?
- ✓ Have you chosen clear interviewer and participant identifiers and used them consistently?
- ✓ Have you included the interview questions in full in your transcription?
- ✓ Have you saved your transcribed data using a separate file for each interview?
- ✓ Does your filename maintain confidentiality and preserve anonymity while still allowing you to recognise important information easily?
- ✓ Have you ensured your data files maintain confidentiality and preserve anonymity?
- ✓ Have you checked your transcript for accuracy and, where necessary, 'cleaned up' the data?
- ✓ (If you intend to use CAQDAS) Will the package you are going to use help you to manage and analyse your data effectively? In other words, will it do what you need it to do?
- ✓ (If you intend to use CAQDAS) Are your saved transcriptions compatible with the CAQDAS package you intend to use, so you will not lose any features from your word-processed document when you import the data?
- ✓ (If you intend to use CAQDAS) Have you checked your transcript for accuracy and 'cleaned up' the data *prior* to importing into your chosen CAQDAS package?
- ✓ Have you stored a separate backup or security copy of each data file, for example on your phone?

Planning for analysis

When transcribing audio recordings or your own notes you need to consider how you intend to analyse your transcriptions. If you only have access to a black and white printer, there is little point in using different coloured fonts to distinguish between participants in a group interview or to distinguish non-verbal responses such as nervous laughter in your transcripts as these will be difficult to discern when working from the paper copies.

You also need to be careful about using these and other word-processing software features if you are going to analyse the data using CAQDAS. These programs often have precise file formats, which can mean that word-processing software features such as *bold* and *italics* generated by your word-processing software will disappear when your data file is imported (Silver and Lewins 2014). For example, although you may transcribe your interviews using a word processor such as Microsoft Word, your chosen CAQDAS package may require this textual data to be saved as a text-only file (.txt) or using rich text format (.rtf), resulting in the loss of some of these features. These along with other aspects of transcribing interviews are summarised as a checklist in Box 13.3.

13.5 Aids to help analysis

Your earlier recording of contextual information about the interviews or observations that you conduct (Section 10.5) will also help you to recall the circumstances and context of each as well as informing your interpretation. Various researchers have suggested ways of

recording information and developing reflective ideas to supplement your written notes or transcripts as you analyse your data (e.g. Brinkmann and Kvale 2015). These include:

- interim or progress summaries;
- transcript summaries;
- document summaries;
- self-memos;
- research notebook;
- reflective diary or journal.

The way you use these analytical aids will be dependent on your preferred approach to recording your ideas and reflections, and the context of your research. Where you produce transcripts of interviews or observations, it will be helpful to write a transcript summary for each one; similarly where you use documents, it will be helpful to write document summaries. Your university may require you to keep a reflective diary, and you may also find it helpful to write interim summaries, self-memos or keep a research notebook.

Interim or progress summaries

As your analysis progresses it is helpful to write an **interim summary** after one or a set of related interviews or observations or period of using secondary data. In this your summaries will show the development of your thoughts and aid your analysis and the direction of your subsequent data collection. Alternatively, your interim summary may become a working document that you add to and modify as your research project progresses. An interim summary may include:

- what you have found so far;
- how much confidence you have in your findings and explanations to date;
- what you intend to do to improve the quality of your data and/or to seek to substantiate your apparent explanations or to seek alternative explanations;

Transcript summaries

After you have written up your notes, or produced the transcript of an interview or observation, you can also produce a summary of the key points that have emerged. A **transcript summary** compresses long statements into briefer ones, the main sense of what has been said or observed being paraphrased into fewer words. Through summarising you will become conversant with the principal themes emerging. You may be able to identify and note possible relationships between themes so that you can return to these to seek to establish their wider credibility. It will also be useful to include some contextual information about the person(s) you interviewed or observed, the setting in which this occurred and whether anything occurred during the interview or observation which might have affected the nature of the data that you collected (Boxes 10.5, 13.4). Once you have produced a summary of the key points that emerge from the interview or observation and its context, you can attach a copy to the file of your written-up notes or transcript for further reference.

Document summaries

Where you use any sort of document, it is helpful to produce a **document summary**. This can be used to summarise and list the document's key points for your research, outline how it relates to your work and why it is significant. As your research progresses, there is



Box 13.4 Focus on student research

Noting an event that affected the nature of data collection

Birjit was the moderator for a focus group whose participants were the customers of a large department store. Approximately halfway through the allotted

time, an additional participant joined the group. This person almost immediately took control of the discussion, two other participants appearing to become reticent and withdrawing from the discussion. Despite this, all Birjit's questions were answered fully, and she felt the data she had obtained was valuable. However, she recorded the point at which the new participant joined the group in a post-transcript summary in case any divergence was apparent between the nature of the data in the two parts of the focus group.

a likelihood that you will forget some of your thoughts about your previous data collection and analysis, so that a document summary, like other analytical aids discussed in this sub-section, will act as a reminder of your earlier ideas.

Self-memos

Self-memos allow you to record ideas about any aspect of your research, as you think of them. Where you omit to record an idea as it occurs to you it may well be forgotten. The occasions when you are likely to want to write a memo include:

- when you are writing up interview or observation notes, or producing a transcript;
- when you are coding and categorising data;
- as you continue to categorise, analyse and interpret these data;
- when you are constructing a narrative;
- when you are writing up your research project.

Most CAQDAS programs include a self-memoing tool for adding comments or write memos as you are analysing your data (Silver and Lewins 2014). This dates self-memos automatically so you can also trace the development of your ideas.

Ideas may also occur during an interview or observation. In this case you may record the idea briefly as a margin or scratch note and write it as a memo to yourself after the event. Similarly, ideas may occur as you work through a documentary source or create a research diary entry. We suggest you use your phone or carry a notebook to record your ideas, whenever and wherever they occur.

Self-memos may vary in length from a few words to one or more pages. They can be written as simple notes and do not need to be set out formally. It is useful to date them and include cross-references to appropriate places in your written-up notes or transcripts, where appropriate. Alternatively, an idea that is not grounded in any data (which may nevertheless prove to be useful) should be recorded as such. Self-memos should be filed together and where appropriate they should be linked to specific data. Memos may also be categorised and can help you in the later stages of your analysis. They may also be updated as your research progresses, so your bank of ideas continues to have currency and relevance.

Research notebook

An alternative approach for recording your ideas about your research is to keep a **research notebook**. You may keep such a notebook alongside the creation of self-memos. Its purpose

will be similar; to record your ideas and reflections, and act as an aide-mémoire about your intentions for the direction of your research. Using a chronological format may help you to identify the development of certain ideas (such as data categories, propositions or hypotheses) and the way in which your research has progressed.

Reflective diary or journal

In Section 1.5 we recommended you also keep a reflective diary or journal about your experiences of undertaking research, what you have learnt from these experiences, how you will seek to apply this learning and what you need to do to develop your competence further. Reflection occurs in several ways. It can occur during an event, as you reflect on your approach while you are conducting an activity such as when interviewing or observing. Reflection may also occur after an activity has taken place as you consider on what occurred and how you might be able to do better next time. A more fundamental type of reflection, known as reflexivity, involves you in monitoring and reflecting on all aspects of the research project from initial ideas to submission of the project report. It includes recognition of the relational, cultural and political practices associated with interviewing, the implications of these, and to report variations between interview account (Reissner and Whittle 2021). Your reactions, your interactions with those taking part and your attitudes and beliefs may each impact on your interpretation of the data that are shared with you. Engaging in forms of reflexivity may enable you to develop greater insights as you explore and analyse these data. Developing a reflexive focus in your reflective diary may therefore prove to be a valuable aid to further your research (Section 1.5).

13.6 Thematic Analysis

Overview

Thematic Analysis is often considered a general analytic approach or method in which patterns of meaning are developed through processes of coding (Braun and Clarke 2006; 2022). The process is found in other approaches to qualitative analysis, albeit in more particularised ways, as we outline in subsequent sections. The essential purpose is to search for themes, or patterns, that occur across a data set (such as a series of interviews, observations, documents or diaries). Thematic Analysis involves you coding your qualitative data to identify themes or patterns for further analysis, related to your research question.

Thematic Analysis offers a systematic yet flexible and accessible approach to analyse qualitative data (Braun and Clarke 2022). It is systematic as it provides an orderly and logical way to analyse qualitative data, leading to rich descriptions, explanations and theorising. Thematic Analysis can be used to help you:

- 1 comprehend often large and disparate amounts of qualitative data;
- 2 integrate related data drawn from different transcripts and notes;
- 3 identify key themes or patterns from a data set for further exploration;
- 4 produce a thematic description of these data; and/or
- 5 develop and test explanations and theories based on apparent thematic patterns or relationships;
- 6 draw and verify conclusions.

Thematic Analysis is flexible as it is not tied to a particular research philosophy. You may use Thematic Analysis irrespective of whether you are adopting an objectivist or subjectivist position (Section 4.2). Your assumptions will, however, affect how you use it to interpret your data. As a critical realist you may use Thematic Analysis to seek to understand factors underpinning human attitudes and actions. Alternatively, as an interpretivist you may use it to explore different interpretations of a phenomenon.

Thematic Analysis can be used irrespective of your approach to theory development. In a deductive approach, the themes you wish to examine are linked to existing theory. Your research question is also more likely to be firmly established and this and your research objectives may be used to derive the themes through which to examine your data. This may lead you to focus on parts of your data set rather than seek to analyse it all.

In an inductive approach, your themes will be derived from the data. You will search the entire data set for themes to explore related to your research interest but will not impose a framework of themes to examine your data set based on existing theory. Depending on which themes emerge and reoccur in an inductive approach, you may also modify your research question. You may also use an abductive approach, either commencing analysis with theoretically derived themes that you then modify or add to as you explore your data or using existing theories to structure inductively identified concepts (Bishop et al. 2020).

Process

The nature and flexibility of Thematic Analysis mean that it is comparatively straightforward. Where you use Thematic Analysis, your energy can be invested in making sure your analysis is rigorous, rather than spending lots of time checking you are applying a more particularised approach to qualitative analysis according to strict rules advocated for its use. In practice, this procedure is neither simple nor linear. Rather, it is concurrent and recursive, involving you analysing data as you collect them and going back over earlier data and analysis as you refine the way in which you code and categorise newly collected data and search for analytical themes.

Thematic analysis involves six phases:

- data familiarisation;
- data coding;
- initial theme generation;
- theme development and review;
- theme refining, defining, and naming;
- (writing up).

Data familiarisation

You will start to become familiar with your data as you produce transcripts of the interviews or observations you conduct, or as you read through documents or diaries or review visual images. The act of transcribing a data item yourself, although laborious, allows you to develop familiarity. It should also prompt you to generate summaries, self-memos or entries in your notebook that aid your analysis.

Familiarisation with your data involves a process of immersion that continues throughout your research project. You will need to read and re-read your data during your analysis. You will be interested to look for meanings, recurring themes and patterns in your data.

Without familiarity, you will not be able to engage in the analytical procedures that follow. Producing transcripts and data familiarisation are therefore important elements in analysing data.

Data coding

Coding is used to categorise data with similar meanings. **Coding** involves labelling each unit of data within a data item (such as a transcript or document) with a code that symbolises or summarises that extract's meaning. Your purpose in undertaking this process is to make each piece of data in which you are interested accessible for further analysis (Boxes 13.5 and 13.6). Qualitative data sets are frequently large and their content complex. A qualitative data set may include references to actions, behaviours, beliefs, conditions, events, ideas, interactions, outcomes, policies, relationships, strategies, etc. Without coding these data you may struggle to comprehend all the meanings in your data in which you are interested. Coding is therefore an important means to manage your data so that you can rearrange and retrieve them under relevant codes. This process effectively involves fragmenting your original data items and regrouping units of data with similar meanings together to be able to examine them in relation to other groups of similar units of data.

A **code** is a single word or a short phrase, which may also be abbreviated in use (Boxes 13.5 and 13.6). A coded extract of data is referred to as a unit of data. A **unit of data** may be a number of words, a line of a transcript, a sentence, a number of sentences, a complete paragraph, other chunk of textual data or visual image that had one or more codes attached to it (Boxes 13.5 and 13.6). The exact size of a unit of data will be determined by its meaning. Some units of data will overlap, and some will be coded using more than one code (Box 13.5).

The process of coding allows you to link units of data that refer to the same aspect or meaning, or to link aspects or meanings that you want to compare and contrast. It allows you to rearrange your original data into groupings for the next stage of analysis. Any unit of data may be coded with as many different codes as you think is appropriate, creating a web of connections to aid your analysis (Boxes 13.5 and 13.6). It is often important to understand the context of the data you are analysing. Where it is important to include some contextual background, you can code larger units of data such as whole paragraphs, as opposed to smaller units such as a few words or single sentences. You should also note that codes may be referred to as categories: these terms are sometimes used interchangeably and sometimes to refer to different aspects of the analytical process – see the next sub-section.

If you think that a new piece of data has a similar meaning to a previously coded unit of data, it should be labelled with the same code. If you think that a new piece of data does not have a similar meaning to a previously coded unit of data, you will need to devise a new code for it. Throughout the process of coding you need to keep a list of codes you are using and a working definition for each, to ensure consistency.

At this point you may be asking two questions. How much of my data should I code – all or only some of it? Where should my codes come from? Both of these questions are related to your approach to theory development and your research question – whether you are setting out to use an inductive or deductive approach and how well you have defined your research question. We answer each of these questions in turn.

How much of your data you code will depend upon your research approach and research question. Where you use a purely deductive approach, you will commence with a framework of codes derived from prior conceptual or theoretical work. In this case you are likely to commence coding by applying theoretically codes to your data and developing an initial thematic map of how you think the codes relate. This is where you start

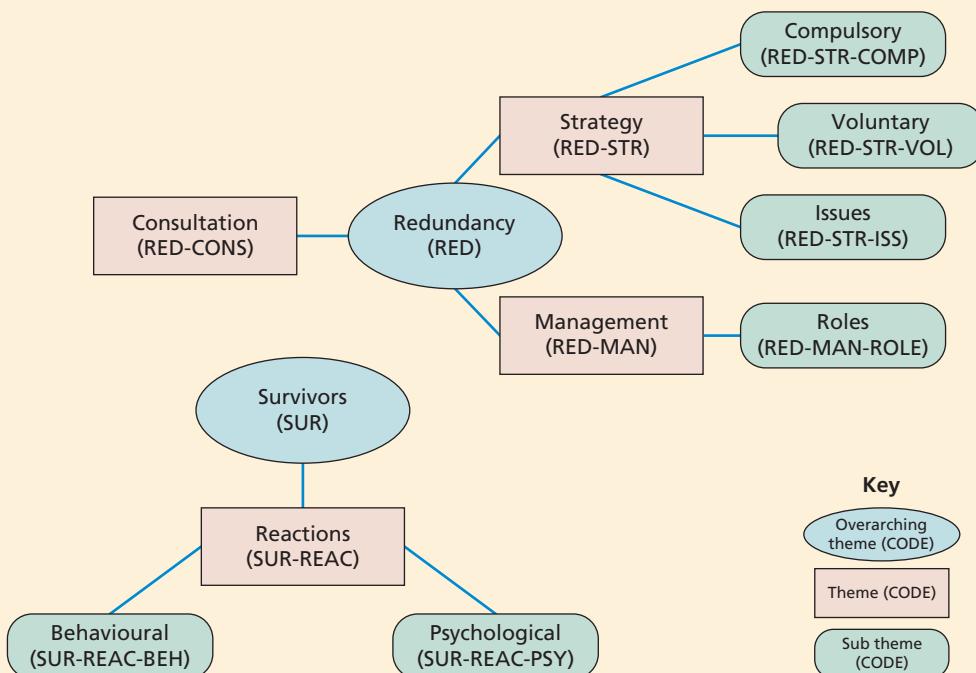


Box 13.5 Focus on student research

Coding an interview deductively using *a priori* codes

Adrian's research project was concerned with how human resource management professionals managed a downsizing process in their own organisations and the impact on surviving employees. He derived initial codes from existing theory in the academic literature

and constructed an initial thematic map. He attached his most detail level codes (sub themes) to appropriate units of data in each transcript. His coding was hierarchical, the codes he used being shown in brackets in the thematic map. These were then attached to the interview transcript, using sentences as units of data. Like our jigsaw example at the start of this chapter, those units of data that were coded with more than one category suggested interrelationships. (Later when he wrote his methodology chapter, Adrian included a copy of his finalised thematic map.)



Adrian's initial thematic map

| CODE | TRANSCRIPT | LINE |
|--------------|---|------------------|
| RED-CONS | 27MM The first stage is to find out what particular employees want for themselves and how they want this to happen. Staff are seen by their line manager and/or a member of personnel. | 1 2 3 |
| RED-MGT-ROLE | Employees might want to talk to someone from personnel rather than talk with their line manager – well, you know, for obvious reasons, at least as they see it – and this would be acceptable to the organisation. This meeting provides them with the opportunity to | 4 5 6 7 |
| RED-STR-VOL | | |





Box 13.5 Focus on student research (*continued*)

Coding an interview deductively using *a priori* codes

| CODE | TRANSCRIPT | LINE |
|--------------|---|------|
| RED-STR-ISS | opt for voluntary redundancy. We do not categorise employees into anything like core or non-core, although we will tell a group | 8 |
| RED-CONS | of employees something like 'there are four of you in this particular function and we only need two of you, so you think | 9 |
| RED-CONS | about what should happen'. Sometimes when we attempt to give employees a choice about who might leave, they actually ask us to make the choice. This is one such situation where a compulsory | 10 |
| RED-STR-COM | selection will occur. We prefer to avoid this compulsory selection | 1 |
| SUR-REAC-PSY | because of the impact on those who survive – negative feelings, guilt and so on. | 2 |
| | | 3 |
| | | 4 |
| | | 5 |
| | | 6 |
| | | 7 |

thinking about the relationship between the overarching themes, themes, and sub themes. Once you have a set of themes and sub themes, you will begin to see the significance of individual themes. You will then review and refine these and produce your first thematic map (Box 13.5). These themes are subsequently reviewed, refined and, eventually, clearly defined by comparing them with the data. A final thematic map is then produced, which can be included in the methodology of your project report.

Where you use a purely inductive approach you will be likely to code all your data, as you explore all possible meanings to guide the direction of your research. If you are following the Gioia methodology (Gioia et al. 2012) you will start by coding at the most detailed level, identifying first order concepts from the data and coding them using data-centric terms. As your analysis progresses you will start to see similarities and differences between these concepts allowing them to be grouped into more abstract second order themes. You will now be asking yourself whether these emerging themes help you understand and explain what you are observing. Once you have a workable set of concepts and themes you then try and distil these even further into a smaller number of aggregate dimensions and can build a data structure (Box 13.6). In a similar way to a thematic map, your data structure not only helps you to configure your data but also, when included in your methodology, provides a graphical representation of how you moved from your data to themes helping you demonstrate rigour in your research.

A purely inductive approach may mean that you spend a great deal of time coding every possible unit of data before you decide on a particular research focus. Using a purely inductive approach is appropriate for a very exploratory study but you would need to ensure that

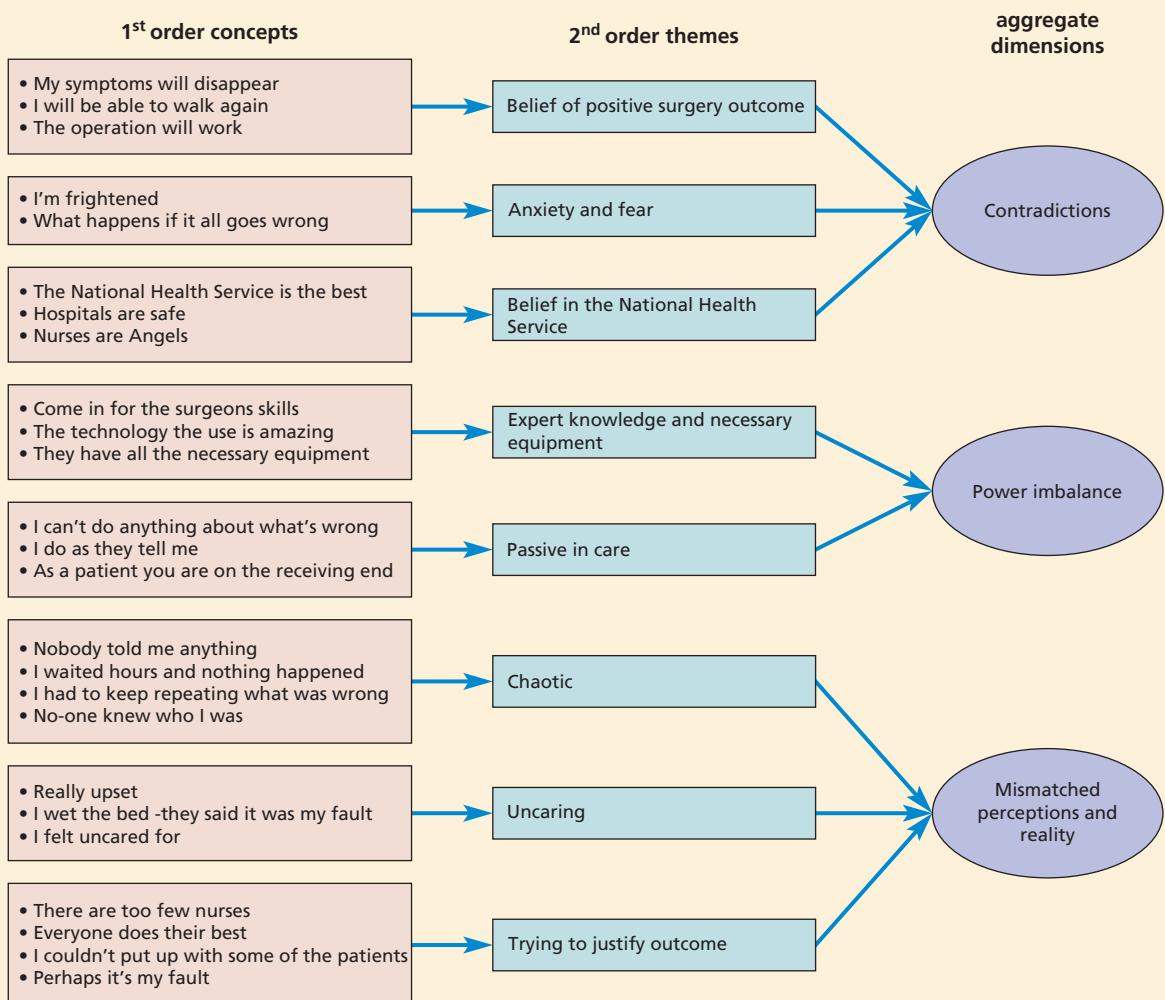


Box 13.6 Focus on student research

Coding an interview inductively

Carole was interested why patients who were in hospital for elective surgery trusted their doctors and nurses. Having received ethical approval from the hospital where she worked, she interviewed 14 patients before and after their hospital stay using face to face

semi-structured interviews and fully transcribed her audio recordings. Reading and re-reading her transcripts she identified first order concepts labelling them using the participants' own words (*in-vivo codes*). These she grouped into themes iteratively, constantly comparing patients' accounts to select and reduce the data to a core set of second order themes. She then looked for evidence of explanations in her data, these becoming her aggregate themes. She summarised these as a data structure diagram in her research project.



you have ample time to conduct it, perhaps related to a major research project. Where you use an inductive approach and have defined a research question, you should be able to use this question to help select which data to code. In this case, while all your data may be potentially interesting, your research question will help you focus on which data to code. Using a purely deductive approach may lead you to conclude that your list of prior codes is inadequate and that you need to devise other codes to be able to code your data adequately to begin to answer your research question and address your research objectives.

This discussion indicates where your codes may come from. There are three main sources of codes, which dependent on your approach to theory development can be used on their own or in combination. Codes may be:

- actual terms used by your participants, recorded in your data. These are often referred to as 'in vivo' codes;
- labels you develop from your data (Box 13.7);
- derived from existing theory and the literature. These are often referred to as 'a priori' codes.

These sources of codes are shown in Figure 13.1 to illustrate their relationship.

Your approach to coding will be guided your research question and objectives. Another researcher with different objectives might derive different codes from the same data. You will be likely to develop new codes as you conduct more interviews or observations and expand your data set. You will also be likely to gain new insights about the utility of existing codes, that suggest possible amendments alongside new codes during the process of analysis. This will require you to re-read all your earlier transcripts and re-code them according to your current list of codes. This process is termed constant comparison and is undertaken to ensure consistency in the way you code and analyse your data set.

Your codes will indicate the occurrence or non-occurrence of a theme and the strength of opinion in some instances. Some codes may be attached to a large number of units of data, proving too broad for further analysis without being subdivided. For example, a research project where some codes (themes) had large amounts of data attached to them, while others attracted relatively small amounts of data. This can lead to the large codes being subdivided into further codes (sub-themes) during the analysis (Box 13.5). Codes attracting small numbers of units of data may be merged with similar ones or retained until later in the process of analysis in case they prove to be more important than they appear initially.

You may use CAQDAS to help you to code your data (Section 13.14) or you may use a manual approach (Box 13.7). Where you use a manual approach, you can label a unit of data with the appropriate code (or codes) in the margin of your transcript or set of notes

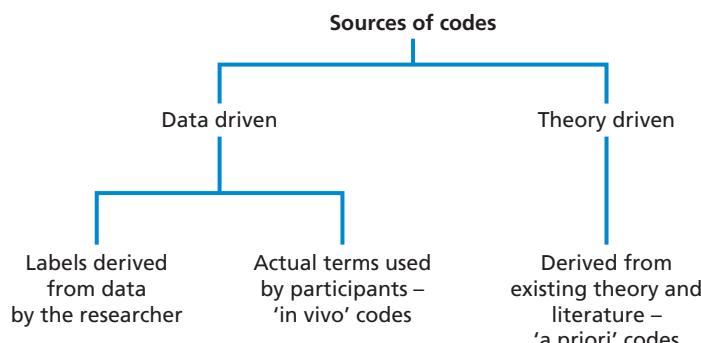


Figure 13.1 Sources and types of code



Box 13.7 Focus on management research

Inductively coding diary entries

Hjálmsdóttir and Bjarnadóttir's (2021) research explores the gendered realities of work–life balance during the Covid-19 pandemic. In particular, they consider the how the changes brought about by the pandemic reflected and affected the gendered division of unpaid labour including childcare and household chores.

In their paper they analyse the open diary entries from 37 mothers who were in heteronormative relationships, in which their participants kept old-style diary entries. While writing these, they were asked to reflect on their day, the impact of Covid-19 on their life, the division of household duties and

responsibilities, and other aspects they wished to include.

Their written reflections were analysed drawing on Braun and Clarke's (2006) thematic analysis. Both authors sorted the diary text by date and participant and read it several times, adding notes and then discussing the diaries' content with each other. Initial coding of each diary entry was open and undertaken inductively, focussing on the participants' experiences as recorded. The authors then collated similar codes and associated text segments to look for and identify repeated patterns of meaning within the data. This generated two main themes:

- complexities of work life balance in COVID-19 times including particularly gendered interactions of mental work, stress, and work-life balance;
- emotional labour performed by women including how conscious they were of family members' wellbeing.

(Box 13.6). Text with the same or similar codes can then be copied and collated to identify repeated patterns. When doing this, it is essential to label each unit of data carefully so that you know its precise source. An alternative is to index each segment by recording precisely where they occur in your transcripts or notes (e.g. interview 7, page 2, line 16) on cards headed with particular codes. This has the effect of reducing and rearranging your data into a more manageable and comprehensible form. In undertaking this stage of the analytic process, you are engaging in a selective process, guided by the aim of your research and your research objectives.

Theme generation, development and review

Although seen as two distinct stages of analysis, in practice you will be searching for themes, developing them and reviewing them as you collect and code data. Your progress summaries, transcript summaries, document summaries, self-memos and/or entries in your research notebook and reflective diary will help you to record your ideas about potential themes, patterns and relationships in your data. A **theme** is a broad category incorporating several codes that appear to be related and indicate an idea that is important to your research question. A theme may also be a single code indicating an idea of general importance to your research question. Searching for themes condenses your coded data by grouping these with similar codes into analytic categories.

The search for themes begins fully when you have coded all your data. At this point you will have created a long list of codes to make sense of and draw meaning from your data. Advice about the number of codes varies considerably. Some suggest working with up to 30 codes, while others suggest creating as many codes as you require to interpret every relevant meaning in your data. This may mean creating up to a couple of hundred codes, or possibly more. Our view is that data should not be constrained by a particular number of codes. Rather, your number of codes will be related to the meanings you wish to explore

in your data set, the nature of your research approach and the focus of your research question. However, if you find yourself creating very large numbers of codes you will need to consider whether your coding is too detailed. Always refer back to your research question, research aim and research objectives to focus your approach to data analysis.

In some discussions, terms like codes and themes are used interchangeably and this can lead to confusion (Ritchie et al. 2014). We are sometimes asked about the difference between a code and a theme. One way we have found it helpful to explain this is to say that data are organised by coding them while codes are organised by drawing them together as themes. You may also see the term ‘thematic code’: this is simply an alternative way to refer to the type of coding we describe here, where your coding leads to identification of themes as opposed to coding for actions as advocated by Charmaz (2014) in constructivist grounded theory (Section 13.9).

Searching for themes involves you making judgements about your data and immersing yourself in them. As you search these codes, initial questions you can ask include:

- What are the key concepts in these codes?
- What, if anything, seems to be recurring in these codes?
- What seems to be important, whether it recurs often or not?
- What patterns and/or trends are evident in the coded data?
- Which codes appear to be related?
- How do a particular set of codes appear to be related?

Next, as you start to decide on themes through which to analyse your data further, additional questions may include:

- What is the essence of each theme?
- How might themes be related to each other?
- Which themes appear to be main themes and, within these, which appear to be sub-themes? (There may also be third level themes evident in your analysis.)
- How can the relationship between themes be represented (as a hierarchy or a network) to produce a thematic map?
- Is there an overarching theme (or more than one) that unites the analysis?

You should not expect this process to be unproblematic. In attempting to achieve a thorough understanding of your data set, some further questions you may ask include:

- How well does this initial thematic map represent the relationships between themes?
- Which themes, if any, do not fit within this thematic representation?
- Does the way the data have been coded need to be revised; if so which data and how?
- Which themes need to be refined, discarded or newly introduced?
- How may the thematic representation be modified to represent my data better?

In the first set of questions, you begin to decide on themes to further your analysis. In the second set of questions, you begin to define your themes and their relationships. Some themes (2nd order themes) will become overarching themes (aggregate dimensions); some may become themes (2nd order themes), linked to a overarching theme (aggregate dimension) yet others may be sub themes (1st order concepts), linked to a theme (2nd order theme). In the third set of questions, you evaluate your themes (2nd order themes) and the relationships between them. This will mean refining your themes and testing propositions.

Theme refining, defining and naming

Refining defining and naming themes is an important part of your analytical process. The themes that you devise need to be coherent to provide you with a well-structured analytical

framework to pursue your analysis. As you develop themes you should reorganise your coded data extracts under the relevant theme or sub-theme. This will allow you to evaluate whether these coded data are meaningful to one another within their theme and, more widely, whether (and how) themes are meaningful in relation to one another. This is likely to be an iterative process, as you re-read and reorganise your data. As you continue to examine your data, you will be likely to refine these themes.

You may decide that some of your initial themes should be combined to make a new theme while others should be separated into different themes, or even discarded. By reading the coded data attached to a theme, you will be able to evaluate whether they support the continuation of the theme. This will also allow you to decide if the data are too dissimilar so that these should be separated into two or more themes. It will also allow you to decide whether two or more themes contain similar meanings and so should be collapsed into a single theme. As you refine your themes in this way you will also be able to revise the relationships between them.

As you seek to reveal patterns within your data and to recognise relationships between themes, you will be able to develop testable propositions (Box 13.8). The appearance of an apparent relationship or connection between themes will need to be tested to establish whether there is an actual relationship. However, while this is sometimes referred to as ‘testing a hypothesis’, it is not the same as the statistical hypothesis or significance testing we discuss in relation to quantitative analysis (Section 12.9).

It is important to test the propositions that emerge inductively by seeking alternative explanations and negative examples that do not conform to the pattern or relationship being tested. Alternative explanations frequently exist, and only by testing the propositions that you identify will you be able to move towards formulating valid conclusions and an explanatory theory, even a simple one (Miles et al. 2019). Dey (1993: 48) points out that ‘the association of one variable with another is not sufficient ground for inferring a causal or any other connection between them’. The existence of an intervening variable may offer a more valid explanation of an association that is apparent in your data (Box 13.9).



Box 13.8 Focus on student research

Research propositions

During the process of qualitative data analysis, a student evaluating the use of online retailing formulated the following proposition:

Customers' willingness to trust specific online retailer depends on their previous customers' review for that retailer.

A student exploring mortgage borrowers' decision making drew up this proposition:

Potential mortgage borrowers' choice of lending institution is strongly affected by the level of customer service that they receive during the initial inquiry stage.

Another student investigating cause-related marketing formulated the following proposition:

Companies engaging in cause-related marketing are motivated principally by altruism.

A relationship is evident in each of these propositions. Each was tested using the data that had been collected.



Box 13.9 Focus on student research

The impact of an intervening variable

Kevin's research project involved looking at the use of subcontractors by an organisation. A relationship appeared to emerge between the awarding of contracts to a particular subcontractor and the size of that contractor in terms of number of employees; in particular, those contractors with larger numbers of employees appeared to be awarded more contracts. This could have led Kevin to conclude the awarding of contracts to a particular subcontractor was related to their size and that the organisation tended to use larger subcontractors.

Reality was not so simple. The organisation had originally used over 2,500 subcontractors but had found this exceedingly difficult to manage. To address this issue the organisation had introduced a system of preferred contractors. All 2,500 subcontractors had been graded according to the quality of their work, with those whose work had been consistently of high quality being awarded preferred contractor status. This meant that they were invited by the organisation Kevin was researching to tender for all relevant contracts. The intervening variable was therefore the introduction of preferred contractor status dependent upon the quality of work previously undertaken. The fact that most of these subcontractors were larger was not the reason why the organisation had awarded them contracts.

By rigorously testing your propositions against your data, looking for alternative explanations and seeking to explain why negative cases occur, you will be able to move towards the development of valid/credible and well-grounded conclusions. The validity/credibility of your conclusions needs to be verified by their ability to withstand alternative explanations and negative cases. **Negative cases** are those that do not support your explanations or the induction of your grounded theory. These should be seen positively as these will help to refine your explanations and direct the selection of further cases to collect and analyse data.

This will help you to avoid interpretations that prove to be unreliable because you only notice evidence that supports your own opinions. It relates to our earlier discussion of reflexivity (Section 13.5) and the need to recognise your own attitudes and beliefs about the topic being researched, to understand how this affects your judgement while analysing these data. Brinkmann and Kvale (2015: 278) refer to this process as seeking to achieve 'reflexive objectivity'.

Evaluation

Thematic Analysis is a generic systematic approach to qualitative data analysis that is accessible and flexible and not overly prescriptive. It is suitable to use with several qualitative research strategies, where you are not following a named version of a strategy that prescribes precise analytic procedures, as in the Grounded Theory Method. Thematic Analysis is adaptable, so if the research strategy you are using requires you to search for particular themes you may consider using it. The process of searching for themes is common to other analytical approaches, as we consider in the following sections. It can be used for deductive, inductive and abductive approaches to theory development.

13.7 Template Analysis

Overview

Template Analysis is a specific approach to thematic analysis, with a few key differences to that outlined previously (Section 13.6). In Template Analysis only a proportion of the data items are coded before developing an initial coding structure and interpretive themes, these data items being chosen for their representativeness or heterogeneity to try to overcome it (King and Brookes 2017). This contrasts with some other approaches to Thematic Analysis where all data items (transcripts or other text) are coded before the search for interpretive themes begins fully, thereby avoiding early thematic interpretation prematurely shaping or skewing the direction of the research in this emergent approach.

Within Template Analysis the **coding template** is a hierarchical representation of the codes and interpretive themes, providing the central analytical tool. A researcher using Template Analysis will start by coding a sufficient part of their data to develop an initial coding template (Box 13.10). This may mean coding a small number of transcripts to develop an initial set of themes. These are then arranged and rearranged until a satisfactory initial template is developed, representing a hierarchy of higher order themes, subthemes and lower order thematic codes. Subsequent transcripts are then coded using the codes in this initial template. This is modified as new data suggests deficiencies in the codes, leading eventually to the development of a final coding template. Template Analysis is a standalone analytical technique, rather than being part of a wider methodological approach. Consequently, it may be used irrespective of whether you are adopting an objectivist or subjectivist position or whether you adopt a deductive or inductive approach to theory development. It may commence with a number of *a priori* codes, which are then supplemented by the use of *in vivo* codes.

Process

King and Brookes (2017) describe a procedure for Template Analysis composed of six stages, involving:

- 1 familiarisation with data;
- 2 preliminary coding;
- 3 clustering codes;
- 4 production of an initial coding template;
- 5 development of this template;
- 6 application of the final template.

The first stage, familiarisation with data, is the same as the initial stage of Thematic Analysis. During this stage, you will need to become familiar with your data by transcribing these and carefully reading each transcript several times, to understand and gain insights into these data.

As you become familiar with your transcribed data, you can look for units of data that relate to your research question and begin to code these. Like with Thematic Analysis, you may begin by using *a priori* codes that you have identified from existing literature, or alternatively *in vivo* codes derived from your data (Section 13.6). Initial use of *a priori* codes may also be supplemented by the subsequent development and use of *in vivo* codes.

As you develop codes and code your data you will start to see how these codes may be related to each other. At this stage you will be clustering your codes and arranging them hierarchically. This is the same process of developing themes that we described for Thematic Analysis, albeit that it occurs earlier in Template Analysis. This leads into the next stage where you produce an initial coding template. This template will show the clusters of codes you have produced in a hierarchical fashion to display the relationships between them, with each cluster being headed by a theme or subtheme.

Box 13.10 provides an example of an initial coding template, showing the hierarchical relationship between the themes. In this example, three levels of themes have been used. The numbering system and placing of lower-level thematic codes towards the right-hand side helps indicate the hierarchical relationships in this coding template. Codes are grouped together in lower levels to show how higher-order themes are constituted.

As your data collection and analysis proceeds, you will also develop your template further. This is an iterative process that involves modifying it until you devise a structure that represents all relevant ideas in your data and the relationships between them, both hierarchically and laterally where appropriate (King and Brookes 2017). Analysing your interview transcripts or observation notes will lead to some earlier themes being revised and even changes to their level or place in the template hierarchy. Where you consider introducing a new code or theme, or altering the level of an existing code or theme in the template, you need to verify this action and explore its implications in relation to your previous coding activity. This is usually more straightforward using CAQDAS (Silver and Lewins 2014). As part of this, it is helpful to use self-memos to note the reasons for these changes.

King and Brookes (2017) outline five principal ways in which a template may be reorganised and revised:



Box 13.10 Focus on student research

Part of an initial template to analyse an advertising campaign's impact

Joss was asked to gather and analyse perceptions from a range of professionals in an organisation about a recent advertising campaign it had commissioned. After conducting a few interviews, transcribing the data and undertaking preliminary coding she embarked on the production of an initial template. She had used existing literature to inform her interview guide and used this to commence her coding and the production of this initial template. This initial template reflected her use of a priori codes to commence analysis, with higher-order themes shown in CAPITALS and lower-order ones in

lower case and *italic script*. An extract of her initial template follows:

1 CONTEXTUAL FACTORS

- 1.1 Reasons for campaign
- 1.2 Environment
 - 1.2.1 Political
 - 1.2.2 Economic
 - 1.2.3 Socio-cultural
 - 1.2.4 Technological
 - 1.2.5 Legal

1.3 Nature of the product

- 1.3.1 Cost
- 1.3.2 Features
- 1.3.3 Target groups

2 NATURE OF THE CAMPAIGN

- 2.1 Media
- 2.2 Coverage

3 AWARENESS BY TARGET GROUPS AND OTHERS

- 3.1 Those in target groups
- 3.2 Others

- Insertion of a new code or theme into the hierarchy as the result of a relevant issue being identified through data collection;
- Deletion of a code or theme from the hierarchy if it is not needed;
- Merging codes or themes that were originally considered distinctive;
- Altering the classification of codes or themes, either by promoting to a higher level or demoting within the coding template;
- Changing the scope of a code or theme. Inserted, deleted, merged and altered codes or themes may have implications, resulting in the need to move a code or theme within the coding template, change its purpose or split it into two or more new codes or themes.

Box 13.11 shows how the themes and codes in the initial coding template in Box 13.10 were altered as the process of data collection and analysis progressed. Several have been deleted and new ones inserted that better reflect the terms used by participants. Some initial themes or codes have been merged. For example, the original, second-level theme, ‘Reasons for campaign’ has been merged with the first-level theme, ‘Contextual factors’, to form a new first-level theme, ‘Perceiving the need for the campaign’. The original second-level themes, ‘Media’ and ‘Coverage’ have both been reclassified to become first-level themes. As a result of this reclassification, the scope of these themes has been enlarged and new subsidiary themes created to encompass this.

Your template is likely to undergo revision until all data have been coded and possibly beyond. The final template should represent all units of data that are relevant to your research question so that no further changes are required. To check this you should work through all your codes and ensure they are appropriately represented through the final template. Once this is achieved all of your data can then be applied to the template. This provides a basis for further analysis and interpretation, allowing the nature of the themes within the template to be fully explored and the relationships between them to be tested in the same way as we described for Thematic Analysis (Section 13.6). This will allow the relative importance of themes to be explored, their different roles that in the overall structure to be recognised (for example, some may contextualise others), the similarity or diversity of participant perspectives to be evaluated, and whether predicted or expected relationships exist or are contradicted. The creation of the final form of a template is



Box 13.11

Focus on student research

Part of a final template to analyse an advertising campaign’s impact

As Joss continued to collect data she used her coding template to conduct her analysis. The coding template was revised as these data were analysed. An extract of her final template follows:

1 PERCEIVING THE NEED FOR THE CAMPAIGN

1.1 Market changes

1.1.1 Globalisation

1.1.2 Competition

1.1.3 Segmentation

1.1.4 Technological convergence

1.1.5 Compliance

1.2 Product promotion

1.2.1 Product awareness

1.2.2 Product differentiation

1.2.3 Product upgrades

2 EVALUATING MEDIA

2.1 Social media

2.2 Television

2.3 Radio

2.4 Printed media

3 EXPLORING COVERAGE

3.1 National

3.2 Regional/Local

3.3 Market segments

therefore not the end of the analytical and interpretive process but a means to explore this further to verify explanations and develop theory.

Evaluation

Like Thematic Analysis, Template Analysis offers a systematic, flexible and accessible approach to analyse qualitative data. It adopts a higher level of structure earlier on than Thematic Analysis through the development of an initial coding template. Using a template may also help you to select *a priori* themes to explore and to identify emergent issues that arise through the process of data collection and analysis that you may not have intended to focus on as you commenced your research project (King and Brookes 2017). Template Analysis's flexibility in developing a coding template early on and then revising this in relation to each subsequent data item or number of items allows you to undertake the stages of analysis (e.g. coding, devising and linking themes, exploring relationships, sense-making) in a more holistic way. However, some researchers may feel constrained by using a template while working though transcripts and may become too focused on applying the template to the data rather than using the data to develop the template (King and Brookes 2017).

13.8 Explanation building and testing

In this section we outline three further processes where the emphasis is on building (or predicting) and testing an explanation. These are analytic induction, deductive explanation building and pattern matching.

Analytic induction

Overview

Analytic induction is an inductive version of explanation building. A key characteristic of this process is that it uses an incremental approach to build and test an explanation or theory. **Analytic induction** seeks to develop and test an explanation by intensively examining the phenomenon being explored through the successive selection of purposive cases. This means the process of collecting and analysing data will be composed of a number of repeated steps to find a valid explanation of the phenomenon being studied (Johnson 2004).

Analytic induction emphasises a cycle of developing and testing propositions that are inductively grounded in participants' data rather than deductively testing existing theory. It may use theory in conjunction with grounded data to formulate the propositions that guide each step to help to find a valid explanation (Bansal and Roth 2000). The analytical procedures are not highly developed or formalised, so you may also find the generic procedures outlined for Thematic Analysis (Section 13.6) helpful to guide your analysis within each case.

Process

Data will have been collected from an initial purposive case study, usually by conducting exploratory interviews or observations. Your data should be analysed to devise codes and themes, and recognise relationships between them to develop an initial definition of a

proposition that seeks to explain the phenomenon being studied. This initial proposition is then tested through the purposive selection (Section 7.9) of a second, related case study, involving further exploratory interviews or observations.

Given the loosely defined nature of your initial proposition, it is likely that it will either need to be redefined or the scope of the phenomenon to be explained will need to be narrowed. Redefining the proposition leads to a third iteration or step in the analytic induction process, involving the purposive selection of a third case study to explore the phenomenon and test your redefined proposition. If at this stage your redefined proposition appears to explain the phenomenon, you may either cease data collection on the basis that you believe you have found a valid explanation or seek to test the explanation in other purposively selected cases to see whether it is still valid.

You are likely to encounter one or more cases where your proposition is not adequate to explain the phenomenon you are studying. These are referred to as negative or deviant cases. When you encounter a negative case you will need to take this into account in redefining your proposition, and to test this in the context of another purposively selected case. This process continues until a redefined proposition is generated that reasonably explains the phenomenon in relevant cases where you have collected and analysed data. In practice, several redefinitions of the proposition may be necessary to develop a valid explanation.

Evaluation

As an inductive and incremental way of collecting and analysing data qualitatively, analytic induction can lead to the development of well-grounded explanations. It encourages the collection of data that are thorough and rich by exploring the actions and meanings of those who participate in this process, through in-depth interviews or observation, or some combination of these methods.

However, it is neither a quick nor an easy approach to conducting qualitative analysis. While it may lead to a well-grounded and unassailable explanation, where all negative cases are either accounted for by the final revised explanation or excluded by redefining the phenomenon being studied, this outcome is only likely to occur if that technique is used in a thorough and rigorous manner. This will involve a search for cases that are related to the phenomenon being studied, the in-depth collection of data within each case and the rigorous analysis of these data to devise a final revised proposition that explains the phenomenon being studied throughout these cases.

Your sample selection needs to be undertaken carefully. Selecting diverse cases related to the phenomenon being studied can help to overcome issues related to theoretical generalisability. For example, if you were seeking to explain how small enterprises respond to regulatory change you could select a sample of cases (organisations) from different business sectors and in relation to a range of regulatory changes, where feasible. Despite this, it may be criticised because of issues about its limited representativeness and generalisability. However, this criticism misses the point of inductive research; namely to find explanations that are well grounded in the context being researched. These explanations will exhibit high levels of reliability and internal validity. Others may subsequently seek to test these explanations in other settings.

Deductive explanation building Overview

Deductive explanation building involves an incremental attempt to build an explanation by testing and refining a predetermined theoretical proposition.

Process

As with analytic induction, the process comprises a number of repeated steps to find a valid explanation (Yin 2018):

- 1 Devise a theoretically based proposition, which you seek to test.
- 2 Undertake data collection using an initial, purposive case study to compare the findings in relation to your theoretically based proposition.
- 3 Where necessary, amend the theoretically based proposition in the light of the findings from the initial case study.
- 4 Select a further, purposive case study to undertake a further round of data collection in order to compare the findings to your revised proposition.
- 5 Where necessary, further amend the revised proposition in the light of the findings from the second case study.
- 6 Undertake further iterations until a satisfactory explanation is reached.

Evaluation

Like pattern matching this technique uses a deductive approach involving testing a theoretical proposition or prediction. Where you are able to utilise existing theory to produce such a proposition or prediction this may make the process of explaining the phenomenon less onerous than using analytic induction. Given the commonality of using a deductive approach in these techniques, we offer a combined evaluation after outlining pattern matching.

Pattern matching

Overview

Pattern matching involves predicting a pattern of outcomes based on theoretical propositions to explain a set of findings (Yin 2018). You develop your conceptual or analytical framework, utilising existing theory, and then test the adequacy of the framework deductively as a means to explain your findings. If the pattern of your data matches that which has been predicted through the conceptual framework you will have found an explanation, where possible threats to the validity of your conclusions can be discounted.

Process

The first use is matching patterns for dependent variables arising from another, independent variable. For example, based on theoretical propositions drawn from the literature you specify a number of related outcomes (dependent variables) that you expect to find as a result of the implementation of a particular change management programme (independent variable) in an organisation where you intend to undertake research. Having specified these expected outcomes, you then engage in data collection and analysis. Where your predicted outcomes are found, it is likely that your theoretically based explanation is appropriate to explain your findings. If, however, you reveal one or more outcomes that have not been predicted by your explanation, you need to seek an alternative (Yin 2018).

The second use is matching patterns for variables that are independent of each other. In this case you would identify two or more alternative explanations to explain the pattern of outcomes that you expect to find (Box 13.12). Consequently, only one of these predicted explanations may be valid. If one explanation is found to explain your findings then the



Box 13.12 Focus on student research

Pattern matching for explanations

Linzi's research objective was to explain why productivity had increased in a case study organisation even though a number of factors had been held constant (technology, numbers of staff employed, pay rates and bonuses, and the order book) during the period of the increase in productivity. She developed two alternative explanations based on different theoretical propositions to explain why this increase in productivity had occurred in the organisation. Her explanations were related to the following propositions:

- 1 Productivity increase is due to better management, which has been able to generate greater employee engagement, where this proposition is based on theory related to strategic human resource management.
- 2 Productivity increase is due to fears about change and uncertainty in the future, where this proposition is based on theory related to organisational behaviour and the management of change.

These propositions offered her two possible and exclusive reasons why the described phenomenon had occurred, so that where evidence could be found to support one of these, the other, which did not match her outcomes, could be discounted.

others may be discarded. Where you find a match between one of these predicted explanations and the data you have collected and analysed, you will have evidence to suggest this is an explanation for your findings. Further evidence that this is a correct explanation will flow from finding the same pattern of outcomes in other similar cases (Yin 2018).

Use of prior theory will help to determine an initial set of codes for analysis, although these will be subject to change (insertions, deletions and merging) depending on their utility. As you collect data, attach units of data to codes and examine them for emergent patterns, your analysis will be guided by your initial theoretical propositions and explanations. These propositions still need to be tested thoroughly by seeking negative examples and alternative explanations that do not conform to the pattern or association being tested. Your analysis will follow that outlined for Thematic Analysis (Section 13.6)

Evaluation

Pattern matching and deductive explanation building both involve a defined and systematic procedure, linked to specifying theoretical propositions prior to commencing data collection and analysis. Even though the initial theoretical proposition in deductive explanation building may need to be revised during the research, this procedure is shaped by the use of prior theory.

The use of prior theory in either procedure should enable you to develop a well-defined research question and set of objectives. It also means you can start with a clear framework to guide your research linked to testing a theoretical proposition or propositions. The use of prior theory should also help to shape your interview questions.

Your use of predicted explanations should mean that the pathway to answer your research question and objectives is reasonably defined. This will depend on two factors:

- your thoroughness in using existing theory to define clearly the theoretical propositions and conceptual framework that will guide your research project;
- the appropriateness of these theoretical propositions and the conceptual framework for the data that you reveal.

13.9 Grounded Theory Method

Overview

Grounded Theory Method is part of the Grounded Theory research strategy (Section 5.5). It avoids using *a priori* codes derived from existing theory and commences inductively, developing codes from the data collected (Section 13.6). Data collection and analysis are interrelated, the concepts emerging from previously collected and analysed data being used to direct future data collection. Grounded Theory is seen as systematic, or even prescriptive, because it sets out research practices that should be followed. Its use in practice is criticised when researchers only implement some of these elements, not all (Box 5.8).

The elements of Grounded Theory as a research strategy include early commencement of data collection, concurrent collection and analysis of data, development of codes from the data, and the use of constant comparison, self-memos, theoretical sampling, theoretical saturation and theoretical sensitivity, leading to the development of a theory that is grounded in the data. We suggest re-reading about these in Section 5.5 before reading further in this section.

Here we focus on the analytical techniques used in the Grounded Theory Method. Although a number of connected analytical techniques are defined in Grounded Theory Method, the exact nature of these varies between sources that outline them (e.g. Bryant and Charmaz 2007; Charmaz 2014; Corbin and Strauss 2015; Glaser and Strauss 1967) and even between editions of the same book (Corbin and Strauss 2008, 2015; Strauss and Corbin 1998). While all subscribe to practices including concurrent collection and analysis of data, use of inductive codes, constant comparison and theoretical sampling, some versions are more structured and precisely defined (e.g. Strauss and Corbin 1998) while others are more flexible (Charmaz 2014). In the Grounded Theory Method of Strauss and Corbin (1998), the disaggregation of data into units is called open coding, the process of recognising relationships between categories is referred to as axial coding and the integration of categories around a core category to develop a grounded theory is labelled selective coding. In the subsequent editions (Corbin and Strauss 2008, 2015), open coding and axial coding have been merged and selective coding has been relabelled as integration. Alternatively, the more flexible approach to the Grounded Theory Method of Charmaz (2014) consists of two major phases of coding: initial coding and focused coding, while she also discusses and evaluates axial coding (Strauss and Corbin 1998) and the theoretical coding approach developed by Glaser (1998) (Figure 13.2).

Rather than being confused by these variations in technique, we need to step back and recognise this should not be a complicated process. Corbin in Corbin and Strauss (2015) succinctly summarises the process of analysing grounded data. She emphasises the central role of constant comparison which involves comparing units of data with other data to see whether these are similar or different. Similar data are given the same code to group these together. Similar codes are subsequently grouped together as themes, although in Grounded Theory these are often called categories. The properties or dimensions of each category are developed as further data are collected and analysed. These categories that withstand analytical development are eventually integrated around a single category referred to as the core category. Choice of this core category will depend on your research question. This core category and its relationships to these other categories are used to develop a grounded theory.

As an introduction to the analytical techniques associated with using the Grounded Theory Method, we focus on those of Strauss and Corbin (1998) and Charmaz (2014). Where you decide to use a Grounded Theory strategy (Section 5.5) you may find it useful

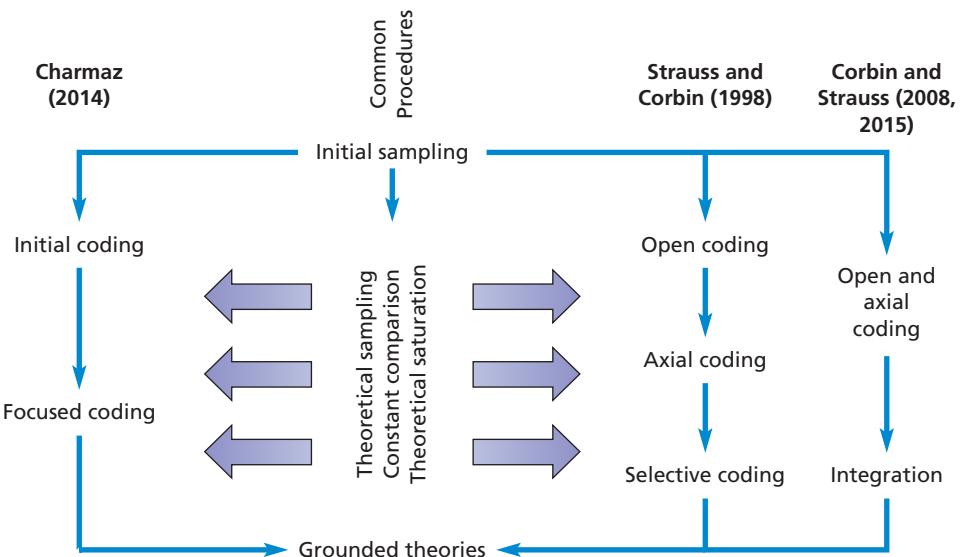


Figure 13.2 Alternative approaches to the Grounded Theory Method
 © 2022 Mark Saunders, Philip Lewis and Adrian Thornhill

to consult these two books and Strauss and Corbin (2015). However, the key to the success in using the Grounded Theory Method is choosing one approach, undertaking this without too much adaptation, and developing your skills in using it (Kenealy 2012). We would advise you to discuss this choice with your project tutor.

Process

Having outlined the elements of Grounded Theory in Section 5.5, we concentrate here on the techniques used in combination to analyse data through different levels of coding. Theoretical sampling (Section 7.9) is used to choose pertinent cases at each phase of data collection and analysis. Your initial sample will be selected in relation to your research question or topic. Each subsequent case will be selected to explore analytical ideas and categories emerging from coding data in the previous case or cases. The purpose of this will be to further develop analytical categories and codes to explore relationships between these to develop a grounded theory. Underpinning this is the process of constantly comparing the data being collected with the codes and categories being used, to develop an emerging theory that is grounded thoroughly in these data. Memo writing throughout your Grounded Theory study allows you to summarise, clarify and develop ideas that relate to the codes you develop, the categories you derive, the relationships between these, the emergence of theory and other aspects related to your study. Theoretical sampling continues until theoretical saturation is reached. This occurs when data collection ceases to reveal new data that are relevant to a category, where the properties or dimensions of categories have become well developed and understood, and relationships between categories have been verified (Figure 13.2).

Initial coding or open coding

Initial coding or **open coding** is similar to the coding process outlined for Thematic Analysis (Section 13.6). Your data are disaggregated into conceptual units and coded

with a label. The same *in vivo* code is used for similar units of data. However, because the Grounded Theory Method commences without an explicit basis in existing theory, the result may be the creation of a multitude of conceptual codes related to the lower level of focus and structure with which you commence your research (Box 13.13). The emphasis is to derive meanings from the actions, interactions, subjects and settings being studied.

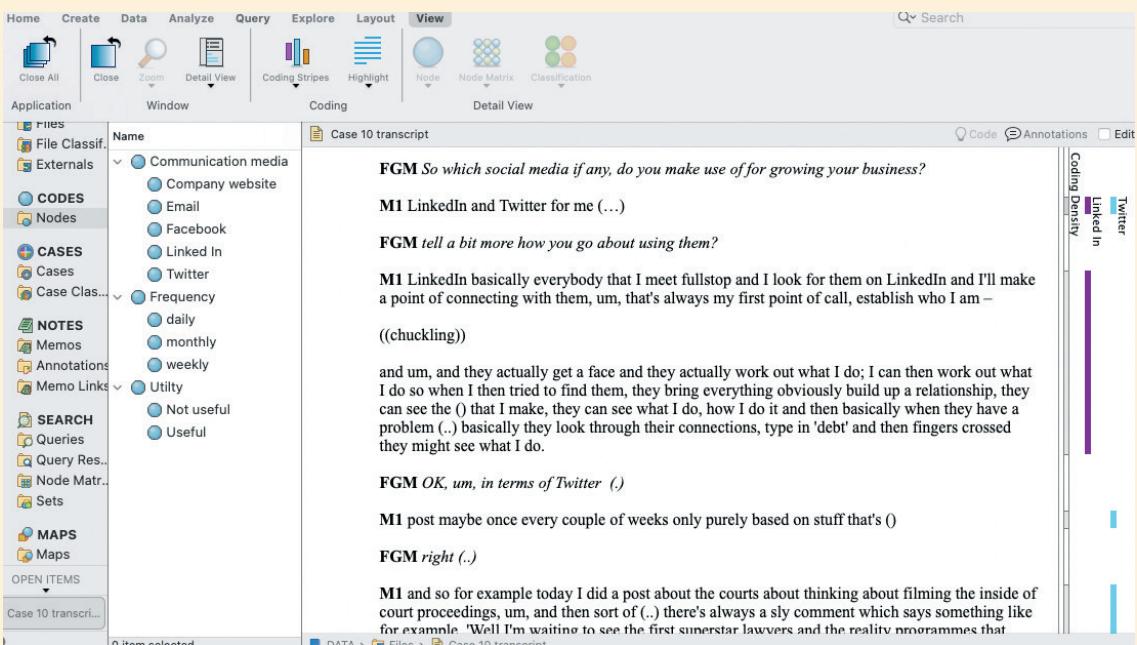
Box 13.13 Focus on student research

Using open coding

Jemma's research was concerned with small- and medium-sized enterprises (SMEs) and their use of social media. She was particularly interested in how they used social media to communicate with potential clients and customers. At the start of her research she undertook two face-to-face focus groups with owner managers of SMEs. The audio recordings of each focus group were subsequently transcribed in Microsoft Word and saved as a docx file. Within the file, Jemma labelled herself, the focus group moderator, as 'FGM'; the male participants as 'M1', 'M2' and so on; and the female participants as 'F1', 'F2' and so on. Each file was then imported into the CAQDAS software NVivo™. Open

codes relating to different communication media such as 'LinkedIn', 'Facebook', 'Twitter', 'Email', 'Letter' and 'Company website' were attached to the transcript, each participant's response being a separate unit of data. Codes were also attached regarding whether the participant felt the social media was 'Useful' or 'Not useful' and the frequency with which it was used.

Based upon analysis using these and other codes Jemma noticed that these SMEs were using social media sites such as LinkedIn widely and frequently to showcase their businesses and to build relationships with customers. However, she noted that clear links between the use of LinkedIn and increased revenues were difficult to establish. Facebook was used less widely in a business context than LinkedIn, being seen predominantly for personal friendships. Twitter was found to be most effective when used in conjunction with other social media such as the business's website. Jemma decided to follow up her initial findings using in-depth interviews.



Charmaz (2014) also advocates **coding with gerunds** rather than coding for themes, to stay close to the meanings in your data and understand these through the actions or interactions that take place in the data. A gerund is a word that ends in ‘-ing’ that is made from a verb but used like a noun. In Box 13.5, this would result in different codes being used in place of the thematic codes to reflect the actions or interactions that occurred; for example: ‘avoiding compulsory redundancy’, ‘requesting voluntary redundancy’ and ‘discussing issues’.

Using a Grounded Theory strategy may mean that your initial research question is broadly focused, although still within manageable exploratory confines. The need to understand meanings and to generate codes to encompass these, means you are likely to conduct your early analysis by looking at smaller rather than larger units of data. Coding your data to identify analytical concepts and categories will help you to consider where data collection should be focused in the future (theoretical sampling). It may also help you to develop the focus of your research question. The initial multitude of code labels will need subsequently to be compared and focussed into broader, related groupings or categories as you develop the analytical process. As you develop a narrower focus through this process, you will be able to refine and limit the scope of your research question.

Focused coding

In Charmaz’s (2014) approach, **focused coding** involves deciding which of your initial codes to use to develop the analytic and explanatory focus of your coded data. This results in a smaller number of codes being attached to larger units of data, serving the same purpose as searching for themes in Thematic Analysis (Section 13.6). Data from various initial codes are re-coded to a smaller number of more focused codes. During initial coding some of the codes you develop may appear to have greater analytic potential, to help you to explain your data and to develop a grounded theory related to your research question. Selecting these codes will lead you to work through all of your coded data again to see if they are suitable to begin to develop a more explanatory focus. Charmaz suggests that codes with the capability to become focused codes, and categorise larger units of your data, are likely to be those that proved the most important or used frequently during initial coding. It is worth noting that codes that are frequently used during initial coding, may not always prove to have the greatest analytical potential, just as codes that become important may not initially attracted large amounts of data.

Progressing from initial coding to focused coding is unlikely to be a simple, linear process (Charmaz 2014). Working out and working through which initial codes may be the best ones to use as focused codes may lead you to re-code your data and develop a new set of codes. If this occurs, do not despair: it will take time, but will allow you to get closer to and understand your data with greater insight. Such reflection and re-working may occur irrespective of which qualitative analytic technique you choose. As you gain insights about what your data mean, you should use these to evaluate which codes will have the analytical capability to become focused codes to progress your analysis. These conceptually more useful focused codes should allow you to code and compare data across different interviews and observations. You will develop your analysis subsequently by constantly comparing the codes you are using to categorise your data with the data you have collected and work towards an emergent explanation of what your data mean to you.

Charmaz’s (2014) approach to the Grounded Theory Method can be seen as less prescriptive and less tightly defined than other approaches. She adopts a constructivist approach, which assumes that people construct their social realities, with both the participants’ and the researcher’s interpretations being socially constructed. Analysis develops by constantly comparing data to codes and codes to data, codes with other codes, and data with other data to develop higher levels of abstraction rather than necessarily using axial coding or

selective coding. Analysis is shaped by the researcher's interaction with and interpretation of these constant comparisons.

Axial coding

Axial coding is a way of rearranging the data that were fragmented during open or initial coding into a new whole, based on a hierarchical structure. It refers to the process of looking for relationships between the categories of data that have emerged from open coding as a process of theoretical development. As relationships between categories are recognised, they are rearranged into a hierarchical form, with the emergence of subcategories. The essence of this approach is to explore and explain a phenomenon (your research topic) by identifying what is happening and why, the environmental factors that affect this, how it is being managed within the context being examined and the outcomes of action that has been taken.

In some Grounded Theory Method prescriptions, axial coding may involve identifying structural elements such as the situation involved, the issue at the centre of this situation, the interactions that took place and the outcomes or consequences of these actions to develop a hierarchical structure. This approach may be appropriate where you wish to use a prescribed analytical framework to develop your analysis such as Strauss and Corbin (1998). But Charmaz (2014) believes that some will find it too prescriptive and will prefer to use a simpler, more flexible approach. Here, her use of initial coding and focused coding, combined with the use of theoretical sampling, constant comparison and theoretical saturation, is likely to prove more suitable.

Once these relationships have been recognised, you will then seek to verify them against actual data that you have collected. Strauss and Corbin (1998) recommend that you undertake this by formulating questions or statements, which can then be phrased as hypotheses, to test these apparent relationships. As you undertake this process you will be testing these hypotheses by looking for both supporting evidence and negative cases that demonstrate variations from these relationships.

Selective coding

After a lengthy period of data collection, which may take several months, you will have developed a number of principal categories and related subcategories (Strauss and Corbin, 1998). The stage that follows is called **selective coding**. This identifies one of these principal categories, labelled the 'central' or 'core' category, and relates the other principal categories to it, integrating the research and developing a grounded theory (Corbin and Strauss 2015; Strauss and Corbin 1998). In the previous stage the emphasis was placed on recognising the relationships between categories and their subcategories. In this stage the emphasis is placed on recognising and developing the relationships between the principal categories that have emerged from this grounded approach in order to develop an explanatory theory.

Evaluation

A number of implications have emerged from our brief outline of the main procedures involved in the use of grounded theory. These may be summed up by saying that the use of the Grounded Theory Method will involve you in processes that will be time-consuming, intensive and reflective.

Before you commit yourself to the Grounded Theory Method, you need to consider the time that you have to conduct your research, the level of competence you need,

your access to data and the logistical implications of immersing yourself in such an intensive process. There may also be a concern that little of significance will emerge at the end of the research. The Grounded Theory Method has the scope to provide you with a systematic analytical technique where you wish to use an emergent research approach. It is part of a wider strategy that you can follow to guide your research project from its inception, through the processes of data collection and analysis, to completion. The theory that you develop should be well grounded in the meanings expressed by your participants and the context of the research setting. Successful application of this approach is related to ensuring you understand one or other of the published versions of the Grounded Theory Method and you are following its procedures exactly as specified.

13.10 Narrative Analysis

Overview

Narrative Analysis is a collection of analytical approaches to analyse different aspects of narrative while preserving the data in its narrative form. These may be combined in practice, depending on your research question and purpose, and the nature of your data. Unlike Thematic Analysis, Template Analysis or the Grounded Theory Method, where original data are fragmented by coding and then assigned to analytical categories, narrative data are preserved and analysed as a whole unit or narrative sequence. Categories, themes and facets of content may still be identified and coded, but this occurs within their narrated context, maintaining the sequential and structural elements.

While a narrative tends to be analysed as a whole, the nature of what constitutes a narrative varies considerably. Narratives may vary from a segment of text or speech to a whole life story provided by a narrator. Analysis may focus on extracts from interview transcripts, which each provide a short narrative about a related topic or incident in which the researcher is interested. These tend to be short stories that have a clear purpose, encompassing a situation, action and outcome, expressed in a structure that has a beginning, middle and end. Analysis may also focus on passages of speech or dialogue, to analyse how the narrative is constructed. For more extended narratives, analysis can focus on accounts of life stories or organisational events, placing emphasis on sequential and structural elements. Analysis may also involve you constructing a narrative from fragments of data collected from multiple sources, such as different documents or research interviews. It can also be constructed from other narratives to provide a unified account for further analysis, sometimes referred to as re-storying.

Because Narrative Analysis is a collection of analytical approaches, with variations evident in the ways they have been used in practice by researchers, it is not sensible to describe a procedural outline as we have done in earlier sections. Rather, we briefly outline two forms: Thematic Narrative Analysis and structural narrative analysis (Maitlis 2012; Riessman 2008).

Outline of thematic narrative analysis

Thematic narrative analysis identifies analytical themes within narratives focusing on the content of a narrative, rather than on the way in which it is structured. It therefore emphasises ‘what’ the narrative is about rather than ‘how’ it is constructed.

Thematic narrative analysis (Box 13.14) can be used to analyse an individual narrative or multiple, related narratives. In either approach, you need to pay attention to the chronological sequence and contextual background of themes you identify. Understanding sequence and context is important to be able to develop a rich and full explanation when analysing an individual narrative. Analysis of multiple narratives can commence by analysing each narrative separately or by working across all the narratives at the same time. Multiple narratives will be related by a common focus, such as professional calling (Box 13.14) or an organisational event, with each narrative provided by a different participant. In analysing multiple narratives separately, initial emphasis will be on in-depth analysis before comparing and contrasting findings across narratives. Analysing multiple narratives individually can reveal how variations in context affect the actions taken and outcomes recorded, or how differences in the actions taken and outcomes recorded may vary despite contextual similarities.

Analysis of multiple narratives can also commence by searching for and coding themes across these narratives. This is more suitable where you commence your research deductively with a predetermined theoretical framework of analytical categories or themes for which to search. In this approach, you identify those themes that occur across some or all of the narratives, where variations occur, and how contextual factors affect these. This will help you to develop an explanation that evaluates the applicability of existing theory



Box 13.14 Focus on management research

Stories of calling: How 'called' professionals construct narrative identities

In an article published in *Administrative Science Quarterly*, Bloom et al. (2021) analysed the narratives of 236 individuals in four professions to discover how they incorporated life events into their stories that support their identity claims. The professions were pastors, physicians, international aid workers and teachers.

Bloom and colleagues' approach to narrative inquiry research involved conducting and audio recording semi-structured narrative interviews to gather life stories, which were then transcribed. Initial transcripts were re-read in their entirety while also listening to the audio recording, with non-verbal expressions being noted where appropriate. Each transcript was then coded using open coding, coding stories that were narrated as particular events and personal memories. Wherever possible in vivo codes were used to remain close to the participants' language.

Over time, as more transcripts were analysed, common interpretations of transcripts emerged between the researchers. Second order coding focussed on interpreting of participants' stories, seeking to elucidate recurring themes, what these meant to participants and what they revealed about how participants came to experience work as a calling. The researchers also developed their conceptual understanding of what the chronological ordering of participants' stories and the underlying themes revealed. When the interviews were no longer revealing new kinds of stories or chronologies they stopped recruiting new participants.

Bloom and colleagues found that participants identified their callings in two distinct ways: discernment and exploration. The first of these, 'discerners', journeyed to their professional destiny which was their calling. In contrast, 'explorers' searched actively for work that they loved, destiny playing no role. Once participants in each group had identified their calling, both groups of participants sought to demonstrate their legitimacy through their mastery and receiving affirmation, whilst also crafting personal authenticity in how they enacted the role.

to your data as well as being grounded in these data, while preserving the integrity of your narratives.

Analysing narratives to identify themes while keeping each narrative intact can use the method of coding we discussed earlier in this chapter. One adaptation you might use is to colour-code analytical themes in each narrative. Through this you will be able to identify its occurrence across different narratives, without fragmenting these data. A further adaptation that you may find useful to keep your narratives intact is to make several electronic copies of each set of narratives and to code a particular theme on one set of copies.

Outline of structural narrative analysis

Structural narrative analysis analyses the way a narrative is constructed by examining use of language to understand how it affects a listener or an audience. The emphasis is therefore on ‘how’ the narrative is constructed and language is used rather than ‘what’ it is about.

While thematic narrative analysis is likely to be easier to use and therefore to be used more often, structural narrative analysis can add further insights. To use this approach you will need to develop some understanding of the socio-linguistic and cognitive theories that underpin it (see Riessman 2008). These have led to methods to analyse the structures of spoken narratives. A widely used way is to analyse how narrative accounts are sequenced and structured using the technique developed by Labov and Waletzky (1967) and Labov (1972). In this you look for the presence of six elements and the way these have been used:

- an abstract (which states the point of the story);
- an orientation (which describes the situation including when and where it took place and who was involved);
- a complicating action (which describes the sequence of events including a critical point);
- an evaluation (where the narrator explains the meaning of the narrative);
- a resolution (how the issue is solved – the outcome); and
- a coda (which ends the narrative and relates it to the present).

This provides a framework to evaluate narratives, recognising not every element may be present in a narrative and the nature and sequencing of these elements are likely to vary. It is, however, worth noting that the purpose of much of the research undertaken using this and other techniques to analyse the structure of narratives is to understand how people in different groups form narratives, rather than form judgemental evaluations. This can include how acts of speech may lead to certain actions or to falsely negative perceptions; sometimes to change professional practice.

Structural narrative analysis can also be used to analyse interactions between individuals to understand the relationship between the construction of a narrative and its effect on the attitudes and subsequent actions of those who receive it. This can encompass a wide range of interactions; for example, between managers and other employees; across occupational groups; between organisational levels; across cultural and transnational boundaries. More generally, structural narrative analysis may be suitable for you to analyse the narratives you collect through conducting interviews or recording naturally occurring conversations.

Evaluation

We noted that adopting a narrative inquiry strategy may be advantageous in certain circumstances (Section 5.5). These include contexts where the experiences of your participants

can best be understood by collecting and analysing these as complete stories or narrative sequences. The ways in which events in a narrative are linked, the actions that follow and their implications are more likely to be revealed by encouraging a participant to narrate their experiences by asking them to respond to a series of pre-formed questions. Narrative analysis allows chronological connections and the sequencing of events as told by the narrator to be preserved, with the potential to enrich understanding and aid analysis.

13.11 Discourse analysis

Overview

Discourse analysis explores how discourses construct or constitute social reality and social relations through creating meanings and perceptions. It covers a variety of approaches that analyse the social effects of the use of language. ‘Discourse’ refers to the spoken or written use of language, often referred to as talk or text. In discourse analysis, the emphasis is on studying how language is used to shape this meaning-making process, to construct social reality. A **discourse** is therefore not just seen as neutrally reflecting social practice or relations but as constructing these (although the notion of ‘constructing’ is contentious and we return to it later). In this way, **Discourse Analysis** explores how discourses construct or constitute social reality and social relations through creating meanings and perceptions.

This conceptualisation allows the complexity and diversity of social practice and relations to be recognised through the existence of different, often competing and sometimes conflicting discourses. Different discourses construct perceptions about organisations and organisational relations. Language (discourse) can be used intentionally to attempt to create ideologically mounted positions, intended to be in the interests of those who produce and disseminate them. A unitarist view would emphasise the commonality of interest within an organisation (or society) and use some means (focusing on discourse) to persuade its members of this approach. By contrast, a pluralist view would see an organisation (or society) as a collection of competing interests. Even within the pluralist view, some discourses may be seen to dominate while others are marginalised.

Discourse analysis involves studying textual sources or passages such as organisational documents as used in archival or documentary research (Section 5.8). It will often involve using multiple texts that are interrelated to understand the nature and development of a discourse. The diffuse, interactional and often taken-for-granted nature of a discourse means that although it cannot be explored comprehensively, by using a range of interrelated sources it should be possible to gain access to aspects of its formation, propagation and acceptance (Phillips and Hardy 2002).

Transcripts of naturally occurring talk can also be used to explore a discourse. Such data may be collected through observation as part of an ethnographic strategy (Section 5.5). As such discourse occurs in naturally occurring talk, it is preferred to contrived talk through interviewing (Chapter 10) where the intervention of the researcher in asking questions, eliciting responses and analysing the data is likely to affect the authenticity of the discourse being analysed (Hepburn and Potter 2007). There may, of course, be a use for interview data in a subsequent, supplementary capacity.

The way a discourse emerges and constructs social reality through influencing social relations and practices is likely to be rooted in a particular period or event, such as the foundation of an organisation or an organisational change. Discourse analysis may therefore require an understanding of historical context. Some approaches to also draw on existing theoretical perspectives to explore the nature of a discourse and to contextualise its impact on social practice and relations.

Outline

Discourse analysis encompasses a range of approaches and does not specify a particular set of techniques to conduct analysis. The focus of ranges from ‘finely-grained’ analysis of text or talk to grand theoretical abstractions about the nature of social practice. A finely grained approach focuses on the analysis (deconstruction) of an individual text, or of a transcript of ‘talk’ that occurred during a social interaction located within a particular situation. The purpose of close reading of a text (or passage of talk) is to understand how the use of language indicates meaning and to categorise the nature of this discourse. Hyatt (2005, 2013) provides advice about conducting this type of analysis. His ‘Critical Literacy Analysis’ (2005) and ‘Critical Policy Discourse Analysis’ (2013) include a range of criteria for analysing text. Although these analyses are devised within the context of education, the generic analytical criteria are transferable or translatable to other contexts. If you are considering using discourse analysis you may find it useful to consult these articles.

Interdiscursive and intertextual analyses

Further (and complementary) approaches include interdiscursive and intertextual analyses. **Interdiscursivity** refers to the relation of one discourse to another, including the way one discourse may influence another discourse. For much of the Covid-19 pandemic many governments justified their responses with a discourse of ‘following the science’. Box 13.15 provides an example of this in the context of the USA’s roll-out of their Covid-19 booster shots programme and the World Health Organization’s (WHO) response.

Intertextuality refers to the way a text or texts overtly or covertly borrow from and are informed by other texts. Overt borrowing from another text is acknowledged through use of quotations and citations. Covert borrowing involves adopting ideas or ideological positions and arguments from other texts without overtly acknowledging this. These focus on how discourses and texts are used in the construction of other discourses and texts, to identify how discourses change and develop, and to understand how attempts are made to give credibility to such changes or developments. Interdiscursive and intertextual approaches to analysis emphasise the importance of contextual knowledge in understanding how discourses develop and evolve, and in appreciating factors that can bring about change. Using these forms of analysis therefore involves using multiple texts.

Our discussion so far has emphasised the role of social constructionism in discourse analysis. By this we mean the assumption that the social world is socially constructed through discourse and that discourse analysis analyses how use of language constructs versions of social reality (including dominant, marginalised and competing discourses). However, the extent to which social reality is socially constructed is contested. To this end, Holstein and Gubrium (2011: 342) reflect a dictum of Karl Marx in saying ‘that people actively construct their worlds but not completely on, or in, their own terms’. This points to the (ontological) distinction between objectivism and subjectivism (Section 4.2). According to realist philosophical positions, objective entities exist that are external to social actors, which impact on their social constructions. It is therefore important to understand external factors that affect human attitudes and actions, whether or not social actors are aware of these influences on the ways in which they make sense of their social world.

Critical discourse analysis

Critical discourse analysis adopts a critical realist philosophy (Section 4.4), drawing a distinction between the natural world and the social world, with the implication that social



Box 13.15 Focus on research in the news

US begin to offer Covid booster shots from next month

By Lauren Fedor and Mamta Badkar

The Biden administration plans to start offering Americans a third dose of mRNA Covid-19 vaccines next month amid a wave of new cases tied to the Delta coronavirus variant, after government scientists concluded booster shots were needed to stem another outbreak.

In a joint statement published on Wednesday, the heads of the US government's leading health agencies, including the Centers for Disease Control and Prevention, the Food and Drug Administration and the National Institutes of Health, said Washington was 'prepared to offer booster shots for all Americans' eight months after their second dose. The boosters would become available from September 20.

The decision comes as the US battles a fresh wave of Covid cases, with the more contagious Delta variant spreading rapidly in several states, particularly among people who have not received any jabs. The health officials, including the US surgeon general Vivek Murthy and Anthony Fauci, Joe Biden's chief medical adviser, said those at the front of the queue for boosters would include healthcare providers, nursing home residents and other senior citizens who had been among the first to receive jabs.

'Our top priority remains staying ahead of the virus and protecting the American people from Covid-19 with safe, effective and long-lasting vaccines, especially in the context of a constantly changing virus and epidemiologic landscape,' the health officials said.

'We will continue to follow the science on a daily basis, and we are prepared to modify this plan should new data emerge that requires it.' The World Health Organization questioned the US decision, saying the current data 'does not indicate that boosters are needed'.

Mike Ryan, executive director of the WHO's health emergencies programme, criticised the US for allocating third jabs for Americans when so many other countries were struggling with a lack of vaccine supply. 'The reality is, right now, today, if we think of this in terms of an analogy, we're planning to hand out extra life jackets to people who already have life jackets, while we're leaving other people to drown without a single life jacket,' Ryan said. 'That's the reality.'

Jeff Zients, the White House coronavirus response co-ordinator, pushed back against the criticisms, telling reporters that in June and July, the US administered 50m shots to Americans while at the same time shipping over 100m doses to other countries. Zients said the US intended to donate more than 200m additional doses to the rest of the world in the coming months. 'To end this pandemic, we have to protect the American people, and we have to continue to do more and more to vaccinate the world. Both are critical.'



Source: Abridged from 'US begins to offer Covid booster shots from next month', Lauren Fedor and Mamta Badkar (2021) *Financial Times*, 18 August. Copyright © The Financial Times Ltd

actors' understanding of the social world is affected by the natural world and is not entirely socially constructed. Fairclough (2010: 4–5) captures this when he writes,

The socially constructive effects of discourse are thus a central concern, but a distinction is drawn between construal and construction; the world is discursively construed (or represented) in many and various ways, but which construals come to have socially constructive effects depends upon a range of conditions which include for instance power relations but also properties of whatever parts or aspects of the world are being construed. We cannot transform the world in any old way we happen to construe it; the world is such that some transformations are possible and others are not. So CDA is a 'moderate' or 'contingent' form of social constructionism.

Critical discourse analysis examines relations between discourse and other objects in the world that are recognised as existing, including the exercise of power by those who control resources. Discourse is seen as being affected or conditioned by social reality, knowingly or unknowingly, as well as socially construing it. As a result, it incorporates the need to not only analyse incidents of discourse (analysis of social interactions or text) but also to understand how wider discursive and social practices influence and are influenced by discourse. An incidence of discourse is 'simultaneously a piece of text, an instance of discursive practice, and an instance of social practice' (Fairclough, 1992: 4) (Figure 13.3). The approach involves analysing discourse at the level of text or social inaction, discursive practice – including the use of interdiscursive and intertextual analyses, and social practice to achieve an integrated and critical understanding.

Evaluation

Discourse analysis offers a valuable analytical approach where your research involves social action and interaction within a particular setting such as an organisation. This can include topics such as organisational communication, culture, decision-making, governance, power, practices, processes, relations or trust. It may therefore provide you with an insightful means to analyse data resulting from the use, for example, of an Action Research,

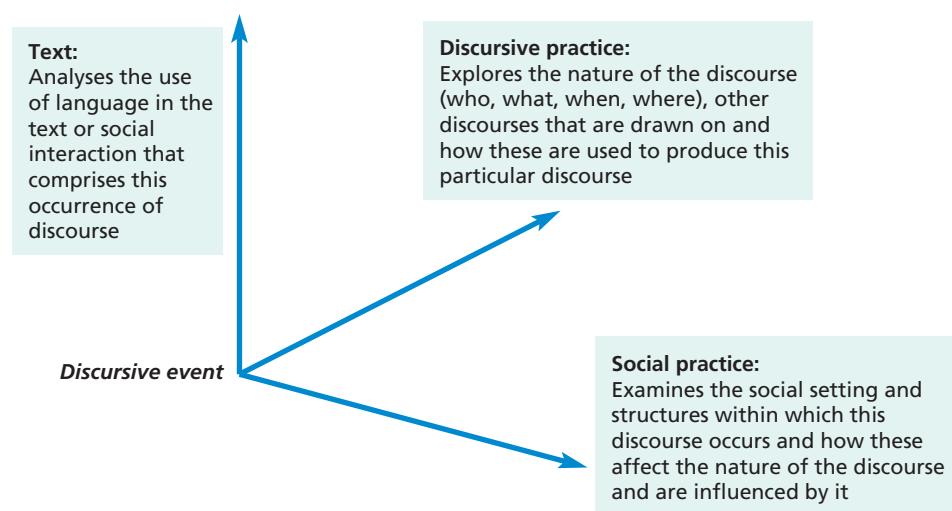


Figure 13.3 A three-dimensional analytical framework for critical discourse analysis
Source: Developed from Fairclough (1992)

documentary or ethnographic strategy where you have transcripts relating to the use of language in discourse. Discourse analysis can be used as your primary analytical technique, depending on your research question, research design and nature of your data, or in support of other analytical techniques where appropriate.

If you consider using discourse analysis, you will need to develop greater familiarity with, and understanding of, the approaches. In particular you will need to be able to explain and justify your choice of approach in relation to your research. Discourse analysis, like some other analytical approaches, has developed to suit several purposes, incorporating different philosophical and theoretical assumptions suitable for different types of data and using different methods. Discourse analysis therefore offers a potentially valuable research approach but consideration about using this approach will benefit from adequate and early preparation!

13.12 Visual analysis

Overview

Visual analysis makes use of existing visual images (Section 8.2), known as ‘found’ images, and images created by the researcher or research participants (Sections 9.6, 9.7 and 10.11). They may be categorised as static, such as photographs and drawings; or moving, such as video, film and television. These images provide a potentially powerful ‘way of seeing’ and means of gaining new insights or perspectives; as well as an effective way to record data (including contextual data) whose breadth and depth might otherwise be difficult to encapsulate and describe in a time constrained research situation. As a representational form in their own right, images can both complement and enhance textual description and analysis. However, we also recognised earlier that analysing visual data is associated with particular analytical implications that make this a problematic research method.

All images need to be interpreted, with implications at different levels. At the practical level, interpretation involves using words to describe and analyse an image, with the consequence in many cases that the image is displaced, leaving us solely with and entirely reliant on the interpretation offered in the textual description. At the explanatory level, this interpretation will only be one of many possible. Pink (2007) says that the interpretation of an image will depend on the subjectivity of the interpreter and that any interpretation is likely to be contextual and time specific. Other people in different situations and at other times may offer different interpretations.

We may also question what the image represents. While images may be viewed superficially as just a way of making things visible, they are constructed or created by someone for an intended purpose or from a particular perspective, whether this is explicitly understood or not. The use of images to capture or reveal an underlying reality is therefore contested. From this perspective, images may be constructed to justify the assumptions and practices of those who commission their creation. Such images will be carefully composed or constructed to achieve this desired effect. For others, images may spontaneously capture attributes of a situation, suggesting a view into a situated reality (Pink 2007; Rose 2016). However, even where an image is apparently taken spontaneously in a natural setting, it will still be affected by its framing, depth of focus and point of view, with the consequence that different images may create different effects.

Where you use found images or those created by others, you will find yourself in the situation of making a subjective interpretation of another person’s subjective representation. In doing this you will also be creating a word-based interpretation of a visual representation. Without sufficient cultural or contextual understanding this will be problematic. To undertake this task less problematically, you will need to understand:

- who took or created the image or images that you wish to analyse;
- how these images were taken or created;
- the purpose for which these images were taken or created;
- whether these images have been used previously and, if so, how;
- any intended audience for these images, and;
- the intended effect(s) of the image maker and those who commissioned these images.

Achieving this depth of insight will necessitate you engaging reflexively (Section 1.5) during your research, where you not only seek to understand how your own preconceived ideas might influence the way you interpret visual images but also to recognise how multiple interpretations may result from the intended and unintended effects of the images you use, and from different perspectives. In recognising multiple interpretations and different perspectives, you need to make these explicit in your analysis and discussion.

Analysing images as visual data in their own right

Where images are analysed as visual data in their own right a number of analytical techniques may be used, drawn from a diverse range of disciplinary domains (Bell and Davison 2013; Rose 2016). Some techniques are data based and inductive, while others use an existing theoretical perspective through which to focus analysis. In this sub-section we briefly outline two approaches to analyse visual images. These are content analysis and semiotic analysis.

Content analysis

Content analysis is used to analyse large numbers of images and involves quantification of the visual data you derive from the images you analyse. This approach involves (Rose 2016):

- identifying categories of visual data in which you are interested;
- developing a systematic coding scheme;
- coding your visual images using this scheme;
- undertaking quantitative analysis.

If you decide to use content analysis you will need to be aware of its precise analytical rules and procedure, to which you will need to adhere (Section 12.3). Subsequent quantitative analysis ranges from calculating the frequency of different categories for a variable (Section 12.8) to examining relationships between variables created (Section 12.10).

Semiotic analysis

Semiotic analysis is the analysis of signs and their meanings in relation to the social world and social processes. A **sign** is ‘something’ (a word, phrase, sound, cultural artefact or visual image) that represents something other than itself such as a culture, warning, feeling or brand. The inclusion of ‘something’ twice in this definition is intentional indicating it consists of two parts. These are the signifier and the signified. The **signifier** is the word, phrase or sound used, or image or artefact shown and the **signified** is the concept or meaning suggested or implied. A simple and relatively straightforward example here might be the way in which images of some animals, birds and mythological creatures have been used in branding and advertisements signify power, strength, dependability, or wisdom. There is, however, no automatic relationship between a signifier and what is signified. Any meaning derived will depend on the context, and the conventions held by those who see and interpret the signs, be moderated by cultural differences, and assessed in relation



Figure 13.4 Person with folded hands emoji

Source: ya_blue_ko/Shutterstock

to other, related signs. For example, the folded hands emoji (Figure 13.4), although not usually explicitly religious in meaning, can mean ‘thank you’ or alternatively indicate ‘pleading’, ‘prayer’ or two people ‘high fiving’ each other.

Approaches to semiotics propose different ways to analyse signs. These include the:

- form a sign takes (iconic, indexical, symbolic);
- meaning a sign conveys (denotative, connotative);
- relationship between signs (syntagmatic, paradigmatic).

The form a sign takes is based on the work of the American philosopher Charles Sanders Peirce, whose work on pragmatism we refer to in Section 4.4. This differentiates between iconic, indexical and symbolic signs. An **iconic sign** is one where the signifier resembles the object being signified. Using the example of road signs for tourist attractions, the UK’s Department for Transport (2015: 100–102) use a range of icons to signify types of destination (Figure 13.5). For example, the location of a historic house is signified by the



Figure 13.5 UK road signs for tourist attractions

Source: Designalldone/DigitalVision Vectors/Getty Images

image of a large house, the location of a castle by an icon that clearly resembles a castle and so on! An **indexical sign** inherently indicates the object being signified by showing evidence of what it represents such as the sign for a farm trail. This does not use an iconic representation but, instead, shows the image of a heavy horse to signify a farm. Similarly, the location of a zoo is signified by the image of an elephant and that of a football ground by a football.

Signs are often multimodal and roadside tourist signs are also a good example of this. In the UK roadside tourist signs have a brown background and often include text. The use of colour helps to signify the purpose of the sign and the use of text helps to anchor what is being signified. While none of the examples used so far include text, others combine a sign and text on a brown background to anchor meaning. A **symbolic sign** is more abstract having no resemblance to what is being signified and must be culturally learned to be capable of signifying meaning to those who see it. For example, traffic signals in many countries are based on the conventional use of three colours, red for stop, green for keep moving, and amber for stop when it is safe to do as the light will turn red soon.

The meaning a sign conveys can be classified as denotative and connotative. In a **denotative sign** the meaning being suggested or implied is literal and will be reasonably obvious or visible. For example, an image of a hotel bedroom on a hotel chain website will signify the quality of the furnishings and the amount of space in the room. A **connotative sign** is either a substitute for or a part of the thing it stands for signifying a secondary meaning. In many people's minds an image of a sunlit seashore and perfectly blue sky is associated with being on holiday and such an image may be shown on a holiday company's website to stand for a relaxing holiday, while an image of a table laid for dinner with a white cloth, silver cutlery, wine glasses, candles and flowers can represent a high-quality dining experience. An individual sign may also signify both denotative and connotative qualities.

The relationship between signs analyses the ways signs work with one another. In this respect, semiotic analysis may be syntagmatic or paradigmatic. **Syntagmatic analysis** explores relations between signs and the ways in which meanings are signified as different signs are combined into structures or sequences in visual images. In a static image such as a printed advertisement, signs become meaningful in relation to other signs that surround them in the advert. The intention of an advertisement will be to transfer the signified meanings from one or more signs to other signs related to the product, so that, for example, signs associated with health and well-being may be projected in an advertisement for a food product that wishes to be seen as healthy and good for you. Alternatively, participant produced images can be used to explore the meanings they give to a specified activity, such as entrepreneurship. In moving images such as video and film, signs occur sequentially being meaningful in relation to those that occur before and after them. Video advertisements such as television adverts may therefore be analysed similarly to explore how signs are used to signify meanings in relation to the product being promoted.

Rather than focusing on structural or sequential relations between signs, **paradigmatic analysis** explores relationships between signs by examining how the substitution of alternative signs for one sign will alter that sign's signified meaning in relation to other signs. Using our example of the advertisement for a healthy food product, if those who commissioned this advert wanted to project the idea of fitness and bodybuilding in relation to consuming this food instead of health and well-being, the signs used to signify this intended meaning would alter and consequently affect the market for this product.

Attempting to comprehend and analyse signified meanings and their relations to other signs is likely to be difficult. This will be due to the complex or abstract nature of some signs and the complexity involved when such signs are used in relation to one another. Complexity leads to the likelihood of multiple meanings as signs are interpreted. Semioticians use a term for this: **polysemy**, or multiple meanings, and a sign with more than one meaning is polysemic. However, while many signs and complex ones in particular are polysemic, their use and interpretation will be influenced by wider cultural and social conventions (such as red indicates stop). These conventions refer to wider systems of meaning that reflect shared understandings and expectations. For example, identifiable groups such as accountants, entrepreneurs, human resource practitioners, marketing professionals and public relations practitioners will each share a set of conventionalised understandings that are referred to as a **code**, affecting the way members of that group understand their professional world, how to interpret its signs and behave within it. More broadly, codes of conventionalised understandings will operate at the societal level, affecting the ways in which signs are used and interpreted. Such codes will be influenced by prevailing **ideologies** or ways of thinking, known as **dominant codes** (Rose 2016). However, while dominant codes will underpin the way in which many signs are used with the intention of producing an intended or preferred meaning, other interpretations will still exist and some of these will encompass a critical perspective about the use of signs to promote ideological interests.

Semiotic analysis can help you analyse and interpret visual images, but its strengths and issues need to be recognised. Rose (2016) expresses these succinctly. In relation to its strengths she says that, ‘A semiological analysis entails the deployment of a highly refined set of concepts that produce detailed accounts of the exact ways the meaning of an image are produced through that image’ (2016: 107). In relation to its issues she says that the very richness of its analytical concepts can appear to be terminologically dense while, ‘for all its analytical richness, semiology does not offer a clear method for its application’ (2016: 110). She does though provide her own outline to use this analytical approach, which we have developed as a checklist in Box 13.16.

Analysing images as visual representations

Images may also be used to represent analytical aspects and evoke meanings that would otherwise be difficult to convey in a research project (Box 13.17). In this approach, selected images and text are combined in a research report, to enhance one another’s ability to represent perspectives that would be difficult to describe using words alone.



Box 13.16 Checklist

For semiotic analysis

- ✓ Identify the signs in an image.
- ✓ Assess what each sign signifies.

- ✓ Analyse how these signs relate to one another using the concepts outlined in this sub-section.
- ✓ Explore how these signs relate more widely to systems of meaning such as codes and ideologies.
- ✓ Evaluate the use of these signs in relation to this wider interpretation.

Developed from Rose (2016)

Where photographic images and text are combined in this way this is referred to as a **photo essay**. In a video format, voice-over narration may be used to achieve a similar effect, referred to as a **video essay**. For some research reports it may be permissible to create a digital document or multimedia website and integrate both static and moving images with text. However, it is worth checking your assessment criteria before embarking on this form of project report.

To produce a photo essay you need to consider a number of aspects. Your research philosophy, research strategy and approach to theory development will be likely to shape the nature and purpose of the photo essay you create. For example, your research may be



Box 13.17 Focus on student research

Symbolic branding of tourist destinations

Sarah was interested in how tourist destinations used images to brand themselves. In reviewing the literature she discovered that, despite an increase in destination branding research, few studies had focussed on how towns and villages branded themselves (Rowley and Hanna 2020). She decided to focus on visual branding of seaside towns and how signs (symbolic representations) were incorporated in their advertising and used at the actual destination. Having undertaken an online analysis of the use of symbolic representations in their branding by seaside towns; she had visited three towns with distinct brand identities and taken photographs of the use of symbolic representations in the town. At the end of this, she selected a few her images from each town to produce a photo-essay, which she subsequently incorporated in her project report. This photo-essay combined text with her digital images.

Her digital images illustrate how symbolic representations (signs) were incorporated into the seaside townscape. One of these locations was Lyme Regis, a tourist destination on the south coast of England, which had long been famous for its association with fossils, a variety of images of an ammonite fossil appearing prominently in much of its advertising. A selection of the images and associated text from her photo-essay for this location are reproduced here.



Lamppost

On arrival in Lyme Regis I was struck by the considerable use of representations of the ammonite fossil that had appeared as a logo or image in many of the web pages related to this destination. A stylised ammonite fossil was the central feature of the lampposts, both on the promenade and in the Langmoor and Lister Gardens, reminding the visitor of the town's links to fossils and fossil hunting. I took this photograph to illustrate both the stylised nature of the ammonite incorporated into the lampposts and its prominence. There were also more subtle uses of the ammonite in the town. One often repeated representation of this is in my photograph of the one of the numerous metal railing posts, each of which incorporated a cast of an ammonite fossil cast. Although not immediately obvious I found that, once noticed, I saw the ammonite on every post.

The town's link to fossils and fossil hunting was emphasised further in the sculpture trail in the Langmoor and Lister Gardens. My photograph shows





Box 13.17 **Focus on student research (continued)**

Symbolic branding of tourist destinations



Ammonite on a railing post

an ammonite in giant form as the sculpture 'Ammonite Jurassic Cracked' by Hamish Mackie. As highlighted by

the associated plaque, this offers a more contemporary take on the ammonite and my photograph tries to capture the atmosphere this creates in the gardens' setting at sunrise. It also shows how it relates to the more stylised ammonite in the lamppost.



Ammonite Jurassic Cracked; sculptor: Hamish Mackie

Several other sculptures also highlighted the town's relationship with ammonites, fossil hunting and its importance in the history of palaeontology. These included one of the 19th Century Lyme Regis fossil hunter Mary Anning, whose life inspired Francis Lee's 2020 film "Ammonite".

Source of images in box: © 2022 Mark N.K. Saunders

concerned with revealing an underlying reality or it may be guided by a desire to visualise multiple realities and interpretations. Where your research commences deductively the images you use are likely to be influenced by existing theory. Where your research is inductive, the images you use are likely to be more exploratory.

A photo essay may be organised thematically or it may create a narrative account. In the first of these, a specific image, or set of images, will be included to illustrate a particular research theme (Section 13.6). These themes may be determined from the literature prior to creating images, or they may emerge through the process of collecting these. A photo essay that creates a narrative account is known as a **photo novella** and will introduce a storyline into the way photographic images are used in relation to one another. In a photo novella, images and text may be presented in a similar style to those in a comic book. Box 13.18 offers a series of questions to help you create your photo essay.

Where you decide to use images as visual representations you will be more likely to focus on theoretical perspectives. This will affect the way in which images and text are used in relation to one another and the balance between image and text. This is likely to involve you using longer explanatory and theoretical captions or statements in relation to



Box 13.18 Checklist

For creating a photo essay

- ✓ How will you arrange the images and text on a page?
- ✓ Will each image and its accompanying text be given equal prominence, or will you choose to give prominence to a core image and arrange others in a subsidiary way?
- ✓ How will you present the text in relation to the images?
- ✓ Will images or text be more prominent, or will these have equal prominence?
- ✓ Will you use images with only brief captions, or you may incorporate a few selected images into a largely textual account?

an image or set of images. For example, a longer caption or accompanying statement may explain who took the image, where it was taken, when it was taken and any significance related to the time at which it was taken, the nature and significance of any interaction or activity shown in the image, and how the image portrays an aspect or issue related to the image taker's experience or that of any person shown in the image (Box 13.17). More theoretically, accompanying captions or statements to images may explain the nature of each theme represented by an image, why each theme shown is significant, how themes relate to one another and how these may be integrated to generate a deeper theoretical insight. Used in this way, a researcher may create a photo essay with an explanatory and theoretical purpose (Pink 2007) (Box 13.17).

Photo essays are also produced in collaborative and participant-led visual research projects where participants analyse their own visual images and select those that represent their perspectives or experiences that are important to them (Section 10.11). Another participant-led narrative form that involves moving images is the video essay. In this, participants may produce a script for filming, record video, analyse and edit video footage and produce a video essay.

Evaluation

Visual data can be used as visual representations, rather than eliminating them during an early stage of analysis, where these are literally lost from view. While images such as some of those created by participants are deliberately and meaningfully used in visual interviews to elicit further data (Section 10.11), other images such as advertisements may be integrated in the analytical process through techniques such as semiotic analysis, while others that act as visual representations may be incorporated in the output from research through the production of photo essays.

This highlights that you will need to think through the purpose of using images, and related to this your research question, to be able to analyse these appropriately. A range of techniques exist to analyse visual images, but their suitability varies according to the research aim. Combined with the subjective nature of visual images, your choice of an appropriate analytical technique will be crucial. We recognised earlier in this section that the analysis of visual images should be undertaken methodically, reflexively and with a clear appreciation of your analytical purpose. To achieve this we recommend you:

- think carefully about the nature of your images and their suitability for enabling your research question to be answered and objectives met;

- explore possible visual analysis techniques that you might use and read further about each, as well as looking for published research that made use of these techniques;
- evaluate these possible analytical techniques and identify one or more techniques to analyse your images;
- use this technique or techniques in your analysis;
- evaluate your use of this technique or techniques as you undertake your analysis;
- if necessary, return to an earlier point in this list where your evaluation indicates the need to revise your approach.

13.13 Data display and analysis

Overview

Data display and analysis is an approach based on the work of Miles et al. (2019). For them, the process of analysis consists of three concurrent activity flows:

- data condensation;
- data display;
- drawing and verifying conclusions.

It is suited to an inductive approach to building theory, although it is also compatible with developing theory deductively.

Outline

Data condensation

Data condensation is the process of selecting, simplifying, abstracting and transforming the qualitative data you have collected from all sources. The activity is argued to make these data stronger and occurs throughout the research project. It commences before data are collected with decisions about the conceptual framework, research questions, sample selection and data collection methods. As data are collected, condensation occurs through analysis techniques such as production of interview or observation summaries, document summaries, coding and categorising data and perhaps constructing a narrative.

These activities continue until the project report is submitted and should be seen as a form of analysis that ‘sharpens, sorts, focusses, discards and organizes data’ allowing conclusions to be drawn and verified (Miles et al. 2019: 8). Within this it may, occasionally, to quantify these qualitative data by comparing relative sizes, in effect undertaking some quantification.

Data display

Data display involves organising and assembling your data into summary diagrammatic or visual displays to help understand what is happening. The simplest form of data display is your transcript, but such extended text on its own is cumbersome to process. Miles et al. (2019) outline a number of ways of displaying data referring to three main families of data display which organise data into a more compact form so you can see what is happening and are often available in CAQDAS. These comprise matrices, networks and other graphic displays. Matrices are generally tabular in form, with defined columns and rows, where

data are entered selectively into the appropriate cells of such a matrix, to facilitate further data analysis (Box 13.19). Networks are collections of codes (often called nodes in CAQDAS software) or boxes that are joined or linked by lines, perhaps with arrows to indicate relationships. The boxes or codes contain brief descriptions or labels to indicate variables or key points from the data. Graphic displays include a variety of graphs such as bar charts and diagrams often used with quantified data such as word clouds (Sections 12.6, 12.7).

Displays allow you to make comparisons between elements in your data and identify potential relationships, key themes, patterns, and trends that may be evident. These will



Box 13.19 Focus on student research

Using CAQDAS to explore how key words are used in context

Marcus' research was concerned with how staff were responding to the managed changes in the organisation where he worked. He had collected his data using an online questionnaire, which contained the open question: 'If there is anything further you would like to add in relation to the changes at OrgCo, please type your comment in the box.' Noting that some of the responses to this question were over 500 words in length, Marcus downloaded all of them verbatim from the online survey tool into his word processor.

He then spellchecked them, correcting words that had been misspelled or used American spellings to the English spelling. This ensured he would pick up all occurrences of particular words such as 'staff' or 'OrgCo', the pseudonym he used to anonymise the organisation. He then loaded the spellchecked data into QSR International's CAQDAS software NVivo. Marcus wanted to see which respondents had mentioned staff or staffing in their responses to the question and the context in which the words had been used. He therefore searched for both exact matches of the word 'staff' and stemmed words such as 'staffing' within his data.

Scanning the responses suggested that those respondents who had answered this question often appeared to talk about how staff were treated at OrgCo. Marcus decided to investigate further.

| Reference | Coverage | Text Context |
|-------------|----------------|--|
| Reference 4 | 0.01% coverage | so that skilled and experienced staff can be retained. I am |
| Reference 5 | 0.01% coverage | car park AND bring more staff (some of whom used to |
| Reference 6 | 0.01% coverage | policy but how they treat staff only pays lip service i.e |
| Reference 7 | 0.01% coverage | re locating 100's of staff to a central point necessitating |
| | | to a central point necessitating staff to drive miles further than |

be worthy of further exploration and analysis. In this way, the use of data displays can help you to interpret your data and to draw meaning from it. As with data condensation, the design and use of displays is part of your analysis continue until the project report is submitted.

Drawing and verifying conclusions

From the start of your analysis you will be interpreting what things mean by noting possible patterns and associated potential explanations. You will be drawing, albeit tentative, conclusions, which you will note while you continue your analysis looking for data to either refute or ground them. As your conclusions and the meanings associated with them become increasingly specific and well supported by your data, you will need continue to verify them testing them for their plausibility and confirmability with your data. Your final conclusions are unlikely to appear until after you have finished your data collection.

Evaluation

Miles et al. (2019) argue there are several advantages associated with using forms of data display in analysis. Qualitative data collection tends to produce hours of audio recorded interviews or extensive piles of notes which are transcribed to produce extended text. This form of display is difficult to analyse because it is both extensive and poorly ordered. Based on the logic that ‘you know what you display’, the analysis of data and the drawing of conclusions from these will be helped by using matrices, networks or other graphical displays using data drawn from your extended text. These forms of display are relatively easy to generate using CAQDAS, and can assist your analytical thinking as you work through several iterations to develop a visual form that represents your data well.

Use of data display and analysis can provide you with a set of procedures to analyse your qualitative data or, alternatively, one or more of the techniques that Miles et al. (2019) outline can be used as part of your approach to data analysis. They describe the analysis of qualitative data as an interactive process, and in this sense their approach includes many aspects of analysis that complement the analytical techniques we discussed earlier. Their approach is systematic and structured, and the procedures they outline are often associated with a fairly high level of formalisation. However, unlike grounded theory, the exact procedures to be followed within their activity flows of data reduction, display and conclusion drawing and verification are not specified. Miles et al. (2019) refer to their book as a ‘sourcebook’, and as such it offers a variety of possible techniques that may be appropriate within the context of your own research question and objectives.

13.14 Using CAQDAS

Overview

CAQDAS (Computer-Aided Qualitative Data Analysis Software, sometimes abbreviated to QDAS) refers to programs containing a range of tools to facilitate the analysis of qualitative data. The use of CAQDAS offers a few advantages in relation to the analytical procedures we have discussed. In particular, when used systematically, it can aid continuity and increase both transparency and methodological rigour.

Silver and Lewins (2014) say that they are often asked which CAQDAS package is best, a question which they say is impossible to answer. They provide reasons for this response.

First, many tools are common to all CAQDAS packages, although each package will have specific characteristics, supporting particular functions in different ways. Second, they say that while there is some discussion about whether specific packages affect the way analysis is conducted, these programs are not designed to cater for particular methodological or analytical approach, and it is the researcher who remains in control of the way the tools in a program are used to facilitate analysis and interpretation. Third, the way in which you are taught to use a particular program may not recognise its full potential. You need to recognise the key aspects of the analytical approach which you wish to use (Section 13.3) and select those tools in your chosen CAQDAS package that facilitate these analyses while being aware of any that are not appropriate.

Depending on your analytical approach some CAQDAS programs may potentially be more useful. Consequently, you will need to develop some familiarity with different programs to be able to evaluate their applicability for the analytical approach you wish to use. However, attempting to achieve this may be problematic if the programs you wish to explore are not readily available. It is worth noting though that some CAQDAS programs that may be suitable for your purpose are available to download as freeware and that commercial software producers offer free downloads of trial versions. Silver and Lewins (2014) say that regardless of your preferred choice, each CAQDAS program will provide you with a range of tools to use to help you to manage and analyse your qualitative data. Even in the situation where you only have one program available to use this should provide you with tools to help you manage and analyse your data.

Function

The general function of a CAQDAS program is to facilitate the management and analysis of a qualitative research project. Its specific functions relate to its software tools that enable particular analytical processes. Based on Silver and Lewins (2014) these may be summarised as:

- *Managing the research project*: all data files can be stored in or linked through a project file created within the software, allowing access to all elements of the project and the establishment of an audit trail.
- *Writing analytic memos, comments, notes, etc.*: thoughts about the data and the research process can be recorded systematically and developmentally.
- *Exploring the data*: data can be explored prior to coding by noting and commenting on points and places of interest.
- *Searching the data initially*: can search for words, phrases, etc. within and across data items to further familiarity.
- *Developing coding*: supports process of developing and applying codes, according to your research philosophy, methodological approach, approach to theory development and analytical technique.
- *Coding*: supports coding of data according to analytical approach.
- *Retrieving coded data*: offers scope to revisit and evaluate coding to data, to facilitate analysis and the future direction of data collection.
- *Revising codes and coding*: offers scope to revisit and re-code data.
- *Organising data*: offers scope to organise and re-organise the qualitative data collected to facilitate analysis.
- *Hyperlinking*: provides ability to link units of data to other units, files etc. for analytical purposes.
- *Searching and interrogating*: facilitates linking and grouping codes, conceptualising and testing relationships.

- *Mapping*: provides visualisation of relationships and representing explanations.
- *Producing outputs*: produces reports allowing you to view material in hard copy or to export it to other applications such as word-processing and spreadsheet programs, as well as producing tabular reports, charts and graphical representations.

What is not apparent from this list is that the functions contained in some CAQDAS packages are better at supporting certain types of qualitative data analysis procedures than others. A wide range of qualitative data exists, and your research may involve collecting one particular type or some combination of these. Text makes up a major type of qualitative data, but this comprises different types such as documents, narratives and transcripts, affecting what you wish to achieve through analysis. Audio, still images and video sources are also important types of qualitative data. This means that you may need to experiment with more than one package before you find the CAQDAS that meets your needs. Your final choice of CAQDAS package will be dependent on a range of factors, including, not least, the relative benefits you will gain relative to the time you need to invest to learn a CAQDAS program. These factors are summarised in Box 13.20 as a checklist.

Where you decide to use a CAQDAS program and have selected a package, you will need to familiarise yourself with it before you start collecting your data. This will avoid the problem of trying to learn the features of the package at the same time as you analyse your data, although you will of course continue to learn about these as you conduct this analysis.

Exploring the latest versions of CAQDAS

Published information about CAQDAS programs is likely to become out of date fairly quickly. Fortunately, there is a wealth of up-to-date information available from the CAQDAS Networking Project's website hosted by the University of Surrey.¹ If you are considering



Box 13.20 Checklist

Choosing a CAQDAS package

- ✓ How much data do you have that needs to be analysed qualitatively?
- ✓ How important are these qualitative data in relation to any other data you have collected for your research project, and will you want to integrate any quantitative data into the qualitative software package you use?
- ✓ What type(s) of qualitative data do you need to analyse: audio, documentary, narratives, transcripts or visual?

- ✓ How much time do you have to learn how to use the package?
- ✓ What is the timeframe for your research project?
- ✓ How much support is available in your university to help you learn to use the package?
- ✓ What is the operating system of your computer?
- ✓ Do you want a package that will help you manage your thinking and assist you in developing your own codes?
- ✓ Do you want a package that will allow you to explore the way language is used in your data?
- ✓ Do you want a package that allows you to display relationships within your data diagrammatically?
- ✓ Do you want a package that will allow you to quantitatively describe the content of your data?

¹The URL for the CAQDAS Networking Project is <https://www.surrey.ac.uk/computer-assisted-qualitative-data-analysis>.

Table 13.3 URLs for a range of CAQDAS developers

| Name | URL | Brief comments |
|----------------|---|---|
| ATLAS.ti | http://www.atlasti.com | Windows and MAC versions. Versatile and flexible. Supports multimedia |
| HyperRESEARCH™ | http://www.researchware.com | Windows and MAC versions. Simple to use. Case-based structure. Supports multimedia |
| MAXQDA | http://www.MAXQDA.com | Windows and MAC versions. Intuitive and easy to use. Mixed methods and content analysis features. Supports multimedia. |
| NVivo | https://www.qsrinternational.com¹ | Windows and MAC versions. Range of editions with added features. Versatile with large range of searching possibilities. Supports multimedia |
| ProSuite | http://www.provalisresearch.com | Windows and MAC versions. Integrated text analytics tools. Computer assisted coding of documents and images (QDA Miner), content analysis and text mining of text (WordStat), statistical analysis (SimStat) |
| Transana | http://www.transana.org | Windows and MAC versions. Specifically designed for qualitative analysis of audio, still image and video data. Ability to synchronise multiple video streams during playback and to synchronise playback with transcripts |

Source: Developed from QUIC Working Paper software reviews available from the CAQDAS Networking Project and Qualitative Innovations in CAQDAS Project (QUIC) website hosted by the University of Surrey and/or software producers websites. Each comment in this table only provides a very brief indication and is not intended to promote or discourage the use of a particular software program, or to advocate the use of one program over other compatible programs. You are advised to evaluate the features and applications of current versions of CAQDAS at the time of your project in relation to the requirements of your research.

¹Full URL for NVivo is <https://www.qsrinternational.com/nvivo-qualitativedata-analysis-software/home>

using CAQDAS, we would strongly recommend a visit to this website, which, in addition to a wealth of useful articles, also contains web links to commercial software producers' sites including downloadable demonstration versions of the software. We would also advise you to explore the Internet sites of CAQDAS producers to obtain details and demonstrations of the latest versions of these packages and the features that they offer. Some of those most widely used are listed in Table 13.3.

13.15 Summary

- Qualitative data comprise rich and full verbal, textual and/or visual data. They may also be characterised as non-standardised and non-numerical data.
- Data collection, analysis and interpretation are an interrelated and interactive set of processes in qualitative research. Analysis often occurs during the collection of data as well as after it.
- Understanding key aspects of different qualitative analysis techniques should help you to choose an appropriate technique, or combination of techniques, to analyse your qualitative data.
- Qualitative data need to be prepared carefully for analysis, usually involving transcription where the spoken word is involved.

- There are several aids that you might use to help you through the process of qualitative analysis, including: interim summaries, event summaries, document summaries, self-memos, maintaining a research notebook and keeping a reflective diary or reflexive journal.
- There are a variety of ways of undertaking qualitative analysis:
 - Thematic Analysis; Template Analysis; Narrative Analysis; discourse analysis; visual analysis; and data display and analysis may commence inductively or deductively;
 - Pattern matching and deductive explanation building commence deductively;
 - Analytic induction and the Grounded Theory Method commence inductively.
- The use of computer-assisted qualitative data analysis software (CAQDAS) can help you during qualitative analysis regarding project management and data organisation, keeping close to your data, exploration, coding and retrieval of your data, searching and interrogating to build propositions and theorise, and recording your thoughts systematically.

Self-check questions

Help with these questions is available at the end of the chapter.

- 13.1** Why is qualitative analysis described as an ‘interactive process’?
- 13.2** Which sorts of data will you need to retain while you are undertaking qualitative research?
- 13.3** What are the main implications of using:
 - a** a deductive analytical approach for the way in which you conduct the process of qualitative analysis?
 - b** an inductive analytical approach for the way in which you conduct the process of qualitative analysis?
- 13.4** What are the key similarities and differences between Thematic Analysis and Template Analysis?

Review and discussion questions

- 13.5** Assume you are undertaking qualitative research or proposing to. Use the checklist in Box 13.1 to commence evaluating your choice of qualitative analysis. Where your analytical process or method is not specified by your choice of research strategy, work through the points in this checklist in conjunction with the material in Sections 13.6 to 13.13 to evaluate how you will undertake this. Draw up a shortlist of possible ways of analysing your data and make notes of points for and against use of each of one. Further evaluate your list of points for and against to decide which you will use to analyse your qualitative data. You may conduct the latter part of this evaluation with a friend to help you both to think through your respective options by discussing this. Brief each other about the nature of your respective data, your shortlists of possible ways for analysing your data, and lists of points for and against each before discussing these options.
- 13.6** With a friend, use part of a transcript that one of you has produced after undertaking a research interview to undertake the following tasks.
 - a** Based on the aim or purpose for undertaking this interview, or interview themes, independently code the data in this part of the transcript using either inductive or deductive coding.
 - b** Compare the results of your coding.

- c Identify where your coding is similar to and different from that of your friend.
- d Where you identify differences in coding, discuss the assumptions you each made when you coded these data and why you made these.
- e By reflecting on your attempt at coding these data, which codes, if any, would you change and why?

You may repeat this process where both of you has conducted a research interview and produced transcripts.

- 13.7** Evaluate the scope to undertake visual research and analysis in your research project by considering the following.
- a How would conducting visual research help you to answer your research question and address your research objectives?
 - b Why would conducting visual research not be appropriate to help you to answer your research question and address your research objectives?
 - c Assess your reasons for and against using visual research. Where you decide that using visual research might be helpful, continue to consider points d through g below.
 - d How you would use this method? This more in-depth consideration will involve you re-reading Sections 8.2, 9.6, 9.7 and 10.11 as well as 13.12.
 - e What would be the nature and purpose of the images produced through your visual research?
 - f How you will analyse these images and for what purpose?
 - g How will you subsequently use these images (or not) in your research report?
- 13.8** Visit the CAQDAS websites listed in Table 13.3. Using the information available on each website, explore the suitability of each program for the nature of your data and chosen approach to analyse these data. Use a demonstration version of a CAQDAS program that may be suitable for your data and analytical approach, to explore its features. Evaluate how useful you think this program will be to assist you in analysing your data.



Progressing your research project

Analysing your data qualitatively

- Review the qualitative data you have collected. These data may be in the form of documentary sources, recordings or notes of observations, visual images, interview transcripts or notes, or diary entries (Chapters 8, 9 and 10).
- Based on your earlier review of the checklist in Box 13.1 and reading of Sections 13.2 and 13.6 to 13.13, decide which will be most appropriate to use to analyse your data.
- Prepare your data for analysis where necessary (Section 13.4) and consider how you will make use of the aids to help you during analysis (Section 13.5).

- Based on your earlier review of CAQDAS programs (Review and discussion question 13.7), decide whether you will make use of a CAQDAS program and where you decide to use one, explore whether this is available for you to use through your university.
- Following the procedural outline of your chosen way of undertaking your analysis in the appropriate section (Sections 13.6 to 13.13) together with guidance from further reading related to their use, analyse your data to answer your research question and address your objectives.
- As you analyse your data and seek to answer your research questions and address your objectives, you will need to continue to evaluate analysis process and its suitability for your data.
- Use the questions in Box 1.4 to guide your reflective diary entry.

References

- Bansal, P. and Roth, K. (2000) 'Why companies go green: A model of ecological responsiveness', *Academy of Management Journal*, Vol. 43, No. 4, pp. 717–36.
- Bell, E. and Davison, J. (2013) 'Visual management studies: Empirical and theoretical approaches', *International Journal of Management Reviews*, Vol. 15, No. 2, pp. 167–84.
- Bishop, D.G., Trevio, L.K., Gioia, D.A. and Kreiner, G.E. (2020, online) 'Leveraging a recessive narrative to transform Joe Paterno's image: Media sensebreaking, sensemaking, and sensegiving during scandal', *Academy of Management Discoveries*, Vol. 6, No. 4. Available at: <https://journals.aom.org/doi/10.5465/amd.2019.0108>
- Bloom, M., Colbert, A.E. and Nielsen, J.D. (2021) 'Stories of calling: How called professionals construct narrative identities', *Administrative Science Quarterly*, Vol. 66, No. 2, pp. 298–338.
- Braun, V. and Clarke, V. (2006) 'Using thematic analysis in psychology', *Qualitative Research in Psychology*, Vol. 3, No. 2, pp. 77–101.
- Braun, V. and Clarke, V. (2022) *Thematic Analysis: A Practical Guide*. London: Sage.
- Brinkmann, S. and Kvale, S. (2015) *InterViews: Learning the Craft of Qualitative Research Interviewing* (3rd edn). London: Sage.
- Bryant, A. and Charmaz, K. (2007) *The Sage Handbook of Grounded Theory*. London: Sage.
- Charmaz, K. (2014) *Constructing Grounded Theory* (2nd edn). London: Sage.
- Corbin, J. and Strauss, A. (2008) *Basics of Qualitative Research Techniques and Procedures for Developing Grounded Theory* (3rd edn). London: Sage.
- Corbin, J. and Strauss, A. (2015) *Basics of Qualitative Research Techniques and Procedures for Developing Grounded Theory* (4th edn). London: Sage.
- Department for Transport (2015) *Know Your Traffic Signs*. London: Department for Transport. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/519129/know-your-traffic-signs.pdf [Accessed 19 February 2022].
- Dey, I. (1993) *Qualitative Data Analysis*. London: Routledge.
- Fairclough, N. (1992) *Discourse and Social Change*. Cambridge: Polity Press.
- Fairclough, N. (2010) *Critical Discourse Analysis: The Critical Study of Language* (2nd edn). Harlow: Pearson Education.
- Gioia, D.A., Corley, K.G. and Hamilton, A.L. (2012) 'Seeking qualitative rigour in inductive research: Notes on the Gioia methodology', *Organizational Research Methods*, Vol. 16, No. 1, pp. 15–31.
- Glaser, B. and Strauss, A. (1967) *The Discovery of Grounded Theory*. Chicago, IL: Aldine.
- Glaser, B.G. (1998) *Doing Grounded Theory: Issues and Discussions*. Mill Valley, CA: Sociology Press.
- Hepburn, A. and Potter, J. (2007) 'Discourse analytic practice', in C. Seale, G. Gobo, J.F. Gubrium and D. Silverman (eds) *Qualitative Research Practice*. London: Sage, pp. 168–184.
- Hjálmsdóttir, A. and Bjarnadóttir, V.S. (2021) "'I have turned into a foreman here at home": Families and work-life balance in times of COVID-19 in a gender equality paradise', *Gender, Work & Organization*, Vol. 28, No. 1, pp. 268–283.
- Holstein, J.A. and Gubrium, J.F. (2011) 'The constructionist analytics of interpretive practice', in N.K. Denzin and Y.S. Lincoln (eds) *The Sage Handbook of Qualitative Research*. London: Sage.
- Hyatt, D. (2005) 'A critical literacy frame for UK secondary education contexts', *English in Education*, Vol. 39, No. 1, pp. 43–59.
- Hyatt, D. (2013) 'The critical policy discourse analysis frame: Helping doctoral students engage with the educational policy analysis', *Teaching in Higher Education*, Vol. 18, No. 8, pp. 833–45.

- Johnson, P. (2004) 'Analytic induction', in G. Symon and C. Cassell (eds) *Essential Guide to Qualitative Methods and Analysis in Organizational Research*. London: Sage, pp. 165–79.
- Kenealy, G.J.J. (2012) 'Grounded theory: A theory building approach', in G. Symon and C. Cassell (eds) *Qualitative Organizational Research: Core Methods and Current Challenges*. London: Sage, pp. 408–25.
- King, N. and Brookes, J.M. (2017) *Template Analysis for Business and Management Students*. London: Sage.
- Labov, W. (1972) *Language in the Inner City: Studies in the Black English Vernacular*. Philadelphia: University of Pennsylvania Press.
- Labov, W. and Waletzky, J. (1967) 'Narrative analysis: Oral versions of personal experience', in J. Helm (ed.) *Essays on the Verbal and Visual Arts*. Seattle, WA: University of Washington Press.
- Maitlis, S. (2012) 'Narrative analysis', in G. Symon and C. Cassell (eds) *Qualitative Organizational Research: Core Methods and Current Challenges*. London: Sage, pp. 492–511.
- Miles, M.B., Huberman, A.M. and Saldaña, J. (2019) *Qualitative Data Analysis: A Methods Sourcebook* (4th edn). London: Sage.
- Phillips, N. and Hardy, C. (2002) *Discourse Analysis: Investigating Processes of Social Construction*. London: Sage.
- Pink, S. (2007) 'Visual methods', in C. Seale, G. Gobo, J.F. Gubrium and D. Silverman (eds) *Qualitative Research Practice*. London: Sage, pp. 361–376.
- Riessman, C.K. (2008) *Narrative Methods for the Human Sciences*. London: Sage.
- Ritchie, J., Lewis, J., McNaughton Nicholls, C. and Ormston, R. (2014) *Qualitative Research Practice* (2nd edn). London: Sage.
- Rose, G. (2016) *Visual Methodologies: An Introduction to Researching with Visual Materials* (4th edn). London: Sage.
- Rowley, J. and Hanna, S. (2020). Branding destinations: symbolic and narrative representations and co-branding. *Journal of Brand Management*, Vol. 27, No. 3, pp. 328–38.
- Silver, C. and Lewins, A. (2014) *Using Software in Qualitative Research: A Step-by-step Guide* (2nd edn). London: Sage.
- Silverman, D. (2013) *A Very Short, Fairly Interesting and Reasonably Cheap Book about Qualitative Research* (2nd edn). London: Sage.
- Strauss, A. and Corbin, J. (1998) *Basics of Qualitative Research* (2nd edn). Thousand Oaks, CA: Sage.
- University of Leicester (n.d.) *What Is the Jefferson Transcription System?* Available at: <https://www2.le.ac.uk/departments/psychology/research/child-mental-health/cara-1/faqs/jefferson> [Accessed 8 Feb 2022].
- Woolf, N.H. and Silver, C. (2018) *Qualitative Analysis Using NVivo: The Five-Level QDA Method*. New York: Routledge.
- Yin, R.K. (2018) *Case Study Research and Applications: Design and Methods* (6th edn). London: Sage.

Further reading

Braun, V. and Clarke, V. (2022) *Thematic Analysis: A Practical Guide*. London: Sage. This book provides an accessible guide to doing Reflexive Thematic Analysis.

Charmaz, K. (2014) *Constructing Grounded Theory* (2nd edn). London: Sage. This book provides an accessible discussion and evaluation of grounded theory and the grounded theory method.

Chapter 13 Analysing data qualitatively

- Miles, M.B., Huberman, A.M. and Saldana, J. (2019) *Qualitative Data Analysis: A Methods Sourcebook* (4th edn). London: Sage. This sourcebook presents a range of techniques for the analysis of qualitative data.
- Reissner, S. and Whittle, A. (2021, online first) 'Interview-based research in management and organisation studies: making sense of the plurality of methodological practices and presentational styles', *Qualitative Research in Organizations and Management: An International Journal*.
- Riessman, C.K. (2008) *Narrative Methods for the Human Sciences*. London: Sage. This book provides a helpful discussion of approaches to narrative analysis.
- Rose, G. (2016) *Visual Methodologies: An Introduction to Researching with Visual Materials* (4th edn). London: Sage. An interesting and useful book on visual research methods with helpful chapters on semiology, content analysis and discourse analysis, and material on photo elicitation and photo essays.
- Silver, C. and Lewins, A. (2014) *Using Software in Qualitative Research: A Step-by-step Guide* (2nd edn). London: Sage. An authoritative guide to using CAQDAS.
- Symon, G. and Cassell, C. (eds) (2012) *Qualitative Organizational Research: Core Methods and Current Challenges*. London: Sage. This edited book contains a range of chapters on qualitative analysis including template analysis, grounded theory, narrative analysis and discourse analysis.

Big Data & Data Mining

Sharda et al. (2023)

Big Data (136-143, 153-156)

Data Mining Process (256-269)

Descriptive Analytics I: Nature of Data, Big Data, and Statistical Modeling

LEARNING OBJECTIVES

- Understand the nature of data as it relates to business intelligence (BI) and analytics
- Learn the methods used to make real-world data analytics ready
- Learn what Big Data is and how it is changing the world of analytics
- Understand the motivation for and business drivers of Big Data analytics
- Become familiar with the wide range of enabling technologies for Big Data analytics
- Learn about Hadoop, Spark, MapReduce, and NoSQL as they relate to Big Data analytics
- Become familiar with the Data for Good concept
- Understand the need for and appreciate the capabilities of stream analytics
- Learn about the applications of stream analytics
- Describe statistical modeling and its relationship to business analytics
- Learn about descriptive and inferential statistics

In the age of Big Data and business analytics in which we are living, the importance of data is undeniable. The newly coined phrases like “data is the oil,” “data is the new bacon,” “data is the new currency,” and “data is the ruler” are further stressing the renewed importance of data. But what type of data are we talking about? Obviously, not just any data. The “garbage in garbage out—GIGO” concept/principle applies to today’s “Big Data” phenomenon more so than any data definition that we have had in the past. To live up to its promise, its value proposition, and its ability to turn into insight, data has to be carefully created/identified, collected, integrated, cleaned, transformed, and properly contextualized for use in accurate and timely decision making.

to unbalanced/original data set) produced better prediction models for identifying the students who are likely to drop out of the college prior to their sophomore year. Among the four individual prediction models used in this study, support vector machines performed the best, followed by decision trees, neural networks, and logistic regression. From the usability standpoint, despite the fact that support vector machines showed better prediction results, one might choose to use decision trees because compared to support vector machines and neural networks, they portray a more transparent model structure. Decision trees explicitly show the reasoning process of different predictions, providing a justification for a specific outcome, whereas support vector machines and artificial neural networks are mathematical models that do not provide such a transparent view of “how they do what they do.”

Sources: Delen, D., Topuz, K., & Eryarsoy, E. (2020). Development of a Bayesian Belief Network-based DSS for predicting and understanding freshmen student attrition. *European Journal of Operational Research*, 281(3), 575–587; Thammasiri, D., Delen, D., Meesad, P., & Kasap N. (2014). A critical assessment of imbalanced class distribution problem: The case of predicting freshmen student attrition. *Expert Systems with Applications*, 41(2), 321–330; Delen, D. (2011). Predicting student attrition with data mining methods. *Journal of College Student Retention*, 13(1), 17–35; Delen, D. (2010). A comparative analysis of machine learning techniques for student retention management. *Decision Support Systems*, 49(4), 498–506.

► SECTION 3.4 REVIEW QUESTIONS

1. Why is the original/raw data not readily usable by analytics tasks?
2. What are the main data preprocessing steps?
3. What does it mean to clean/scrub the data? What activities are performed in this phase?
4. Why do we need data transformation? What are the commonly used data transformation tasks?
5. Data reduction can be applied to rows (sampling) and/or columns (variable selection). Which is more challenging?

3.5 DEFINITION OF BIG DATA

Using data to understand customers/clients and business operations to sustain (and foster) growth and profitability is an increasingly challenging task for today’s enterprises. As more and more data becomes available in various forms and fashions, timely processing of the data with traditional means becomes impractical. Nowadays, this phenomenon, usually called Big Data is receiving substantial press coverage and drawing increasing interest from both business users and IT professionals. The result is that Big Data is becoming an overhyped and overused marketing buzzword.

Big Data means different things to people with different backgrounds and interests. Traditionally, the term *Big Data* has been used to describe the massive volumes of data analyzed by huge organizations like Google or research science projects at NASA. But for most businesses, it’s a relative term: “Big” depends on an organization’s size. The point is more about finding new value within and outside conventional data sources. Pushing the boundaries of data analytics uncovers new insights and opportunities, and “big” depends on where you start and how you proceed. Consider the popular description of Big Data: Big Data exceeds the reach of commonly used hardware environments and/or capabilities of software tools to capture, manage, and process it within a tolerable time span for its user population. **Big Data** has become a popular term to describe the exponential growth, availability, and use of information, both structured and unstructured. Much has been written on the Big Data trend and how it can serve as the basis for innovation, differentiation, and growth. Because of the technology challenges in managing the large volume of data coming from multiple sources, sometimes at a rapid speed, additional

new technologies have been developed to overcome the technology challenges. Use of the term *Big Data* is usually associated with such technologies. Because a prime use of storing such data is generating insights through analytics, sometimes the term Big Data is expanded as Big Data analytics. But the term is becoming content free in that it can mean different things to different people. Because our goal is to introduce you to the large data sets and their potential in generating insights, we will use the original term in this chapter.

Where does Big Data come from? A simple answer is “everywhere.” The sources that were ignored because of the technical limitations are now treated as gold mines. Big Data may come from Web logs, radio-frequency identification (RFID), global positioning systems (GPS), sensor networks, social networks, Internet-based text documents, Internet search indexes, detail call records, astronomy, atmospheric science, biology, genomics, nuclear physics, biochemical experiments, medical records, scientific research, military surveillance, photography archives, video archives, and large-scale e-commerce practices.

Big Data is not new. What is new is that the definition and the structure of Big Data constantly change. Companies have been storing and analyzing large volumes of data since the advent of the data warehouses in the early 1990s. Whereas terabytes used to be synonymous with Big Data warehouses, now it's exabytes, and the rate of growth in data volume continues to escalate as organizations seek to store and analyze greater levels of transaction details, as well as Web- and machine-generated data, to gain a better understanding of customer behavior and business drivers.

Many (academics and industry analysts/leaders alike) think that “Big Data” is a misnomer. What it says and what it means are not exactly the same. That is, Big Data is not just “big.” The sheer volume of the data is only one of many characteristics that are often associated with Big Data, including variety, velocity, veracity, variability, and value proposition, among others.

The “V”s That Define Big Data

Big Data is typically defined by three “V”s: volume, variety, velocity. In addition to these three, we see some of the leading Big Data solution providers adding other “V”s, such as veracity (IBM), variability (SAS), and value proposition.

Volume Volume is obviously the most common trait of Big Data. Many factors contributed to the exponential increase in data volume, such as transaction-based data stored through the years, text data constantly streaming in from social media, increasing amounts of sensor data being collected, automatically generated RFID and GPS data, and so on. In the past, excessive data volume created storage issues, both technical and financial. But with today’s advanced technologies coupled with decreasing storage costs, these issues are no longer significant; instead, other issues have emerged, including how to determine relevance amid the large volumes of data and how to create value from data that is deemed to be relevant.

As mentioned before, *big* is a relative term. It changes over time and is perceived differently by different organizations. With the staggering increase in data volume, even the naming of the next Big Data echelon has been a challenge. The highest mass of data that used to be called petabytes (PB) has left its place to zettabytes (ZB), which is a trillion gigabytes (GB) or a billion terabytes (TB). Technology Insights 3.1 provides an overview of the size and naming of Big Data volumes.

From a short historical perspective, in 2009 the world had about 0.8 ZB of data; in 2010, it exceeded the 1 ZB mark; at the end of 2011, the number was 1.8 ZB. It is expected to be 44 ZB in 2020 (Adshead, 2014). With the growth of sensors and the Internet of Things (IoT), these forecasts could all be wrong. Though these numbers are astonishing in size, so are the challenges and opportunities that come with them.

TECHNOLOGY INSIGHTS 3.1 The Data Size Is Getting Big, Bigger, and Bigger

The metrics of data size is having a hard time keeping up with new names. We all know kilobyte (KB, which is 1,000 bytes), megabyte (MB, which is 1,000,000 bytes), gigabyte (GB, which is 1,000,000,000 bytes), and terabyte (TB, which is 1,000,000,000,000 bytes). Beyond that, the names given to data sizes are relatively new to most of us. The following table shows what comes after terabyte and beyond.

| Name | Symbol | Value |
|-------------|--------|-----------|
| Kilobyte | KB | 10^3 |
| Megabyte | MB | 10^6 |
| Gigabyte | GB | 10^9 |
| Terabyte | TB | 10^{12} |
| Petabyte | PB | 10^{15} |
| Exabyte | EB | 10^{18} |
| Zettabyte | ZB | 10^{21} |
| Yottabyte | YB | 10^{24} |
| Brontobyte* | BB | 10^{27} |
| Gogobyte* | GeB | 10^{30} |

*Not an official SI (International System of Units) name/symbol, yet.

Consider that an exabyte of data is created on the Internet each day, which equates to 250 million DVDs' worth of information. And the idea of even larger amounts of data—a zettabyte—isn't too far off when it comes to the amount of information traversing the Web in any one year. In fact, industry experts are already estimating that we will see 1.3 zettabytes of traffic annually over the Internet by 2016—and it could jump to 2.3 zettabytes by 2020. By 2020, Internet traffic is expected to reach 300 GB per capita per year. When referring to yottabytes, some of the Big Data scientists often wonder about how much data the NSA or FBI have on people altogether. Put in terms of DVDs, a yottabyte would require 250 trillion of them. A brontobyte, which is not an official SI prefix but is apparently recognized by some people in the measurement community, is a 1 followed by 27 zeros. The size of such a magnitude can be used to describe the amount of sensor data that we will get from the Internet in the next decade, if not sooner.

Sources: Backblaze. (2022). What is a yottabyte? at <https://www.backblaze.com/blog/what-is-a-yottabyte/> (accessed May 2022); Byte. (2022). Unit of information, at <https://en.wikipedia.org/wiki/Byte> (accessed April 2022).

Variety Data today come in all types of formats—ranging from traditional databases to hierarchical data stores created by the end users and OLAP systems to text documents, e-mail, XML, meter-collected and sensor-captured data, to video, audio, and stock ticker data. By some estimates, 80 to 85% of all organizations' data are in some sort of unstructured or semistructured format (a format that is not suitable for traditional database schemas). But there is no denying its value, and hence, it must be included in the analyses to support decision making.

Velocity According to Gartner, velocity means both how fast data is being produced and how fast the data must be processed (i.e., captured, stored, and analyzed) to meet the need or demand. RFID tags, automated sensors, GPS devices, and smart meters are driving an increasing need to deal with torrents of data in near real time. Velocity is

perhaps the most overlooked characteristic of Big Data. Reacting quickly enough to deal with velocity is a challenge to most organizations. For time-sensitive environments, the opportunity cost clock of the data starts ticking the moment the data is created. As time passes, the value proposition of the data degrades and eventually becomes worthless. Whether the subject matter is the health of a patient, the well-being of a traffic system, or the health of an investment portfolio, accessing the data and reacting faster to the circumstances will always create more advantageous outcomes.

In the Big Data storm that we are currently witnessing, almost everyone is fixated on at-rest analytics, using optimized software and hardware systems to mine large quantities of variant data sources. Although this is critically important and highly valuable, there is another class of analytics, driven from the velocity of Big Data, called “data stream analytics” or “in-motion analytics,” which is evolving fast. If done correctly, data stream analytics can be as valuable, and in some business environments more valuable, than at-rest analytics. Later in this chapter, we will cover this topic in more detail.

Veracity *Veracity* is a term coined by IBM that is being used as the fourth “V” to describe Big Data. It refers to conformity to facts: accuracy, quality, truthfulness, or trustworthiness of the data. Tools and techniques are often used to handle Big Data’s veracity by transforming the data into quality and trustworthy insights.

Variability In addition to the increasing velocities and varieties of data, data flows can be highly inconsistent with periodic peaks. Is something big trending in the social media? Perhaps there is a high-profile IPO looming. Maybe swimming with manatees in the Bahamas is suddenly the must-do vacation activity. Daily, seasonal, and event-triggered peak data loads can be highly variable and thus challenging to manage—especially with social media involved.

Value Proposition The excitement around Big Data is its value proposition. A preconceived notion about “Big” data is that it contains (or has a greater potential to contain) more patterns and interesting anomalies than “small” data. Thus, by analyzing large and feature-rich data, organizations can gain greater business value that they may not have otherwise. Although users can detect the patterns in small data sets using simple statistical and machine-learning methods or ad hoc query and reporting tools, Big Data means “big” analytics. Big analytics means greater insight and better decisions, something that every organization needs.

Because the exact definition of Big Data (or its successor terms) is still a matter of ongoing discussion in academic and industrial circles, it is likely that more characteristics (perhaps more “V”s) are likely to be added to this list. Regardless of what happens, the importance and value proposition of Big Data is here to stay. Figure 3.7 shows a conceptual architecture where Big Data (at the left side of the figure) is converted to business insight through the use of a combination of advanced analytics and delivered to a variety of different users/roles for faster/better decision making.

► SECTION 3.5 REVIEW QUESTIONS

1. Why is Big Data important? What has changed to put it in the center of the analytics world?
2. How do you define Big Data? Why is it difficult to define?
3. Out of the “V”s that are used to define Big Data, in your opinion, which one is the most important? Why?
4. What do you think the future of Big Data will be like? Will it leave its popularity to something else? If so, what will it be?

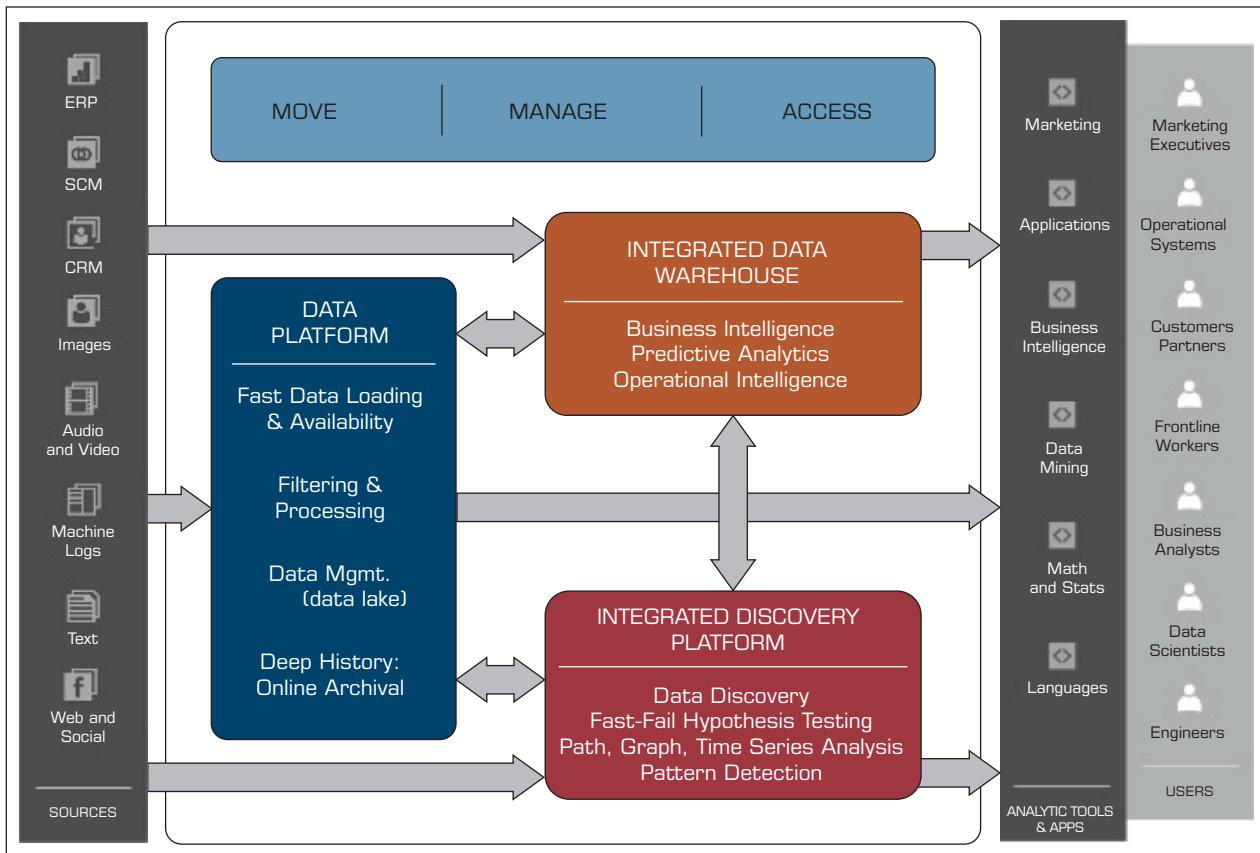


FIGURE 3.7 A High-Level Conceptual Architecture for Big Data Solutions. Source: AsterData—A Teradata Company.

3.6 FUNDAMENTALS OF BIG DATA ANALYTICS

Big Data by itself, regardless of the size, type, or speed, is worthless unless business users do something with it that delivers value to their organizations. That's where "big" analytics comes into the picture. Although organizations have always run reports and dashboards against data warehouses, most have not opened these repositories to in-depth on-demand exploration. This is partly because analysis tools are too complex for the average user but also because the repositories often do not contain all the data needed by the power user. But this is about to change (and has been changing, for some) in a dramatic fashion, thanks to the new Big Data analytics paradigm.

With the value proposition, Big Data also brought about big challenges for organizations. The traditional means for capturing, storing, and analyzing data are not capable of dealing with Big Data effectively and efficiently. Therefore, new breeds of technologies need to be developed (or purchased/hired/outsourced) to take on the Big Data challenge. Before making such an investment, organizations should justify the means. Here are some questions that may help shed light on this situation. If any of the following statements are true, then you need to seriously consider embarking on a Big Data journey.

- You can't process the amount of data that you want to because of the limitations posed by your current platform or environment.
- You want to involve new/contemporary data sources (e.g., social media, RFID, sensory, Web, GPS, textual data) into your analytics platform, but you can't because it

does not comply with the data storage schema-defined rows and columns without sacrificing fidelity or the richness of the new data.

- You need to (or want to) integrate data as quickly as possible to be current on your analysis.
- You want to work with a schema-on-demand (as opposed to predetermined schema used in relational database management systems [RDBMSs]) data storage paradigm because the nature of the new data may not be known, or there may not be enough time to determine it and develop schema for it.
- The data is arriving so fast at your organization's doorstep that your traditional analytics platform cannot handle it.

As is the case with any other large IT investment, the success in Big Data analytics depends on a number of factors. Figure 3.8 shows a graphical depiction of the most critical success factors.

The following are the most critical success factors for **Big Data analytics**:

- 1. A clear business need (alignment with the vision and the strategy).** Business investments ought to be made for the good of the business, not for the sake of mere technology advancements. Therefore, the main driver for Big Data analytics should be the needs of the business, at any level—strategic, tactical, and operations.
- 2. Strong, committed sponsorship (executive champion).** It is a well-known fact that if you don't have strong, committed executive sponsorship, it is difficult (if not impossible) to succeed. If the scope is a single or a few analytical applications, the sponsorship can be at the departmental level. However, if the target is enterprise-wide organizational transformation, which is often the case for Big Data initiatives, sponsorship needs to be at the highest levels and organization wide.



FIGURE 3.8 Critical Success Factors for Big Data Analytics.

3. Alignment between the business and IT strategy. It is essential to make sure that the analytics work is always supporting the business strategy, and not the other way around. Analytics should play the enabling role in successfully executing the business strategy.

4. A fact-based decision-making culture. In a fact-based decision-making culture, the numbers rather than intuition, gut feeling, or supposition drive decision making. There is also a culture of experimentation to see what works and what doesn't. To create a fact-based decision-making culture, senior management needs to:

- Recognize that some people can't or won't adjust
- Be a vocal supporter
- Stress that outdated methods must be discontinued
- Ask to see what analytics went into decisions
- Link incentives and compensation to desired behaviors

5. A strong data infrastructure. Data warehouses have provided the data infrastructure for analytics. This infrastructure is changing and being enhanced in the Big Data era with new technologies. Success requires marrying the old with the new for a holistic infrastructure that works synergistically.

As the size and complexity increase, the need for more efficient analytical systems is also increasing. To keep up with the computational needs of Big Data, a number of new and innovative computational techniques and platforms have been developed. These techniques are collectively called **high-performance computing**, which includes the following:

- **In-memory analytics:** Solves complex problems in near real time with highly accurate insights by allowing analytical computations and Big Data to be processed in-memory and distributed across a dedicated set of nodes.
- **In-database analytics:** Speeds time to insights and enables better data governance by performing data integration and analytic functions inside the database so you won't have to move or convert data repeatedly.
- **Grid computing:** Promotes efficiency, lower cost, and better performance by processing jobs in a shared, centrally managed pool of IT resources.
- **Appliances:** Brings together hardware and software in a physical unit that is not only fast but also scalable on an as-needed basis.

Computational requirements are just a small part of the list of challenges that Big Data impose on today's enterprises. The following is a list of challenges that are found by business executives to have a significant impact on successful implementation of Big Data analytics. When considering Big Data projects and architecture, being mindful of these challenges will make the journey to analytics competency a less stressful one.

Data volume: The ability to capture, store, and process a huge volume of data at an acceptable speed so that the latest information is available to decision makers when they need it.

Data integration: The ability to combine data that is not similar in structure or source and to do so quickly and at a reasonable cost.

Processing capabilities: The ability to process data quickly, as it is captured. The traditional way of collecting and processing data may not work. In many situations, data needs to be analyzed as soon as it is captured to leverage the most value. (This is called *stream analytics*, which will be covered later in this chapter.)

Data governance: The ability to keep up with the security, privacy, ownership, and quality issues of Big Data. As the volume, variety (format and source), and velocity of data change, so should the capabilities of governance practices.

Skills availability: Big Data is being harnessed with new tools and is being looked at in different ways. There is a shortage of people (often called *data scientists*) with skills to do the job.

Solution cost: Because Big Data has opened up a world of possible business improvements, a great deal of experimentation and discovery is taking place to determine the patterns that matter and the insights that turn to value. To ensure a positive return on investment on a Big Data project, therefore, it is crucial to reduce the cost of the solutions used to find that value.

Though the challenges are real, so is the value proposition of Big Data analytics. Anything that you can do as a business analytics leader to help prove the value of new data sources to the business will move your organization beyond experimenting and exploring Big Data into adapting and embracing it as a differentiator. There is nothing wrong with exploration, but ultimately the value comes from putting those insights into action.

Business Problems Addressed by Big Data Analytics

The top business problems addressed by Big Data overall are process efficiency and cost reduction, as well as enhancing customer experience, but different priorities emerge when it is looked at by industry. Process efficiency and cost reduction are perhaps among the top-ranked problems that can be addressed with Big Data analytics for the manufacturing, government, energy and utilities, communications and media, transport, and healthcare sectors. Enhanced customer experience may be at the top of the list of problems addressed by insurance companies and retailers. Risk management usually is at the top of the list for companies in banking and education. Here is a partial list of problems that can be addressed using Big Data analytics:

- Process efficiency and cost reduction
- Brand management
- Revenue maximization, cross-selling, and up-selling
- Enhanced customer experience
- Churn identification, customer recruiting
- Improved customer service
- Identifying new products and market opportunities
- Risk management
- Regulatory compliance
- Enhanced security capabilities

► SECTION 3.6 REVIEW QUESTIONS

1. What is Big Data analytics? How does it differ from regular analytics?
2. What are the critical success factors for Big Data analytics?
3. What are the big challenges that one should be mindful of when considering implementation of Big Data analytics?
4. What are the common business problems addressed by Big Data analytics?

3.7 BIG DATA TECHNOLOGIES

There are a number of technologies for processing and analyzing Big Data, but most have some common characteristics (Kelly, 2012). Namely, they take advantage of commodity hardware to enable scale-out and parallel-processing techniques; employ nonrelational data storage capabilities to process unstructured and semistructured data; and apply

► SECTION 3.7 REVIEW QUESTIONS

1. What are the common characteristics of emerging Big Data technologies?
2. What is Hadoop? How does it work?
3. What are the main Hadoop components? What functions do they perform?
4. What is MapReduce? What does it do? How does it do it?
5. What is Spark and how does it compare to Hadoop? What are the similarities and differences?
6. What is NoSQL? How does it fit into the Big Data analytics picture?
7. What is Data for Good? Select, study, and report on two of the Data for Good initiatives. In your report, make sure to comment on the origins, motivations, and outcomes of the initiatives.

3.8 BIG DATA AND STREAM ANALYTICS

Along with volume and variety, as we have seen earlier in this chapter, one of the key characteristics that defines Big Data is velocity, which refers to the speed at which the data is created and streamed into the analytics environment. Organizations are looking for new means to process streaming data as it comes in to react quickly and accurately to problems and opportunities to please their customers and to gain a competitive advantage. In situations where data streams in rapidly and continuously, traditional analytics approaches that work with previously accumulated data (i.e., data at rest) often either arrive at the wrong decisions because of using too much out-of-context data, or they arrive at the correct decisions but too late to be of any use to the organization. Therefore, it is critical for a number of business situations to analyze the data soon after it is created and/or as soon as it is streamed into the analytics system.

The presumption that the vast majority of modern-day businesses are currently living by is that it is important and critical to record every piece of data because it might contain valuable information now or sometime in the near future. However, as long as the number of data sources increases, the “store-everything” approach becomes harder and harder and, in some cases, not even feasible. In fact, despite technological advances, current total storage capacity lags far behind the digital information being generated in the world. Moreover, in the constantly changing business environment, real-time detection of meaningful changes in data as well as of complex pattern variations within a given short time window is essential to come up with the actions that better fit with the new environment. These facts become the main triggers for a paradigm that we call *stream analytics*. The stream analytics paradigm was born as an answer to these challenges, namely, the unbounded flows of data that cannot be permanently stored to be subsequently analyzed, in a timely and efficient manner, and complex pattern variations that need to be detected and acted on as soon as they happen.

Stream analytics (also called *data-in-motion analytics* and *real-time data analytics*, among others) is a term commonly used for the analytic process of extracting actionable information from continuously flowing/streaming data. A stream is defined as a continuous sequence of data elements (Zikopoulos et al., 2013). The data elements in a stream are often called *tuples*. In a relational database sense, a tuple is similar to a row of data (a record, an object, an instance). However, in the context of semistructured or unstructured data, a tuple is an abstraction that represents a package of data, which can be characterized as a set of attributes for a given object. If a tuple by itself is not sufficiently informative for analysis or a correlation—or other collective relationships among tuples are needed—then a window of data that includes a set of tuples is used. A window of data is a finite number/sequence of tuples, where the windows are continuously updated

as new data become available. The size of the window is determined based on the system being analyzed. Stream analytics is becoming increasingly more popular because of two things. First, time-to-action has become an ever-decreasing value, and second, we have the technological means to capture and process the data while it is created.

Some of the most impactful applications of stream analytics were developed in the energy industry, specifically for smart grid (electric power supply chain) systems. The new smart grids are capable of not only real-time creation and processing of multiple streams of data to determine optimal power distribution to fulfill real customer needs, but also generating accurate short-term predictions aimed at covering unexpected demand and renewable energy generation peaks. Figure 3.10 shows a depiction of a generic use case for streaming analytics in the energy industry (a typical smart grid application). The goal is to accurately predict electricity demand and production in real time by using streaming data that is coming from smart meters, production system sensors, and meteorological models. The ability to predict near future consumption/production trends and detect anomalies in real time can be used to optimize supply decisions (how much to produce, what sources of production to use, and optimally adjust production capacities) as well as to adjust smart meters to regulate consumption and favorable energy pricing.

Applications of Stream Analytics

Because of its power to create insight instantly, helping decision makers to be on top of events as they unfold and allowing organizations to address issues before they become problems, the use of streaming analytics is on an exponentially increasing trend. The following are some of the application areas that have already benefited from stream analytics.

e-Commerce

Companies like Amazon and eBay (among many others) are trying to make the most out of the data that they collect while a customer is on their Web site. Every page visit, every product looked at, every search conducted, and every click made is recorded and

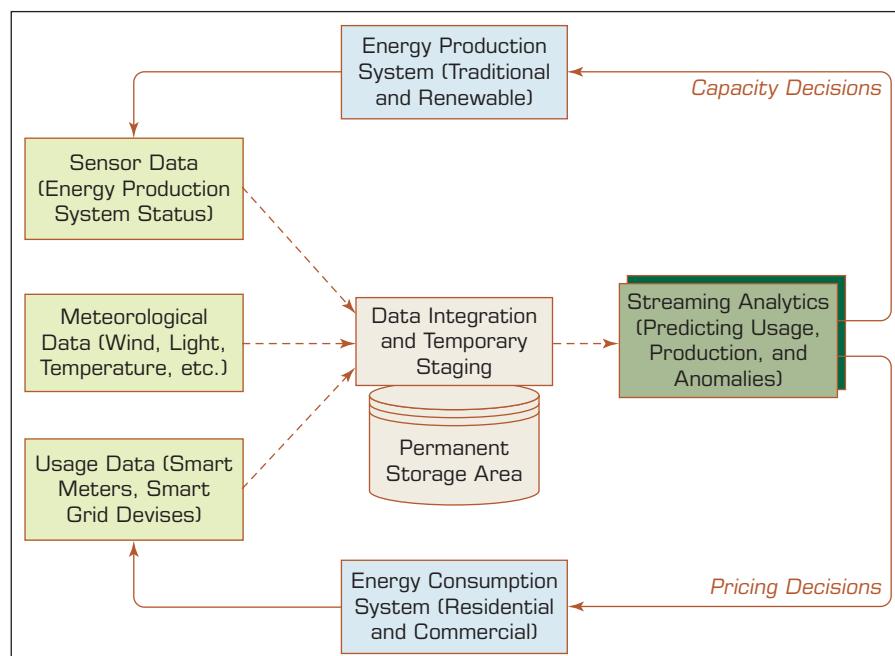


FIGURE 3.10 A Use Case of Streaming Analytics in the Energy Industry.

analyzed to maximize the value gained from a user's visit. If done quickly, analysis of such a stream of data can turn browsers into buyers and buyers into shopaholics. When we visit an e-commerce Web site, even the ones where we are not a member, after a few clicks here and there we start to get very interesting product and bundle price offers. Behind the scenes, advanced analytics are crunching the real-time data coming from our clicks, and the clicks of thousands of others, to "understand" what it is that we are interested in (in some cases, even we do not know that) and make the most of that information by making creative offerings.

Telecommunications

The volume of data that come from call detail records (CDR) for telecommunications companies is astounding. Although this information has been used for billing purposes for quite some time now, there is a wealth of knowledge buried deep inside this Big Data that the telecommunications companies are just now realizing to tap. For instance, CDR data can be analyzed to prevent churn by identifying networks of callers, influencers, leaders, and followers within those networks and proactively acting on this information. As we all know, influencers and leaders have the effect of changing the perception of the followers within their network toward the service provider, either positively or negatively. Using social network analysis techniques, telecommunication companies are identifying the leaders and influencers and their network participants to better manage their customer base. In addition to churn analysis, such information can also be used to recruit new members and maximize the value of the existing members.

Continuous streams of data that come from CDR can be combined with social media data (sentiment analysis) to assess the effectiveness of marketing campaigns. Insight gained from these data streams can be used to rapidly react to adverse effects (which may lead to loss of customers) or boost the impact of positive effects (which may lead to maximizing purchases of existing customers and recruitment of new customers) observed in these campaigns. Furthermore, the process of gaining insight from CDR can be replicated for data networks using Internet protocol detail records. Because most telecommunications companies provide both of these service types, a holistic optimization of all offerings and marketing campaigns could lead to extraordinary market gains.

Law Enforcement and Cybersecurity

Streams of Big Data provide excellent opportunities for improved crime prevention, law enforcement, and enhanced security. They offer unmatched potential when it comes to security applications that can be built in the space, such as real-time situational awareness, multimodal surveillance, cyber-security detection, legal wiretapping, video surveillance, and face recognition (Zikopoulos et al., 2013). As an application of information assurance, enterprises can use streaming analytics to detect and prevent network intrusions, cyberattacks, and malicious activities by streaming and analyzing network logs and other Internet activity monitoring resources.

Power Industry

Because of the increasing use of smart meters, the amount of real-time data collected by power utilities is increasing exponentially. Moving from once a month to every 15 minutes (or more frequently), meter reading accumulates large quantities of invaluable data for power utilities. These smart meters and other sensors placed all around the power grid are sending information back to the control centers to be analyzed in real time. Such analyses help utility companies to optimize their supply chain decisions (e.g., capacity adjustments, distribution network options, real-time buying or selling) based on the

up-to-the-minute consumer usage and demand patterns. In addition, utility companies can integrate weather and other natural conditions data into their analytics to optimize power generation from alternative sources (e.g., wind, solar) and to better forecast energy demand on different geographic granulations. Similar benefits also apply to other utilities such as water and natural gas.

Financial Services

Financial service companies are among the prime examples where analysis of Big Data streams can provide faster and better decisions, competitive advantage, and regulatory oversight. The ability to analyze fast-paced, high-volumes of trading data at very low latency across markets and countries offers a tremendous advantage to making the split-second buy/sell decisions that potentially translate into big financial gains. In addition to optimal buy/sell decisions, stream analytics can also help financial service companies in real-time trade monitoring to detect fraud and other illegal activities.

Health Sciences

Modern-era medical devices (e.g., electrocardiograms and equipment that measure blood pressure, blood oxygen level, blood sugar level, and body temperature) are capable of producing invaluable streaming diagnostic/sensory data at a very fast rate. Harnessing this data and analyzing it in real time offers benefits—the kind that we often call “life and death”—unlike any other field. In addition to helping healthcare companies become more effective and efficient (and hence more competitive and profitable), stream analytics is also improving patient conditions and saving lives.

Many hospital systems all around the world are developing care infrastructures and health systems that are futuristic. These systems aim to take full advantage of what the technology has to offer, and more. Using hardware devices that generate high-resolution data at a very rapid rate, coupled with super-fast computers that can synergistically analyze multiple streams of data, increases the chances of keeping patients safe by quickly detecting anomalies. These systems are meant to help human decision makers make faster and better decisions by being exposed to a multitude of information as soon as it becomes available.

Government

Governments around the world are trying to find ways to be more efficient (via optimal use of limited resources) and effective (providing the services that people need and want). As the practices for e-government become mainstream, coupled with widespread use and access to social media, very large quantities of data (both structured and unstructured) are at the disposal of government agencies. Proper and timely use of these Big Data streams differentiates proactive and highly efficient agencies from the ones who are still using traditional methods to react to situations as they unfold. Another way in which government agencies can leverage real-time analytics capabilities is to manage natural disasters such as snowstorms, hurricanes, tornadoes, and wildfires through a surveillance of streaming data coming from radar, sensors, and other smart detection devices. They can also use similar approaches to monitor water quality, air quality, and consumption patterns and detect anomalies before they become significant problems. Another area where government agencies use stream analytics is in traffic management in congested cities. By using the data coming from traffic flow cameras, GPS data coming from commercial vehicles, and traffic sensors embedded in roadways, agencies are able to change traffic light sequences and traffic flow lanes to ease the pain caused by traffic congestion problems.

CHAPTER
5

Predictive Analytics I: Data Mining Process, Methods, and Algorithms

LEARNING OBJECTIVES

- Define data mining as an enabling technology for business analytics
- Understand the objectives and benefits of data mining
- Become familiar with the wide range of applications of data mining
- Learn the standardized data mining processes
- Learn different methods and algorithms of data mining
- Build awareness of the existing data mining software tools
- Understand the privacy issues, pitfalls, and myths of data mining

Generally speaking, data mining is a way to develop intelligence (i.e., actionable information or knowledge) from data that an organization collects, organizes, and stores. A wide range of data mining techniques are being used by organizations to gain a better understanding of their customers and their operations and to solve complex organizational problems. In this chapter, we study data mining as an enabling technology for business analytics and predictive analytics, learn about the standard processes of conducting data mining projects, understand and build expertise in the use of major data mining techniques, develop awareness of the existing software tools, and explore privacy issues, common myths, and pitfalls that are often associated with data mining.

- 5.1** Opening Vignette: Police Departments Are Using Predictive Analytics to Foresee and Fight Crime 251
- 5.2** Data Mining Concepts and Applications 254
- 5.3** Data Mining Applications 261
- 5.4** Data Mining Process 264
- 5.5** Data Mining Methods 272
- 5.6** Data Mining Software Tools 286
- 5.7** Data Mining Privacy Issues, Myths, and Blunders 292

Definitions, Characteristics, and Benefits

Simply defined, **data mining** is a term used to describe discovering or “mining” knowledge from large amounts of data. When considered by analogy, one can easily realize that the term *data mining* is a misnomer; that is, mining of gold from within rocks or dirt is referred to as “gold” mining rather than “rock” or “dirt” mining. Therefore, data mining perhaps should have been named “knowledge mining” or “knowledge discovery.” Despite the mismatch between the term and its meaning, *data mining* has become the choice of the community. Many other names that are associated with data mining include *knowledge extraction*, *pattern analysis*, *data archaeology*, *information harvesting*, *pattern searching*, and *data dredging*.

Technically speaking, data mining is a process that uses statistical, mathematical, and artificial intelligence techniques to extract and identify useful information and subsequent knowledge (or patterns) from large sets of data. These patterns can be in the form of business rules, affinities, correlations, trends, or prediction models (see Nemati and Barko, 2001). Most literature defines data mining as “the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data stored in structured databases,” where the data are organized in records structured by categorical, ordinal, and continuous variables (Fayyad et al., 1996, pp. 40–41). In this definition, the meanings of the key term are as follows:

- *Process* implies that data mining comprises many iterative steps.
- *Nontrivial* means that some experimentation-type search or inference is involved; that is, it is not as straightforward as a computation of predefined quantities.
- *Valid* means that the discovered patterns should hold true on new data with a sufficient degree of certainty.
- *Novel* means that the patterns are not previously known to the user within the context of the system being analyzed.
- *Potentially useful* means that the discovered patterns should lead to some benefit to the user or task.
- *Ultimately understandable* means that the pattern should make business sense that leads to the user saying, “Mmm! It makes sense; why didn’t I think of that,” if not immediately, at least after some postprocessing.

Data mining is not a new discipline, but rather a new definition for the use of many disciplines. Data mining is tightly positioned at the intersection of many disciplines, including statistics, artificial intelligence, machine learning, management science, information systems (IS), and databases (see Figure 5.1). Using advances in all of these disciplines, data mining strives to make progress in extracting useful information and knowledge from large databases. It is an emerging field that has attracted much attention in a very short time.

The following are the major characteristics and objectives of data mining:

- Data are often buried deep within very large databases, which sometimes contain data from several years. In many cases, the data are cleansed and consolidated into a data warehouse. Data may be presented in a variety of formats (see Chapter 2 for a brief taxonomy of data).
- The data mining environment is usually a client/server architecture or a Web-based IS architecture.
- Sophisticated new tools, including advanced visualization tools, help to remove the information ore buried in corporate files or archival public records. Finding it involves massaging and synchronizing the data to get the right results. Cutting-edge data miners are also exploring the usefulness of soft data (i.e., unstructured text stored in such places as Lotus Notes databases, text files on the Internet, or enterprise-wide intranets).

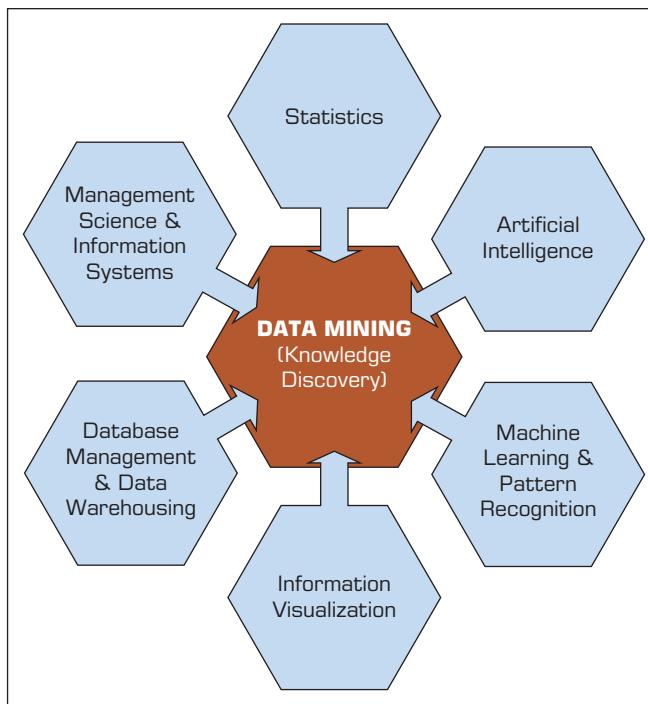


FIGURE 5.1 Data Mining Is a Blend of Multiple Disciplines.

- The miner is often an end user, empowered by data drills and other powerful query tools to ask ad hoc questions and obtain answers quickly, with little or no programming skill.
- Striking it rich often involves finding an unexpected result and requires end users to think creatively throughout the process, including the interpretation of the findings.
- Data mining tools are readily combined with spreadsheets and other software development tools. Thus, the mined data can be analyzed and deployed quickly and easily.
- Because of the large amounts of data and massive search efforts, it is sometimes necessary to use parallel processing for data mining.

A company that effectively leverages data mining tools and technologies can acquire and maintain a strategic competitive advantage. Data mining offers organizations an indispensable decision-enhancing environment to exploit new opportunities by transforming data into a strategic weapon. See Nemati and Barko (2001) for a more detailed discussion on the strategic benefits of data mining.

How Data Mining Works

Using existing and relevant data obtained from within and outside the organization, data mining builds models to discover patterns among the attributes presented in the data set. Models are the mathematical representations (simple linear relationships/affinities and/or complex and highly nonlinear relationships) that identify the patterns among the attributes of the things (e.g., customers, events) described within the data set. Some of these patterns are explanatory (explaining the interrelationships and affinities among the attributes), whereas others are predictive (foretelling future values of certain attributes). In general, data mining seeks to identify four major types of patterns:

- *Associations* find the commonly co-occurring groupings of things, such as beer and diapers going together in market-basket analysis.

- *Predictions* tell the nature of future occurrences of certain events based on what has happened in the past, such as predicting the winner of the Super Bowl or forecasting the absolute temperature of a particular day.
- *Clusters* identify natural groupings of things based on their known characteristics, such as assigning customers in different segments based on their demographics and past purchase behaviors.
- *Sequential relationships* discover time-ordered events, such as predicting that an existing banking customer who already has a checking account will open a savings account followed by an investment account within a year.

These types of patterns have been *manually* extracted from data by humans for centuries, but the increasing volume of data in modern times has created a need for more automatic approaches. As data sets have grown in size and complexity, direct manual data analysis has increasingly been augmented with indirect, automatic data processing tools that use sophisticated methodologies, methods, and algorithms. The manifestation of such evolution of automated and semiautomated means of processing large data sets is now commonly referred to as *data mining*.

Generally speaking, data mining tasks can be classified into three main categories: prediction, association, and clustering. Based on the way in which the patterns are extracted from the historical data, the learning algorithms of data mining methods can be classified as either supervised or unsupervised. With supervised learning algorithms, the training data includes both the descriptive attributes (i.e., independent variables or decision variables) as well as the class attribute (i.e., output variable or result variable). In contrast, with unsupervised learning the training data includes only the descriptive attributes. Figure 5.2 shows a simple taxonomy for data mining tasks, along with the learning methods and popular algorithms for each of the data mining tasks.

PREDICTION **Prediction** is commonly referred to as the act of telling about the future. It differs from simple guessing by taking into account the experiences, opinions, and other relevant information in conducting the task of foretelling. A term that is commonly associated with prediction is *forecasting*. Even though many believe that these two terms are synonymous, there is a subtle but critical difference between the two. Whereas prediction is largely experience and opinion based, forecasting is data and model based. That is, in order of increasing reliability, one might list the relevant terms as *guessing*, *predicting*, and *forecasting*, respectively. In data mining terminology, *prediction* and *forecasting* are used synonymously, and the term *prediction* is used as the common representation of the act. Depending on the nature of what is being predicted, prediction can be named more specifically as classification (where the predicted thing, such as tomorrow's forecast, is a class label such as "rainy" or "sunny") or regression (where the predicted thing, such as tomorrow's temperature, is a real number, such as "65°F").

CLASSIFICATION **Classification**, or supervised induction, is perhaps the most common of all data mining tasks. The objective of classification is to analyze the historical data stored in a database and automatically generate a model that can predict future behavior. This induced model consists of generalizations over the records of a training data set, which help distinguish predefined classes. The hope is that the model can then be used to predict the classes of other unclassified records and, more important, to accurately predict actual future events.

Common classification tools include neural networks and decision trees (from machine learning), logistic regression and discriminant analysis (from traditional statistics), and emerging tools such as rough sets, support vector machines (SVMs), and genetic algorithms. Statistics-based classification techniques (e.g., logistic regression and

| Data Mining Tasks & Methods | Data Mining Algorithms | Learning Type |
|-----------------------------|--|---------------|
| Prediction | | |
| Classification | Decision Trees, Neural Networks, Support Vector Machines, kNN, Naïve Bayes, GA | Supervised |
| Regression | Linear/Nonlinear Regression, ANN, Regression Trees, SVM, kNN, GA | Supervised |
| Time Series | Autoregressive Methods, Averaging Methods, Exponential Smoothing, ARIMA | Supervised |
| Association | | |
| Market-basket | Apriori, OneR, ZeroR, Eclat, Genetic Algorithm (GA) | Unsupervised |
| Link analysis | Expectation Maximization, Apriori Algorithm, Graph-based Matching | Unsupervised |
| Sequence analysis | Apriori Algorithm, FP-Growth, Graph-based Matching | Unsupervised |
| Segmentation | | |
| Clustering | K-means, K-modes, Expectation Maximization (EM), DBSCAN | Unsupervised |
| Outlier analysis | DBSCAN, Expectation Maximization (EM), Bayesian and Gaussian Models | Unsupervised |

FIGURE 5.2 A Simple Taxonomy for Data Mining Tasks, Methods, and Algorithms.

discriminant analysis) have received their share of criticism—that they make unrealistic assumptions about the data, such as independence and normality—which limit their use in classification-type data mining projects.

Neural networks involve the development of mathematical structures (somewhat resembling the biological neural networks in the human brain) that have the capability to learn from past experiences presented in the form of well-structured data sets. They tend to be more effective when the number of variables involved is rather large and the relationships among them are complex and imprecise. Neural networks have disadvantages

as well as advantages. For example, it is usually very difficult to provide a good rationale for the predictions made by a neural network. Also, neural networks tend to need considerable training. Unfortunately, the time needed for training tends to increase exponentially as the volume of data increases, and in general, neural networks cannot be trained on very large databases. These and other factors have limited the applicability of neural networks in data-rich domains.

Decision trees classify data into a finite number of classes based on the values of the input variables. Decision trees are essentially a hierarchy of if-then statements and are thus significantly faster than neural networks. They are most appropriate for **categorical** and **interval data**. Therefore, incorporating continuous variables into a decision tree framework requires *discretization*; that is, converting continuous valued numerical variables to ranges and categories.

A related category of classification tools is rule induction. Unlike with a decision tree, with rule induction the if-then statements are induced from the training data directly, and they need not be hierarchical in nature. Other, more recent techniques such as SVM, rough sets, and genetic algorithms are gradually finding their way into the arsenal of classification algorithms.

CLUSTERING **Clustering** partitions a collection of things (e.g., objects, events, presented in a structured data set) into segments (or natural groupings) whose members share similar characteristics. Unlike in classification, in clustering the class labels are unknown. As the selected algorithm goes through the data set, identifying the commonalities of things based on their characteristics, the clusters are established. Because the clusters are determined using a heuristic-type algorithm, and because different algorithms may end up with different sets of clusters for the same data set, before the results of clustering techniques are put to actual use it may be necessary for an expert to interpret, and potentially modify, the suggested clusters. After reasonable clusters have been identified, they can be used to classify and interpret new data.

Not surprisingly, clustering techniques include optimization. The goal of clustering is to create groups so that the members within each group have maximum similarity and the members across groups have minimum similarity. The most commonly used clustering techniques include *k*-means (from statistics) and self-organizing maps (from machine learning), which is a unique neural network architecture developed by Kohonen (1982).

Firms often effectively use their data mining systems to perform market segmentation with cluster analysis. Cluster analysis is a means of identifying classes of items so that items in a cluster have more in common with each other than with items in other clusters. It can be used in segmenting customers and directing appropriate marketing products to the segments at the right time in the right format at the right price. Cluster analysis is also used to identify natural groupings of events or objects so that a common set of characteristics of these groups can be identified to describe them.

ASSOCIATIONS **Associations**, or *association rule learning in data mining*, is a popular and well-researched technique for discovering interesting relationships among variables in large databases. Thanks to automated data-gathering technologies such as bar code scanners, the use of association rules for discovering regularities among products in large-scale transactions recorded by point-of-sale systems in supermarkets has become a common knowledge discovery task in the retail industry. In the context of the retail industry, association rule mining is often called *market-basket analysis*.

Two commonly used derivatives of association rule mining are **link analysis** and **sequence mining**. With link analysis, the linkage among many objects of interest is discovered automatically, such as the link between Web pages and referential relationships among groups of academic publication authors. With sequence mining, relationships

are examined in terms of their order of occurrence to identify associations over time. Algorithms used in association rule mining include the popular Apriori (where frequent itemsets are identified) and FP-Growth, OneR, ZeroR, and Eclat.

VISUALIZATION AND TIME-SERIES FORECASTING Two techniques often associated with data mining are *visualization* and *time-series forecasting*. Visualization can be used in conjunction with other data mining techniques to gain a clearer understanding of underlying relationships. As the importance of visualization has increased in recent years, a new term, *visual analytics*, has emerged. The idea is to combine analytics and visualization in a single environment for easier and faster knowledge creation. Visual analytics is covered in detail in Chapter 3. In time-series forecasting, the data consists of values of the same variable that is captured and stored over time in regular intervals. These data are then used to develop forecasting models to extrapolate the future values of the same variable.

Data Mining versus Statistics

Data mining and statistics have a lot in common. They both look for relationships within data. Most people call statistics the “foundation of data mining.” The main difference between the two is that statistics starts with a well-defined proposition and hypothesis, whereas data mining starts with a loosely defined discovery statement. Statistics collects sample data (i.e., primary data) to test the hypothesis, whereas data mining and analytics use all the existing data (i.e., often observational, secondary data) to discover novel patterns and relationships. Another difference comes from the size of data that they use. Data mining looks for data sets that are as “big” as possible, whereas statistics looks for the right size of data (if the data are larger than what is needed/required for the statistical analysis, a sample of the data is used). The meaning of “large data” is rather different between statistics and data mining. A few hundred to a thousand data points are large enough to a statistician, but several million to a few billion data points are considered large for data mining studies.

► SECTION 5.2 REVIEW QUESTIONS

1. Define *data mining*. Why are there many different names and definitions for data mining?
2. What recent factors have increased the popularity of data mining?
3. Is data mining a new discipline? Explain.
4. What are some major data mining methods and algorithms?
5. What are the key differences between the major data mining tasks?

5.3 DATA MINING APPLICATIONS

Data mining has become a popular tool in addressing many complex businesses problems and opportunities. It has been proven to be very successful and helpful in many areas, some of which are shown by the following representative examples. The goal of many of these business data mining applications is to solve a pressing problem or to explore an emerging business opportunity to create a sustainable competitive advantage.

- **Customer relationship management.** Customer relationship management (CRM) is the extension of traditional marketing. The goal of CRM is to create one-on-one relationships with customers by developing an intimate understanding of their needs and wants. As businesses build relationships with their customers over time through a variety of interactions (e.g., product inquiries, sales, service requests,

warranty calls, product reviews, social media connections), they accumulate tremendous amounts of data. When combined with demographic and socioeconomic attributes, this information-rich data can be used to (1) identify most likely responders/buyers of new products/services (i.e., customer profiling), (2) understand the root causes of customer attrition to improve customer retention (i.e., churn analysis), (3) discover time-variant associations between products and services to maximize sales and customer value, and (4) identify the most profitable customers and their preferential needs to strengthen relationships and to maximize sales.

- **Banking.** Data mining can help banks with the following: (1) automating the loan application process by accurately predicting the most probable defaulters, (2) detecting fraudulent credit card and online banking transactions, (3) identifying ways to maximize customer value by selling them products and services that they are most likely to buy, and (4) optimizing the cash return by accurately forecasting the cash flow on banking entities (e.g., ATM machines, banking branches).
- **Retailing and logistics.** In the retailing industry, data mining can be used to (1) predict accurate sales volumes at specific retail locations to determine correct inventory levels, (2) identify sales relationships between different products (with market-basket analysis) to improve the store layout and optimize sales promotions, (3) forecast consumption levels of different product types (based on seasonal and environmental conditions) to optimize logistics and, hence, maximize sales, and (4) discover interesting patterns in the movement of products (especially for the products that have a limited shelf life because they are prone to expiration, perishability, and contamination) in a supply chain by analyzing sensory and radio-frequency identification (RFID) data.
- **Manufacturing and production.** Manufacturers can use data mining to (1) predict machinery failures before they occur through the use of sensory data (enabling what is called *condition-based maintenance*), (2) identify anomalies and commonalities in production systems to optimize manufacturing capacity, and (3) discover novel patterns to identify and improve product quality.
- **Brokerage and securities trading.** Brokers and traders use data mining to (1) predict when and how much certain bond prices will change, (2) forecast the range and direction of stock fluctuations, (3) assess the effect of particular issues and events on overall market movements, and (4) identify and prevent fraudulent activities in securities trading.
- **Insurance.** The insurance industry uses data mining techniques to (1) forecast claim amounts for property and medical coverage costs for better business planning, (2) determine optimal rate plans based on the analysis of claims and customer data, (3) predict which customers are more likely to buy new policies with special features, and (4) identify and prevent incorrect claim payments and fraudulent activities.
- **Computer hardware and software.** Data mining can be used to (1) predict disk drive failures well before they actually occur, (2) identify and filter unwanted Web content and e-mail messages, (3) detect and prevent computer network security breaches and (4) identify potentially unsecure software products.
- **Government and defense.** Data mining also has a number of military applications. It can be used to (1) forecast the cost of moving military personnel and equipment; (2) predict an adversary's moves and, hence, develop more successful strategies for military engagements; (3) predict resource consumption for better planning and budgeting; and (4) identify classes of unique experiences, strategies, and lessons learned from military operations for better knowledge sharing throughout the organization.
- **Travel industry (airlines, hotels/resorts, rental car companies).** Data mining has a variety of uses in the travel industry. It is successfully used to (1)

predict sales of different services (seat types in airplanes, room types in hotels/resorts, car types in rental car companies) in order to optimally price services to maximize revenues as a function of time-varying transactions (commonly referred to as *yield management*), (2) forecast demand at different locations to better allocate limited organizational resources, (3) identify the most profitable customers and provide them with personalized services to maintain their repeat business, and (4) retain valuable employees by identifying and acting on the root causes for attrition.

- **Healthcare.** Data mining has a number of healthcare applications. It can be used to (1) identify people without health insurance and the factors underlying this undesired phenomenon, (2) identify novel cost-benefit relationships between different treatments to develop more effective strategies, (3) forecast the level and the time of demand at different service locations to optimally allocate organizational resources, and (4) understand the underlying reasons for customer and employee attrition.
- **Medicine.** Use of data mining in medicine should be viewed as an invaluable complement to traditional medical research, which is mainly clinical and biological in nature. Data mining analyses can (1) identify novel patterns to improve survivability of patients with cancer, (2) predict success rates of organ transplantation patients to develop better organ donor matching policies, (3) identify the functions of different genes in the human chromosome (known as genomics), and (4) discover the relationships between symptoms and illnesses (as well as illnesses and successful treatments) to help medical professionals make informed and correct decisions in a timely manner.
- **Entertainment industry.** Data mining is successfully used by the entertainment industry to (1) analyze viewer data to decide what programs to show during prime time and how to maximize returns by knowing where to insert advertisements, (2) predict the financial success of movies before they are produced to make investment decisions and to optimize the returns (Delen et al., 2007), (3) forecast the demand at different locations and different times to better schedule entertainment events and to optimally allocate resources, and (4) develop optimal pricing policies to maximize revenues.
- **Homeland security and law enforcement.** Data mining has a number of homeland security and law enforcement applications. Data mining is often used to (1) identify patterns of terrorist behaviors, (2) discover crime patterns (e.g., locations, timings, criminal behaviors, and other related attributes) to help solve criminal cases in a timely manner, (3) predict and eliminate potential biological and chemical attacks to the nation's critical infrastructure by analyzing special-purpose sensory data, and (4) identify and stop malicious attacks on critical information infrastructures (often called *information warfare*).
- **Sports.** Data mining was used to improve the performance of National Basketball Association (NBA) teams in the United States. Major League Baseball teams are into predictive analytics and data mining to optimally utilize their limited resources for a winning season (see Moneyball article in Chapter 1). In fact, most, if not all, professional sports nowadays employ data crunchers and use data mining to increase their chances of winning. Data mining applications are not limited to professional sports. In a 2012 article, Delen, Cogdell, and Kasap (2012) developed data mining models to predict National Collegiate Athletic Association (NCAA) Bowl Game outcomes using a wide range of variables about the two opposing teams' previous games statistics (more details about this case study are provided in Chapter 2). Wright (2012) used a variety of predictors for examination of the NCAA men's basketball championship bracket (a.k.a. March Madness).

► SECTION 5.3 REVIEW QUESTIONS

1. What are the major application areas for data mining?
2. Identify at least five specific applications of data mining and list five common characteristics of these applications.
3. What do you think is the most prominent application area for data mining? Why?
4. Can you think of other application areas for data mining not discussed in this section? Explain.

5.4 DATA MINING PROCESS

To systematically carry out data mining projects, a general process is usually followed. Based on best practices, data mining researchers and practitioners have proposed several processes (workflows or simple step-by-step approaches) to maximize the chances of success in conducting data mining projects. These efforts have led to several standardized processes, some of which (a few of the most popular ones) are described in this section.

One such standardized process, arguably the most popular one, Cross-Industry Standard Process for Data Mining—**CRISP-DM**—was proposed in the mid-1990s by a European consortium of companies to serve as a nonproprietary standard methodology for data mining (CRISP-DM, 2013 and Delen, 2014). Figure 5.3 illustrates this proposed process, which is a sequence of six steps that starts with a good understanding of the business and the need for the data mining project (i.e., the application domain) and ends with the deployment of the solution that satisfies the specific business need. Even though these steps are sequential in nature, there is usually a great deal of backtracking. Because the data mining is driven by experience and experimentation, depending on the problem

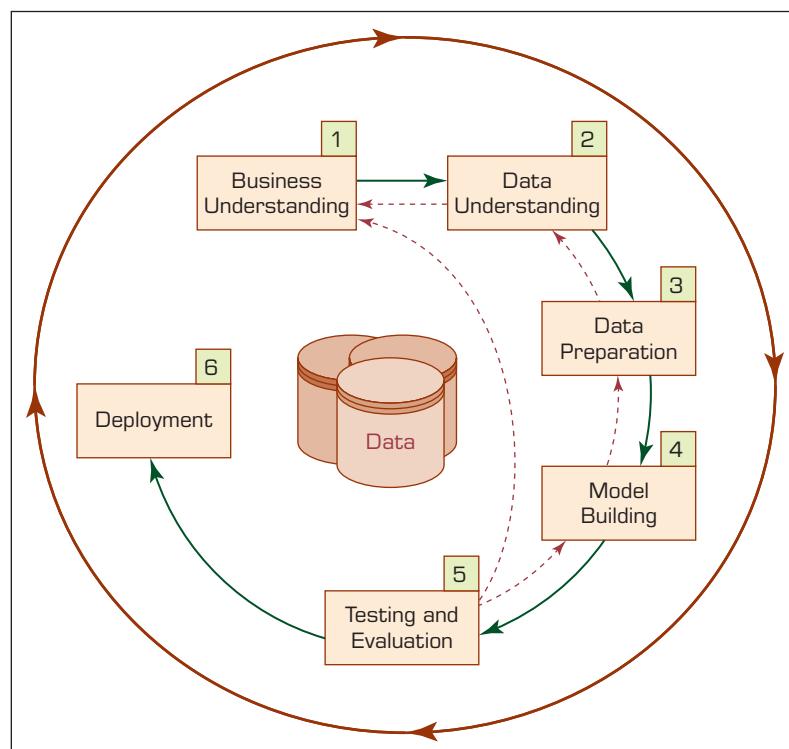


FIGURE 5.3 The Six-Step CRISP-DM Data Mining Process.

situation and the knowledge/experience of the analyst, the whole process can be very iterative (i.e., one should expect to go back and forth through the steps quite a few times) and time-consuming. Because later steps are built on the outcomes of the former ones, one should pay extra attention to the earlier steps in order not to put the whole study on an incorrect path from the onset.

Step 1: Business Understanding

The key element of any data mining study is to know what the study is for. Answering such a question begins with a thorough understanding of the managerial need for new knowledge and an explicit specification of the business objective regarding the study to be conducted. Specific goals such as “What are the common characteristics of the customers we have lost to our competitors recently?” or “What are typical profiles of our customers, and how much value does each of them provide to us?” are needed. Then a project plan for finding such knowledge is developed that specifies the people responsible for collecting the data, analyzing the data, and reporting the findings. At this early stage, a budget to support the study should also be established, at least at a high level with rough numbers.

Step 2: Data Understanding

A data mining study is specific to addressing a well-defined business task, and different business tasks require different sets of data. Following the business understanding, the main activity of the data mining process is to identify the relevant data from many available databases. Some key points must be considered in the data identification and selection phase. First and foremost, the analyst should be clear and concise about the description of the data mining task so that the most relevant data can be identified. For example, a retail data mining project may seek to identify spending behaviors of shoppers who purchase seasonal clothes based on their demographics, credit card transactions, and socioeconomic attributes. Furthermore, the analyst should build an intimate understanding of the data sources (e.g., where the relevant data are stored and in what form; what the process of collecting the data is—automated versus manual; who the collectors of the data are and how often the data are updated) and the variables (e.g., What are the most relevant variables? Are there any synonymous and/or homonymous variables? Are the variables independent of each other—do they stand as a complete information source without overlapping or conflicting information?).

To better understand the data, the analyst often uses a variety of statistical and graphical techniques, such as simple statistical summaries of each variable (e.g., for numeric variables the average, minimum/maximum, median, and standard deviation are among the calculated measures, whereas for categorical variables the mode and frequency tables are calculated), correlation analysis, scatterplots, histograms, and box plots. A careful identification and selection of data sources and the most relevant variables can make it easier for data mining algorithms to quickly discover useful knowledge patterns.

Data sources for data selection can vary. Traditionally, data sources for business applications include demographic data (such as income, education, number of households, and age), sociographic data (such as hobby, club membership, and entertainment), transactional data (sales record, credit card spending, issued checks), and so on. Nowadays, data sources also use external (open or commercial) data repositories, social media, and machine-generated data.

Data can be categorized as quantitative and qualitative. Quantitative data are measured using numeric values, or **numeric data**. It can be discrete (such as integers) or continuous (such as real numbers). Qualitative data, also known as categorical data, contains both nominal and ordinal data. **Nominal data** has finite nonordered values (e.g., gender data, which may have several values). **Ordinal data** has finite ordered values. For

example, customer credit ratings are considered ordinal data because the ratings can be excellent, fair, and bad. A simple taxonomy of data (i.e., the nature of data) is provided in Chapter 2.

Quantitative data can be readily represented by some sort of probability distribution. A probability distribution describes how the data are dispersed and shaped. For instance, normally distributed data are symmetric and is commonly referred to as being a bell-shaped curve. Qualitative data may be coded to numbers and then described by frequency distributions. Once the relevant data are selected according to the data mining business objective, data preprocessing should be pursued.

Step 3: Data Preparation

The purpose of data preparation (more commonly called *data preprocessing*) is to take the data identified in the previous step and prepare it for analysis by data mining methods. Compared to the other steps in CRISP-DM, data preprocessing consumes the most time and effort; most believe that this step accounts for roughly 80% of the total time spent on a data mining project. The reason for such an enormous effort spent on this step is the fact that real-world data are generally incomplete (lacking attribute values, lacking certain attributes of interest, or containing only aggregate data), noisy (containing errors or outliers), and inconsistent (containing discrepancies in codes or names). The nature of the data and the issues related to preprocessing of data for analytics are explained in detail in Chapter 2.

Step 4: Model Building

In this step, various modeling techniques are selected and applied to an already prepared data set to address the specific business need. The model-building step also encompasses the assessment and comparative analysis of the various models built. Because there is not a universally known *best* method or algorithm for a data mining task, one should use a variety of viable model types along with a well-defined experimentation and assessment strategy to identify the “best” method for a given purpose. Even for a single method or algorithm, a number of parameters need to be calibrated to obtain optimal results. Some methods may have specific requirements in the way that the data are to be formatted; thus, stepping back to the data preparation step is often necessary. Analytics in Action 5.1 presents a research study where a number of model types are developed and compared to each other.

Depending on the business need, the data mining task can be of a prediction (either classification or regression), an association, or a clustering type. Each of these data mining tasks can use a variety of data mining methods and algorithms. Some of these data mining methods were explained earlier in this chapter, and some of the most popular algorithms, including decision trees for classification, *k*-means for clustering, and the Apriori algorithm for association rule mining, are described later in this chapter.

ANALYTICS IN ACTION 5.1 Data Mining Helps in Cancer Research

According to the American Cancer Society, half of all men and one-third of all women in the United States will develop cancer during their lifetimes; approximately 1.5 million new cancer cases were expected to be diagnosed in 2013. Cancer is the second-most-common cause of death in the United States and in the world, exceeded only by cardiovascular disease. This year, over 500,000 Americans are expected to die of cancer—more than 1,300 people a day—accounting for nearly one of every four deaths.

Cancer is a group of diseases generally characterized by uncontrolled growth and spread of abnormal cells. If the growth and/or spread are not controlled, it can result in death. Even though the exact reasons are not known, cancer is believed to be caused by both external factors (e.g., tobacco, infectious organisms, chemicals, and radiation) and internal factors (e.g.,

inherited mutations, hormones, immune conditions, and mutations that occur from metabolism). These causal factors may act together or in sequence to initiate or promote carcinogenesis. Cancer is treated with surgery, radiation, chemotherapy, hormone therapy, biological therapy, and targeted therapy. Survival statistics vary greatly by cancer type and stage at diagnosis.

The five-year relative survival rate for all cancers is improving, and decline in cancer mortality had reached 20%, translating into the avoidance of about 1.2 million deaths from cancer since 1991. That's more than 400 lives saved per day! The improvement in survival reflects progress in diagnosing certain cancers at an earlier stage and improvements in treatment. Further improvements are needed to prevent and treat cancer.

Even though cancer research has traditionally been clinical and biological in nature, in recent years data-driven analytic studies have become a common complement. In medical domains where data- and analytics-driven research have been applied successfully, novel research directions have been identified to further advance the clinical and biological studies. Using various types of data, including molecular, clinical, literature-based, and clinical trial data, along with suitable data mining tools and techniques, researchers have been able to identify novel patterns, paving the road toward a cancer-free society.

In one study, Delen (2009) used three popular data mining techniques (decision trees, artificial neural networks, and SVMs) in conjunction with logistic regression to develop prediction models for prostate cancer survivability. The data set contained around 120,000 records and 77 variables. A *k*-fold cross-validation methodology was used in model building, evaluation, and comparison. The results showed that support vector models are the most accurate predictor (with a test set accuracy of 92.85%) for this domain, followed by artificial neural networks and decision trees. Furthermore, using a sensitivity analysis-based evaluation method, the study also revealed novel patterns related to prognostic factors of prostate cancer.

In a related study, Delen, Walker, and Kadam (2005) used two data mining algorithms (artificial neural networks and decision trees) and logistic regression to develop prediction models for breast cancer survival using a large data set (more than 200,000 cases). Using a 10-fold cross-validation method to measure the unbiased estimate of the prediction models for performance comparison purposes, the results indicated that the decision tree (C5 algorithm) was the best predictor, with 93.6% accuracy on the holdout sample (which was the best prediction accuracy reported in the literature), followed by artificial neural networks, with 91.2% accuracy, and logistic regression, with 89.2% accuracy. Further analysis of prediction models revealed prioritized importance of the prognostic factors, which can then be used as a basis for further clinical and biological research studies.

In the most recent study, Zolbanin, Delen, and Zadeh (2015) studied the impact of comorbidity in cancer survivability. Although prior research has shown that diagnostic and treatment recommendations might be altered based on the severity of comorbidities, chronic diseases are still being investigated in isolation from one another in most cases. To illustrate the significance of concurrent chronic diseases in the course of treatment, their study used the Surveillance, Epidemiology, and End Results (SEER) Program's cancer data to create two comorbid data sets: one for breast and female genital cancers and another for prostate and urinal cancers. Several popular machine-learning techniques are then applied to the resultant data sets to build predictive models (see Figure 5.4). Comparison of the results has shown that having more information about comorbid conditions of patients can improve models' predictive power, which in turn can help practitioners make better diagnostic and treatment decisions. Therefore, the study suggested that proper identification, recording, and use of patients' comorbidity status can potentially lower treatment costs and ease the healthcare-related economic challenges.

These examples (among many others in the medical literature) show that advanced data mining techniques can be used to develop models that possess a high degree of predictive as well as explanatory power. Although data mining methods are capable of extracting patterns and relationships hidden deep in large and complex medical databases, without the cooperation and feedback from the medical experts, their results are not of much use. The patterns found via data mining methods should be evaluated by medical professionals who have years of experience in the problem domain to decide whether they are logical, actionable, and novel enough to warrant new research directions. In short, data mining is not meant to replace medical professionals and researchers, but to complement their invaluable efforts to provide data-driven new research directions and to ultimately save more human lives.

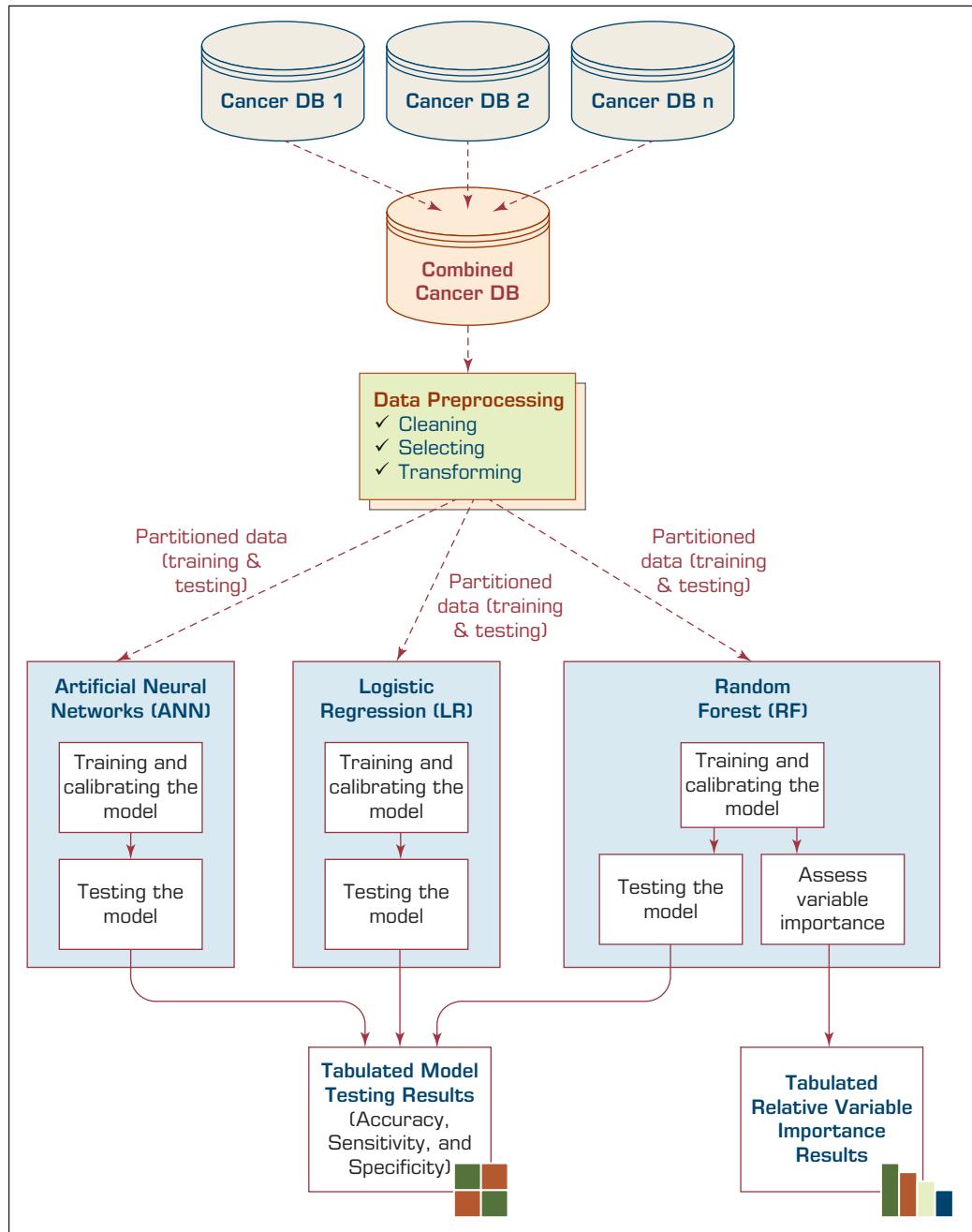


FIGURE 5.4 A Data Mining Methodology for Investigation of Comorbidity in Cancer Survivability.

Sources: Zolbanin, H. M., Delen, D., & Zadeh, A. H. (2015). Predicting overall survivability in comorbidity of cancers: A data mining approach. *Decision Support Systems*, 74, 150–161; Delen, D. (2009). Analysis of cancer data: A data mining approach. *Expert Systems*, 26(1), 100–112; Thongkam, J., Xu, G., Zhang, Y., & Huang, F. (2009). Toward breast cancer survivability prediction models through improving training space. *Expert Systems with Applications*, 36(10), 12200–12209; Delen, D., Walker, G., & Kadam, A. (2005). Predicting breast cancer survivability: A comparison of three data mining methods. *Artificial Intelligence in Medicine*, 34(2), 113–127.

Step 5: Testing and Evaluation

In step 5, the developed models are assessed and evaluated for their accuracy and generality. This step assesses the degree to which the selected model (or models) meets the business objectives and, if so, to what extent (i.e., Do more models need to be developed and assessed?). Another option is to test the developed model(s) in a real-world scenario if time and budget constraints permit. Even though the outcome of the developed models is expected to relate to the original business objectives, other findings that are not necessarily related to the original business objectives but that might also unveil additional information or hints for future directions often are discovered.

The testing and evaluation step is a critical and challenging task. No value is added by the data mining task until the business value obtained from discovered knowledge patterns is identified and recognized. Determining the business value from discovered knowledge patterns is somewhat similar to playing with puzzles. The extracted knowledge patterns are pieces of the puzzle that need to be put together in the context of the specific business purpose. The success of this identification operation depends on the interaction among data analysts, business analysts, and decision makers (such as business managers). Because data analysts may not have the full understanding of the data mining objectives and what they mean to the business and the business analysts, and decision makers may not have the technical knowledge to interpret the results of sophisticated mathematical solutions, interaction among them is necessary. To properly interpret knowledge patterns, it is often necessary to use a variety of tabulation and visualization techniques (e.g., pivot tables, cross-tabulation of findings, pie charts, histograms, box plots, scatterplots).

Step 6: Deployment

Development and assessment of the models is not the end of the data mining project. Even if the purpose of the model is to have a simple exploration of the data, the knowledge gained from such exploration will need to be organized and presented in a way that the end user can understand and benefit from. Depending on the requirements, the deployment phase can be as simple as generating a report or as complex as implementing a repeatable data mining process across the enterprise. In many cases, it is the customer, not the data analyst, who carries out the deployment steps. However, even if the analyst will not carry out the deployment effort, it is important for the customer to understand up front what actions need to be carried out to actually make use of the created models.

The deployment step may also include maintenance activities for the deployed models. Because everything about the business is constantly changing, the data that reflect the business activities also are changing. Over time, the models (and the patterns embedded within them) built on the old data may become obsolete, irrelevant, or misleading. Therefore, monitoring and maintenance of the models are important if the data mining results are to become a part of the day-to-day business and its environment. A careful preparation of a maintenance strategy helps to avoid unnecessarily long periods of incorrect usage of data mining results. To monitor the deployment of the data mining result(s), the project needs a detailed plan on the monitoring process, which may not be a trivial task for complex data mining models.

Other Data Mining Standardized Processes and Methodologies

To be applied successfully, a data mining study must be viewed as a process that follows a standardized methodology rather than as a set of automated software tools and techniques. In addition to CRISP-DM, there is another well-known methodology developed by the SAS Institute, called SEMMA (2009). The acronym **SEMMA** stands for “sample, explore, modify, model, and assess.”

Negotiating Access & Research Ethics

Saunders et al. (2023; 220-280)

Chapter 6



Negotiating access and research ethics

Learning outcomes

By the end of this chapter you should be able to:

- discuss the characteristics of access;
- demonstrate an awareness of issues associated with gaining traditional, Internet-mediated, intranet-mediated and hybrid access;
- evaluate a range of strategies to help you to gain access to organisations, groups and to individual participants;
- discuss the importance of research ethics and the need to act ethically;
- identify ethical issues at each stage of your research;
- recall the principles of data protection and data management;
- ensure you research ethically and progress your research project by gaining access;
- apply the knowledge, skills and understanding gained to your own research project.

6.1 Introduction

Many students want to start their research as soon as they have identified a topic area, forgetting that access and ethics are critical aspects for the success of any research project. Such considerations are equally important whether you are using secondary data or collecting primary data and whether you are conducting your research face-to-face or remotely (perhaps using online questionnaires or Internet-mediated interviews). Consequently, you need to think carefully about how access can be gained to collect your data and about possible ethical concerns that could arise throughout your research project. Without paying careful attention to both of these aspects, what seems like a good idea for your research may prove impractical or problematic, potentially causing harm to those you research.

Business and management research almost inevitably involves people. Ethical concerns are greatest where research involves human participants, irrespective of whether the research is conducted remotely or face-to-face. Your university, as well as an increasing number of



organisations, will require you to obtain formal research ethics committee approval (or a favourable ethical opinion) for your proposed research prior to commencing data collection. Universities and other organisations help facilitate the process of ethical scrutiny and approval by providing ethical guidelines for researchers. We consider ethical guidelines later, but it is worth noting that being clear to others, including participants, about how you will address potential ethical concerns is a crucial aspect of your research project, as our opening vignette illustrates.

Understanding the risk of developing Covid-19

Throughout the pandemic, research has been conducted to understand more about the disease. One study, in which Mark took part, was to understand how many

people in England have antibodies that might protect against Covid-19. Like other participants, Mark was invited to take part by a letter signed by the Parliamentary Under Secretary of State for Health and Social Care, the co-director of the university research institute and the director of the independent research organisation who



Source: MikeDotta/Shutterstock



were undertaking the study on behalf of the government. This initial letter stressed that participation was extremely important, that he had been chosen at random and that his taking part was completely voluntary. It also provided details of the research, which included doing a finger-prick test at home and completing an online questionnaire, how to sign up to the study online, offering a free telephone number to register for those who did not have access to the Internet. The back of the letter offered a variety of additional information including more detail on who was carrying out the study, how those invited to take part were chosen at random, the nature of the test and what would happen with the data. Of particular interest was the assurance the data would be kept confidential and that nobody would be identifiable from the published results. The data could, with agreement, be linked to other health data already held by the National Health Service about those who participated.

Having agreed to take part, like other participants, Mark received his Covid-19 in-home antibody testing kit by post along with a letter explaining how to take part and a detailed information sheet. Among other aspects, the information sheet reiterated that whether or not he decided to participate in the study was entirely up to him and that he could change his mind at any time without giving a reason. It also explained how the data collected would be stored and how each participant would be given a unique security number to make sure they could not be identified from their data. As before, additional contact details were provided as well as details about how to make a complaint.

Following the instructions, Mark completed the antibody test, took a photograph of the results and uploaded this to the website. Here he also completed an online questionnaire.

In this chapter we start by considering the characteristics of access and issues of feasibility and sufficiency in relation to gaining access (Section 6.2). Section 6.3 examines issues associated with researcher status, and Section 6.4 explores issues associated with Internet-mediated access. Section 6.5 discusses a number of established strategies to help you gain access to organisations and to your intended participants within these. Section 6.5 provides an overview of research ethics, the process of ethical reviews and outlines ethical principles for both generally and for Internet-mediated research. Section 6.6 anticipates the scope for ethical issues to occur during the various stages of your research project. Finally, Section 6.7 introduces data protection principles and data management, both of which you will need to consider in order to manage and store your data ethically.

6.2 Characteristics of access

Your ability to collect your own primary, or obtain secondary, data will depend on gaining access to an appropriate source or sources. The appropriateness of a source depends on your research question and objectives and the research design (Chapter 5). In this section we discuss three overlapping characteristics that will help you decide the precisely what is required: the type, nature and level of access (Figure 6.1).

Type of access

Types of access focus upon the nature of interactions with participants used to collect data. **Traditional access** involves face-to-face interactions (to conduct experiments, interviews, focus groups, observations or deliver and collect questionnaires), telephone conversations (for telephone interviews), correspondence (for postal questionnaires) or visiting

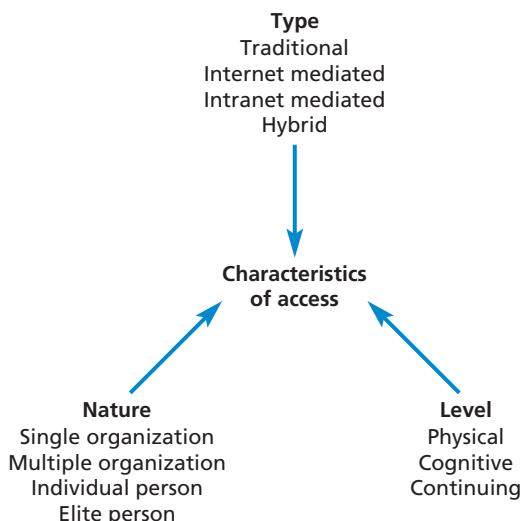


Figure 6.1 Characteristics of access

data archives (such as record offices or organisational archives, where data are not available online). In contrast, **Internet-mediated access** involves the use of different online technologies (e.g. email, video-conferencing, text messaging and apps for cloud-based platforms), to gain virtual access to deliver questionnaires, conduct archival research, discussions, experiments and interviews or to gather secondary data. A variant of this is **intranet-mediated access**, where you seek to gain virtual access (usually as an employee or worker) from within an organisation using its intranet. **Hybrid access** is where you use elements of both Internet mediated and traditional means to gain access (Box 6.6).

Nature of access

Another characteristic of access is the nature of access; whether your research design will involve collecting data from participants in a single organisation or group, across multiple organisations or groups, or if access to an organisation or group is irrelevant. For many research projects it will be sufficient to gain access to one organisation to conduct research and collect data; **single-organisation access**. For other research projects **multi-organisation access** to two or more organisations will be necessary. For example, a researcher using a case study strategy may focus their research in one organisation, hoping to be able to negotiate access to intended participants within this single case (Box 5.12). Another researcher, also using a case study strategy, may decide that meeting her research question and objectives requires multiple case organisations (Section 5.5, Figure 5.5). Gaining research access to one organisation can be difficult. Gaining access to multiple organisations may be even more difficult and time-consuming as it will be necessary to repeat the process. We discuss strategies to gain access that apply to both single-organisation and multi-organisation research in Section 6.4.

Other research projects will not necessitate seeking organisational access. You may need to collect data directly from individuals with the same role such as digital marketing managers from a large number of organisations. Alternatively, you may wish to collect data from people who do not have an organisational affiliation such as staycation tourists. For other research projects you may wish to gain access to individuals who are notable

in their field (such as retired CEOs or social media influencers) but who may not have an organisational affiliation. Due to their notability, we refer to this as **elite person access**, distinguishing it from **individual person access** where an individual's neither affiliated to an organisation, nor considered to be elite.

Levels of access

Even where you wish to conduct your research within a single organisation, gaining access to intended participants is still likely to involve you in a multi-faceted process of negotiation across different **levels of access** (physical/virtual, continuing and cognitive). The level(s) you require will depend on your research question and objectives. For some research projects the person you approach to negotiate access will also be a person you wish to ask to participate in your research. This would be the case where you approach a financial manager in an organisation to ask him or her to take part in a research interview conducted by yourself. However, one of your research objectives may require you to ask members of staff in this finance department to participate in an online survey. In this case you would need to negotiate access not only with the finance manager but also with the members of staff in this department whom you wish to participate.

An initial agreement to conduct research is referred to as gaining **physical access** (Gummesson 2000) where the research will be face-to-face and gaining **virtual access** where it will be online. This can be difficult for three reasons. First, organisations, groups or individuals may not be prepared to engage in additional, voluntary activities because of the time and resources required. Many organisations receive frequent student requests for access and cooperation and would find it impossible to agree to all or even some of these. Second, the request for access may fail to interest the **gatekeeper** or **broker**, the person who controls access, making the decision whether or not to allow the researcher to undertake their research. This may be for a number of reasons, related to:

- a lack of perceived value in relation to the work of the organisation, group or the individual;
- the nature of the topic such as its potential sensitivity or concerns regarding confidentiality or sensitivity of the information required;
- perceptions regarding the researcher's lack of credibility or competence.

Finally, the organisation or group may have no choice but to refuse access for reasons unrelated to the research. An organisation may be undertaking a strategic review or considering whether to restructure its functions and therefore be unwilling to allow access at such a sensitive time. Even when someone is prepared to offer access, this may be overruled at a higher level in the organisation, due to some aspect of organisational politics. Such 'false starts' are invariably disappointing as you will need to find another organisation or group.

Even when you are able to negotiate physical or virtual access there are other levels of access that you will need to consider and plan for in your research strategy as access is usually a continuing process rather than a single event (Gummesson 2000; Marshall and Rossman 2021). Gaining access is often iterative and incremental, in which you gain entry to carry out the initial part of your research, and then seek further access in order to conduct another part (Box 6.1).

Physical or virtual access to an organisation will be granted formally through its management. However, to actually collect data from individuals, you will need to gain **cognitive access** from your intended participants, that is their acceptance and trust. Simply obtaining such access to an organisation is unlikely to be sufficient unless you are also



Box 6.1 Focus on student research

Negotiating access incrementally

Luc wished to undertake a series of Internet-mediated interviews in the departments and sections of a data management company. He initially managed to negotiate access to commence his research in the

management systems support department, where he was granted permission to interview a sample of information systems support workers. As a result of conducting these interviews, he was then granted access within the same department to interview a sample of staff in the information technology section. Following the conduct of these interviews, the department's management team agreed to support his attempt to negotiate further access to interview staff in the company's accounting, human resources, marketing and sales departments.

able to negotiate yourself into a position where you can obtain consent and collect data from participants accounts to answer your research question and meet your objectives.

Gaining cognitive access to intended participants will ultimately be determined by whether they consent to take part in your research. Whether or not potential research participants agree, for example, to complete a questionnaire or take part in a research interview is the subject of **Leverage-saliency theory** (Groves et al. 2000). This recognises that different people will respond to different levers such as a request to participate in research in different ways and that their response is dependent on this being made salient to them. The key is to understand what encourages and what discourages, potential participants to participate; the latter resulting in non-response.

Leverage-saliency theory recognises the different attributes associated with any request to participate in research can influence intended participants decision to take part. For example, the topic of the research may be more or less interesting to different potential participants (Groves et al. 2004). The way in which the request to participate in the research is presented is also likely to affect how intended participants respond to this. Other attributes that may affect participation include the offer of incentives to take part, the purpose and use of the research and the requirements of participating in the research (Groves et al. 2000; Groves et al. 2004; Trussell and Lavrakas 2004).

While thinking how you will leverage responses; it is important to ensure your request appeals to all potential participants and not just those interested in the research topic. This will help reduce **non-response error**, where non-respondents in your intended sample differ in meaningful ways from those who take part and, as a consequence, are likely to have different characteristics and viewpoints to those who do respond and participate.

The access you manage to negotiate will therefore impact on your ability to select suitable participants or secondary data and whether your data are likely to be valid and reliable (Box 6.2) or credible and dependable. Selecting a suitable sample of, for example, customers, clients or employees requires access to organisational data, either directly or indirectly, and a request that outlines precisely how the sample should be selected (Sections 7.2 and 7.3). For longitudinal studies using primary data, access to participants will be needed on two or more occasions.

Negotiating physical or, in the case of Internet-mediated research, virtual access will be important to gain **personal entry** to one or more organisations, being a precursor to developing cognitive access to allow you to collect the necessary data. The suitability of the access negotiated can be evaluated using the concepts of feasibility and sufficiency. **Feasibility** is concerned with whether it is practicable to negotiate access for your proposed research project. Your research design may be grand and elegant, but if it is not



Box 6.2 **Focus on student research**

Gaining access to a suitable sample

Maria wished to discover how component suppliers viewed the just-in-time delivery requirements of large manufacturing organisations that they supplied. Two large manufacturing organisations agreed to introduce her to a sample of their component suppliers

whom Maria could then interview. While undertaking her telephone interviews Maria noted that all of the interviewees' responses were extremely positive about the just-in-time delivery requirements of both large manufacturing organisations. As both manufacturing organisations had selected who she could interview, Maria wondered whether these extremely positive responses were typical of all the component suppliers used by these organisations or whether they were providing an unreliable and untypical picture.

possible to gain access to data then it will need revising. Once you have a proposal that you believe will be feasible in general terms, the next point to consider is whether you will be able to gain sufficient access to fulfil all of your research objectives. **Sufficiency** is the extent to which the access you negotiate will be enough to answer your proposed question and achieve your research objectives. You do not want to have to say, 'I could achieve research objectives a, b and c but not x, y and z!' Or, perhaps more likely, 'I can achieve research objectives a and b, but now I think about this carefully, I'm going to find it difficult to collect much data for c and x, which will then mean I can't do y and z!' You therefore need to consider fully the nature of the access that you will require and whether you will be able to gain sufficient access in practice to meet your objectives, to answer your research question. The clarity, which should result from having considered the nature and extent of the access that you require, will also be helpful in persuading organisations or groups to grant access since they are more likely to be convinced of your credibility and competence.

6.3 Researcher status

Even when you consider feasibility and sufficiency carefully, access is still unlikely to be straightforward, requiring persistence and emotional resilience (Peticcia-Harris et al. 2016). However, with careful planning you will be able to anticipate and, hopefully, overcome problems that occur in practice. Drawing on this, you may wish to consider where you are likely to be able to gain access and amend your topic and research design to reflect the nature of access. As Buchanan et al. (2013: 53–4) note:

Fieldwork is permeated with the conflict between what is theoretically desirable on the one hand and what is practically possible on the other. It is desirable to ensure representativeness in the sample, uniformity of interview procedures, adequate data collection across the range of topics to be explored and so on. But the members of organisations block access to information, constrain the time allowed for interviews, lose your questionnaires, go on holiday and join other organisations in the middle of your unfinished study. In the conflict between the desirable and the possible, the possible always wins.

Problems of access may also vary with regard to your status relative to the organisations, groups or people you wish to research. We now consider these in more detail.

Access issues as an external researcher

If you are approaching one or more organisations or groups where you have little or no prior contact, you will be an **external researcher**. You will need to negotiate access at each level discussed earlier (physical/virtual, cognitive and continuing). Operating as an external researcher is likely to pose problems, although it may have some benefits. Your lack of status in relation to an organisation or group in which you wish to conduct research will mean gaining physical or virtual access is a major issue, and this concern will remain in relation to negotiating cognitive access (Box 6.3). Goodwill on the part of the organisation or group and its members is something that external researchers have to rely on at each level of access. Your ability to demonstrate your competence and integrity, and in particular your ability to explain your research project clearly and concisely, will be critical. Alongside, the gatekeeper can also play an important role: creating awareness of your research, adding credibility by her or his intervention and introducing you and your research project to the relevant people.

Even where you are not seeking to access participants in an organisation or group, you will still need to demonstrate competence and integrity. Here your role as an external researcher may prove beneficial as participants are usually willing to accept an external researcher as objective and without a covert agenda.

Access issues as an internal researcher or participant researcher

As an organisational employee or group member you are likely to undertake research in your own group or organisation as an **internal researcher** or a **participant researcher**. Even in this role, you may face problems negotiating physical or continuing access and still need to



Box 6.3 Focus on student research

The impact of a researcher's organisational status

David recalls a case of mistaken identity. His research involved gaining access to several employers' and trade union organisations. Having gained access to the regional office of one such organisation, David read and noted various organisational documents kept there over a period of a few days. During the first day David was located in a large, comfortable room and frequently brought refreshments by the caretaker of the building.

This appeared to David to be very kind treatment. However, David did not know that a rumour had spread among some staff that he was from 'head office' and was there to 'monitor' in some way the work of the office. On attending the second day, David was met by the caretaker and taken to a small, plain room and no more refreshments appeared for the duration of the research visit. The rumour had been corrected!

Of course, this example of the effect of the researcher's (lack of) organisational status is most unfair on the large number of people who treat those who undertake research within their organisation very well in full knowledge of their status. However, it illustrates the way in which some people may react to perceptions about status.

obtain formal approval to undertake research from your organisation or group. As an internal researcher, potential participants may be suspicious about why you are undertaking your research project and the data collected will be used. Their views about the part of the organisation for which you work and your status may also affect whether you gain cognitive access. Such problems may be exacerbated if you are given a project to research, perhaps by your line manager or mentor, where others are aware that this is an issue about which management would like to implement change. This is particularly likely where resulting change is perceived as being harmful to potential participants. It may also suggest ethical concerns (Section 6.6).

6.4 Internet-mediated access

Internet-mediated access became the norm for much research during the Covid-19 pandemic, both for collecting primary and searching and locating secondary data. However, it still can be challenging to access to participants online who are suitable for your research and who match the characteristics of your intended sample (Sections 7.2 and 7.3). Similarly, it can be difficult to locate suitable secondary data (Section 8.4); both potentially leading to data quality issues.

Quantitative and qualitative data can be collected using the Internet. Online questionnaires can be accessed through a hyperlink displayed in an email or on a web page (Section 11.5). Experimental data may be collected online (Box 6.4). Internet-mediated observation can be conducted, using, for example, an online ethnographic research strategy (Section 9.5). Interviews or discussion groups may also take place online. These may be text based using SMS messaging applications, social networks or emails. They may also be conducted using telecommunication applications, such as Skype™ and Zoom™, to overcome the impersonal nature of a text-based Internet interview (Section 10.10). The Internet also provides access, through gateways and archives, to existing data sets that are available for secondary analysis (Section 8.2).

Online communities have generated extremely large amounts of material, especially qualitative but also quantitative, which subject to permission is often accessible to researchers.



Box 6.4 Focus on management research

Gaining access

Sara Althammer and colleagues (Althammer et al. 2021) conducted an online experiment over three weeks using a self-training intervention to research whether mindfulness interventions promoted work-life balance.

To gain access to potential participants for their experiment, the researchers used a range of different strategies. These included advertising flyers, professional email list servers, a snowball sampling approach

and asking their professional and social contacts. For each they promoted their research as a scientific project that offered a free three-week training intervention using mindfulness to detach from work and enhance work-life balance. Potential participants were asked to sign up online and subsequently complete a daily questionnaire. In total, 379 people signed up online to take part in the research, 99 subsequently not completing any of the daily questionnaires used to collect the data. A further 93 completed fewer than three of these questionnaires and were also excluded, the remaining participants completing questionnaires for an average of 9.46 days.

Based on their findings, the authors concluded that a brief mindfulness self-training intervention could foster detachment, reduce work-life conflict and improve satisfaction with work-life balance.

As these communities organise around an interest or a particular product, service, place or lifestyle, their forums and bulletin (message) boards can be used to post messages and create a discussion over time among members. Groups' email lists can, with permission, be used to start conversations. Linked web pages provide online community resources organised by interest, such as for consumer-to-consumer discussion. **Blogs** (web logs) and to a lesser extent tweets provide a public online journal or diary. Numerous bloggers comment on political events, often from the perspective of their political beliefs. Others comment on their shopping experiences and offer consumer advice or on their travel experiences (Jiang et al. 2021). Many blogs and bulletin boards are organised through content management systems, although these can be accessed through specialised blog search engines (Kozinets 2020).

Determining the type of access

Internet-mediated access is subject to the same issues that affect traditional access. While the Internet, and more specifically the use of web links, messaging apps, email, social networks, webcam and web conferencing, may facilitate communication between you and your participants; the most suitable way to conduct your research and negotiate access will depend on your research question and objectives. In some circumstances you may conclude that it is more effective to gain physical rather than virtual access using associated methods (Box 6.5).



Box 6.5 Focus on student research

Physical or virtual access and methods?

Sab had a keen interest in IT and thought that he would conduct his research using Internet-mediated access and data-collection methods. His research focused on the ways in which senior managers influence board-level strategic decision-making. His interest in this topic had developed after a fortuitous conversation with a senior personnel policy manager who worked for a large organisation, who had explained how in some cases strategy formation was influenced by promoting incremental changes rather than trying to bring about a radical change in one movement. This idea interested Sab and he formulated a research project to explore it in a range of organisational contexts. However, the more he thought about it and discussed it with his project tutor the more he realised that he would have to research it using traditional methods.

After negotiating physical access to interview six senior managers who worked in different functional areas in different organisations, he conducted a face-to-face in-depth, exploratory interview with each one. While

conducting these interviews he realised that the value and depth of the data he collected would have been much less if he had tried to conduct these online. His questioning was shaped by the data each participant shared with him during the interview. Because of the sensitive nature of the topic, most of the interviews took the form of discussions, allowing Sab to clarify points and ask for illustrative examples. As each interview progressed, he found that some of his participants were willing to show him quite sensitive documents in the privacy of the interview room (which was the manager's own office). He found that rapport and trust were vital to the conduct of each interview. He also found that conducting an interview at the organisation helped to focus his mind and enhance his understanding of the organisational context. This in turn helped him to make sense of the data his participants shared with him.

Sab concluded that first negotiating physical access and then developing cognitive access on a person-to-person basis at the start of each interview had been the most appropriate strategy to adopt and also the most effective. However, as he had met with each participant and established rapport and trust, he asked each one if he would be able to email any further questions for clarification. Some agreed but others said that they would prefer to undertake this either by telephone or another face-to-face discussion.

Where you decide to use Internet-mediated techniques, you are likely to need to negotiate virtual access and obtain permission from a broker or gatekeeper to gain access to a sample of organisational members (Box 6.6). Subsequently, you could advertise your research by a post or use an email list to invite potential respondents and including a hyperlink to your questionnaire (Section 11.5). Here you will need to ensure that your intended participants are aware of your research, its purpose, how it will be used, its nature and what will be required if they decide to participate in it. Their decision will be influenced by how well you explain the purpose, use and nature of your research and the requirements of taking part. This highlights how gaining access to an organisation and intended participants within it may involve a hybrid strategy. Such hybrid access strategies may be valuable where you wish to:

- achieve multi-organisation access and need to negotiate access to intended participants within several organisations;
- negotiate continuing access and meet with your organisational broker or gatekeeper and intended participants to develop rapport and demonstrate your competence and establish trust to achieve this.

Where you plan to conduct your research with individuals (individual person access or elite person access) online, you will need to identify an appropriate sample and then to negotiate virtual and cognitive access with these intended participants. The ability to



Box 6.6 Focus on student research

Where topic and strategy determined type of access

Elina's research focused on consumers' purchasing decisions. She was interested in assessing the relative importance of information obtained from online shopping sites and from high street shops in informing purchasing decisions for different product categories. These categories covered all of the products purchased by her age group, such as people on her marketing course.

Elina had formulated a mixed methods research design. She had designed a web questionnaire that asked respondents to identify actual recent purchasing decisions related to the categories in which she was interested. For each of these, where applicable, she asked questions about the product, the sources of information used to inform the purchase decision and the way in which these sources determined the purchasing decision. Following ethical approval from her university, she asked the course leader if they would send a message using the university's virtual learning

environment to each person on her course asking for their help and containing a hyperlink to the questionnaire. The questionnaire included a question asking each respondent if they were willing to help further by completing an electronic diary. Those who answered yes were asked to provide their email address so Elina could send them the diary.

Elina emailed the template of the electronic diary to all those willing to help further. She had designed this to allow respondents to record purchasing decisions related to her list of product categories, the sources of information used to inform these purchases and the way in which these sources determined the decision. Respondents returned the diary as an email attachment.

Elina was aware that her request to maintain an electronic diary of influences on purchasing decisions would sensitise respondents to their use of different information sources, so had distributed the questionnaire first. This she felt would help her judge the extent the participant had been sensitised as well as about the relative impact of these different sources.

Her use of an Internet-mediated access strategy proved successful in gaining access to both questionnaire respondents and a group of people who would keep a diary.

identify your sample will be a key determinant of the feasibility of this approach. If you are unable to find an appropriate sample yourself (Section 6.5), websites such as ‘Call for Participants’ may be helpful. These allow researchers including students to advertise surveys, interviews and other research studies to potential participants worldwide free of charge (Call for Participants 2021). However, you will need to establish whether using this or a similar online platform can provide access to appropriate participants.

6.5 Strategies to gain access

We now consider strategies that can be used in combination to gain access. Invariably, their precise use will depend on will depend on the nature of your research design (Chapter 5), data-collection methods (Chapters 8–11) and the characteristics of the access you require (Figure 6.1). Their use will therefore need to take into account the type, nature and level of access required. Table 6.1 summarises a range of access strategies and their relative focus on these characteristics.

Ensuring familiarity before making contact

Before attempting to gain physical access, it is essential that you familiarise yourself fully with the characteristics of the organisation, group or, for elite interviews, the individual. The knowledge you gain will enable you to signal to the gatekeeper that you have thought

Table 6.1 Access strategies and their focus on different characteristics

| Strategy | Access characteristic | | |
|--|-----------------------|--------|-------|
| | Type | Nature | Level |
| Ensuring familiarity before making contact | ✓ | ✓✓ | ✓ |
| Allowing sufficient time | ✓✓ | ✓✓ | ✓✓ |
| Using existing contacts and developing new ones | ✓ | ✓✓ | ✓ |
| Providing a clear account of the purpose of the research and type and level of access required | ✓ | ✓ | ✓✓ |
| Overcoming concerns | ✓ | ✓ | ✓✓ |
| Identifying possible benefits of participating | ✓ | ✓✓ | ✓ |
| Using suitable language | ✓✓ | ✓✓ | ✓✓ |
| Facilitating replies | ✓✓ | ✓✓ | ✓✓ |
| Developing access incrementally | ✓ | ✓ | ✓✓ |
| Establishing researcher credibility | ✓ | ✓ | ✓✓ |

✓✓ = more focussed, ✓✓ = less focussed

carefully about your research, as you will be able to provide a credible case to justify your request to grant access.

Allowing sufficient time

Physical access may take weeks or even months to arrange, and often the time invested will not result in access being granted (Buchanan et al. 2013). An approach to an organisation or group will result in either a reply or no response at all. If you do not receive a reply and still wish to gain access, you will need to allow sufficient time before making a follow-up request. Seeking access to a large organisation as an external researcher, where you have no contacts, will necessitate considerable detective work to establish the most appropriate person to email or telephone. When making contact, care needs to be taken in composing the email or planning the phone call.

Gaining physical access to people from a large number of organisations offers additional challenges. Where data will be collected using questionnaires, researchers usually either purchase a list of potential respondents or, alternatively, select them from a volunteer panel (Saunders et al. 2017). Access in such cases is dependent upon the willingness of potential respondents to take part, the accuracy of the third-party list and careful planning and monitoring (Box 6.7).

If you can contact a participant directly, such as a manager, an exchange of emails or a telephone call or online chat may be sufficient to gain access. Here you should be clear regarding what you require and the value of your work and your credibility. Even so, you will still need to allow time for your request to be received and considered and an interview meeting to be arranged at a convenient time for your research participant. This may take a number of weeks, and you may have to wait even longer to schedule an interview.

An access request may be passed ‘up’ the organisation or group for approval. Where you are able to use a known contact in the organisation or group this may help, especially where a senior person is willing to act as a sponsor for your research. Even so, you will still need to allow for this process to take weeks rather than days. It is likely that you will be asked to attend a meeting to discuss your research during which your case for access will be evaluated. This all takes time!

Where your intended participants or respondents are not the same people who grant you physical access, you will need to allow further time to gain their acceptance. This



Box 6.7 Checklist

To maximise physical access using third-party lists

- ✓ Have you checked third-party compiled list or volunteer panel for accuracy, even if purchased from a reputable source?
 - ✓ Have you used pilot testing to establish the likely response rate and likely representativeness?
- ✓ Have you logged actual and complete returns regularly against sample requirements so that it becomes clear at an early stage if response targets or representativeness are unlikely to be met?
 - ✓ Have you followed up non-respondents and organisations that helped in distributing the questionnaire with polite but regular reminders to maximise returns?
 - ✓ Do you have a contingency plan to activate if response rates are lower than expected?

Source: Developed from Saunders et al. (2017).

may involve you making **pre-survey contact** by telephoning these people (Section 11.8), engaging in correspondence or holding an explanatory meeting with them. We have found establishing contact and securing cooperation can take weeks rather than days, especially individuals' availability is restricted.

Once you have gained physical access to the organisation or group you will need to gain cognitive access. Whichever data collection procedures you use, time needs to be allowed for maximising cognitive access. For some forms of observation and interviews (Chapters 9 and 10), access will be developed while collecting the data. Where continuing access is required, time will be needed to negotiate, or renegotiate, access at each stage.

Using existing contacts and developing new ones

Most management and organisational researchers suggest that you are more likely to gain access where you are able to use **existing contacts** such as colleagues, friends or others already known to you. Buchanan et al. (2013: 56) note they 'have been most successful where we have a friend, relative or student working in the organisation'. We have also found this to be the case. In order to request access we have approached colleagues, present or past students, course advisors, LinkedIn connections or those who are otherwise known to us through our networks. Their knowledge of us means they can trust our stated intentions and assurances we give about the use of any data provided. Using such existing contacts means we can establish a track record of our credibility that we refer to when approaching other organisations or groups where we do not have contacts.

Use of contacts will depend largely on the strategy (Section 5.5), sample selection (Sections 7.2 and 7.3), research question and objectives. It is likely to be easier to use known contacts in an in-depth study that focuses on a small, purposively selected sample, such as a case study strategy; or for a survey strategy in one organisation. However, use of known contacts may also be possible where you have a large number of appropriate connections through your professional and online networks. There will clearly be some convenience in terms of gaining access through contacts that are familiar; however, these contacts may also be cases in other forms of non-probability samples (Section 7.3).

It may be possible for you to use a previous employer or your work placement organisation for your research project. You will undoubtedly have made a number of contacts who may be able to help negotiate or even grant access. You may have become interested in a particular topic because of the time that you spent in the organisation. Prior to asking for access you need to identify a research question and objectives, and plan your research project. The combination of genuine interest in the topic and relatively easy access to organisational participants should help you undertake a good-quality and useful project.

Where you need to develop **new contacts**, you will need to approach people previously unknown to you. You may consider asking the local branch of an appropriate professional association for the names and contact details of key employees to contact in organisations where it would be suitable for you to conduct research. You could also contact this professional association at national level, or an employers' association for a particular industry, or a trade union, at local or national level. Alternatively, you might contact one or more chambers of commerce, skills training organisation or other business networks. Note, however, such associations and organisations receive hundreds of requests from students each year and so may have insufficient time or resources to respond.

You may also consider making a direct approach to an organisation or group in an attempt to identify the appropriate person to contact in relation to a particular research project. This has the advantage of potentially providing access to organisations or groups that you would like to include in your research project; however, great care needs to

be exercised at each stage of the process (Box 6.8). Once you have obtained the email addresses of possible contacts you will need to send a request to each of them (Box 6.9), following the standards of care used for a formal letter. Use of email may result in some recipients simply deleting the message! Sending a letter to a potential gatekeeper can result in your request for access being considered more carefully.

Providing a clear account of the purpose and type of access required

Asking for access and cooperation without specifying your requirements clearly is likely to lead to either no response or refusal. It is also likely to be considered unethical (Section 6.6). A potential organisation, group or participant needs to know enough about your research to make an informed decision about whether to take part.

Establishing your credibility is vital to gaining access. Using existing contacts means you can trade on your existing level of credibility. However, when you are making contact for the first time, your approach will be highly influential. Your request will need to be well presented and demonstrate your clarity of thought and purpose. Poor preparation will be immediately apparent and is likely to result in access being refused (Section 10.4).



Box 6.8 Focus on student research

Identifying possible contacts and requesting access

Andrew identified a number of organisations that matched the criteria established for businesses he wished to include in his research project.

Organisations' websites were used to identify the corporate headquarters, which he then contacted by telephone. When talking to each organisation, Andrew explained he was a student and gave the title of his course and the name of his university. He also gave a very brief explanation of his research to the person who answered the telephone. This resulted in him being provided with a telephone number or email address or being connected directly to the part of the organisation thought appropriate by the person answering the telephone. Andrew always ended this initial telephone conversation by thanking the person for their help.

At the next stage, Andrew again revealed he was a student and gave the title of his course and the name

of his university. The purpose of the research was explained briefly to the personal assistant who inevitably answered the telephone. Andrew asked for the name and email address of the person whom the personal assistant thought would be the most appropriate person to contact. In most cases the people to whom he spoke at this stage were helpful and provided some excellent leads.

Sometimes, particularly in large organisations, Andrew found that he was not talking to someone in the appropriate part of the organisation. He therefore asked the person to help by transferring the telephone call. Sometimes this led to a series of calls to identify the right person. Andrew always remained polite, thanking the person to whom he spoke. He always gave his name and that of his university and tried to create a positive for what could be perceived as a tiresome enquiry.

Andrew chose to ask for the name and email address of a hoped-for organisation 'lead'. He would send a written request, which could be considered when it was convenient, rather than attempting to talk to them, when they might have been busy. This resulted in many successes, and Andrew added a number of good contacts to his previous list.



Box 6.9

Focus on student research

Email requesting access

Annette was undertaking her research project on the use of lean production systems. Having made telephone contact with the production controller's personal assistant, she was asked to send an email requesting access.

Unfortunately, Annette relied on her email software's spellcheck to proofread her email. This resulted in the

production controller receiving an email containing four mistakes:

- the addition of the word 'I' at the end of the first paragraph;
 - the phrase 'between 30 minutes and half an hour' instead of 'between 30 minutes and an hour' at the end of the second paragraph;
 - two digits being transposed in the mobile telephone number at the end of the last paragraph, resulting in it being incorrect;
 - the second sentence of the final paragraph being poorly worded.

Not surprisingly, Annette was denied access.

Research Project: The Use of Lean Production Systems

Dear Mr Kolowski

Further to my telephone conversation with your personal assistant, Tom Penny, I would like to meet with you and discuss the use of lean production systems at Manufax PLC. The interview is part of a series I am arranging with a carefully selected sample of production managers for my degree in Business Management at the University of Anytown. I

An outline of my proposed interview structure is attached, although it is not my intention to follow it slavishly. I am hoping to conduct these interviews in January and February and envisage that they will last between 30 minutes and half an hour.

I am fully aware of the need to treat the data you give me with the utmost confidentiality. No source, individual or organisational, will be identified or comment attributed without written permission of the originator.

One of my intended outputs will be a report summarising the findings and I will be sending a copy of this to each of the participants in the study.

I hope that you are able to help me and would be extremely grateful if you could let me know by replying to this email. As discussed with Tom Penny, I can then contact them to arrange a suitable time and venue at your convenience. If you prefer to talk to me to agree a suitable time and venue, please telephone me on 07987-6543210. If you require further information, please do not hesitate to get in touch.

Yours sincerely,

Your initial request for access should outline the purpose of your research, how the person contacted could help and what is involved in participating. The success of the request will be helped by being clear, concise and, if writing, the use of short and clear sentences. You should be polite and formal, and try to stimulate interest in your research. You will need to think carefully about whether to start with a telephone request, use email or send a letter by post. Even where your initial contact is by telephone, it is still advisable to confirm the details of your proposed research and requirements in writing (Box 6.9).

Overcoming concerns

Concerns about granting access relate principally to the time and resources that will be needed, the sensitivity of the research and issues of confidentiality and anonymity. Your request for access is more likely to be accepted if the amount of time and

resources you ask for are kept to a minimum. However, you need to be realistic about what is required. Stating a questionnaire will only take 5 minutes to complete, when it actually takes 15 minutes is very likely to annoy your respondents and result in only a few being returned.

Organisations, groups and individuals are less likely to cooperate where the research topic is sensitive or has negative implications. Organisations do not normally wish to present themselves as performing poorly. However, you may be able to highlight a positive approach by, for example, emphasising that your work will be designed to identify individual and organisational learning in relation to the topic (a positive inference). You should avoid sending any request that appears to concentrate on aspects associated with poor performance, failure or are likely to be a sensitive topic for your intended participants.

Where you have promised **confidentiality**, you must not reveal the identity of participants or participating organisations, groups or present your findings in a way that allows them to be identified. This means you should not share your raw data with anyone, not even your project tutor. **Anonymity** means your data is in a form that does not identify individuals and it will not be possible for identification to occur, even by the researcher. This allows for much wider use of the data as data protection legislation often does not apply to such data. An alternative to anonymising data is **pseudonymisation** where individuals are identified only by pseudonym such as a unique code or reference number. This is useful, particularly in cohort studies, as it allows data collected at different times about the same organisation, group or person to be linked by the pseudonym.

Although guarantees are usually given in writing at the time of making the initial request for access, you will need to repeat these assurances to individual participants as you seek their consent (Section 6.7). These will have to be acted upon, and confidentiality and anonymity maintained, when you write your project report (Box 6.18).

Identifying possible benefits

Decisions to participate can be looked at using the lens of **social exchange theory**, the potential participant evaluating the benefits and costs of taking part. Where potential benefits are judged to outweigh costs, a potential participant is more likely to take part. This notion of exchange does, however, have consequences. You therefore need to be clear about the potential benefits (and costs) of taking part.

Practitioners often wrestle with the same subjects as researchers and welcome the opportunity to discuss an issue, in a non-threatening, non-judgemental environment. A discussion may allow them to think through an issue and to reflect on their actions. For those who work in organisations, this may be the first time they have had this opportunity. Your research may offer some useful insights, although this does not mean that you should attempt to 'buy' your way in based on some promise about the potential value of your work.

It may help to offer a summary of your findings to those who grant access. The intention here would be to provide something of value and to fulfil any expectations about the (social) exchange. We believe it is essential that this summary is designed specifically for those who granted access rather than a copy of the research project you submit to your university. It is also possible that feedback from the organisation about your summary may help you further with your research.

Where access is granted in return for supplying a summary of your findings, it is important to be clear about what you will be providing. You therefore need to agree the format,

length, overall content and how you will ensure confidentiality and anonymity. This can vary from a one or two-page summary of key findings to a much more in-depth report.

Using suitable language

Some researchers advise against using certain research terms when requesting access because these may be perceived as threatening or not interesting to the potential participant. Buchanan et al. (2013: 57) suggest using the phrase ‘learn from your experience’ in place of research, ‘conversation’ instead of interview and ‘write an account’ rather than publish.

Your language will depend largely on the nature of the people you are contacting. It should be appropriate to the person being contacted, without any hint of being patronising, threatening or just boring. Given the vital role of initial telephone conversations, introductory emails or letters, we would suggest allowing adequate time to consider and draft these and using someone to check through your message (Section 11.7, Box 11.15). Do not forget that you need to leverage salient aspects to engender interest in your research project.

Facilitating replies

We have found that the inclusion of a number of different contact methods (telephone, mobile phone, email) in our written requests for access helps to ensure a reply. These may not be suitable in all cases depending on the characteristics of the access required. Inclusion of a stamped or postage pre-paid (freepost) addressed envelope may also facilitate a reply.

Developing access incrementally

There are potentially a number of advantages related to developing access incrementally. A request to an organisation for wide-ranging access may be sufficient to cause them to decline entry. Using an incremental strategy gains you initial access to a certain level of data such as to deliver and collect a questionnaire. It also allows you the opportunity to develop a positive relationship with those who are prepared to grant initial access of a restricted nature. As you establish your credibility, you can develop the possibility of achieving a fuller level of access, such as being allowed to facilitate a series of focus groups. A further advantage may follow from the opportunity that you have to tailor your request for further access in relation to your initial research findings. However, an incremental process is time-consuming and can be argued to be unethical as you do not explain your access requirements fully.

Establishing researcher credibility

In Section 6.2 we differentiated between physical and cognitive access. Even if you have been granted physical access to an organisation, group or individuals you cannot assume that those from whom you wish to collect data will cooperate. Gaining their cooperation will mean repeating much of the process that you will have used to gain entry into the organisation. You will need to explain the purpose of your research project, state how you believe that they will be able to help your study and provide assurances about

confidentiality and/or anonymity. This may involve emailing or writing to your intended participants or talking to them individually or in a group. How you do this depends on whether you are conducting the research face-to-face, the number of potential participants and your intended data collection techniques. However, your credibility and the probability of individuals' participation are likely to be enhanced if the request for participation is made jointly with a senior person from the organisation (Box 6.10). If your data-collection technique is likely to be considered intrusive, for example overtly observing people undertaking work tasks, you may need to exercise even greater care and take longer to gain acceptance. The extent to which you succeed in gaining cognitive access will invariably depend on your efforts.

The strategies that we have outlined to help you to gain access to organisations and to those whom you wish to participate in your research project are summarised as a checklist in Box 6.10.

Box 6.10 Checklist

To help to gain access

- ✓ Have you allowed yourself plenty of time for the entire process?
- ✓ Are you clear about the purpose of your research project?
- ✓ Are you clear about your requirements when requesting access (at least your initial requirements)?
- ✓ Can you use existing contacts, at least at the start of your research project, in order to gain access?
- ✓ (If you have been employed or on a work placement) Is the organisation or group an appropriate setting for your research project?
- ✓ Have you approached appropriate local and/or national employers, or employees, professional or trade bodies to see if they can suggest contacts through whom you might gain access?
- ✓ Have you considered making a direct approach to an organisation to identify the most appropriate person to contact for access?
- ✓ Have you identified the most appropriate person and been willing to keep on trying to make contact?
- ✓ Have you drafted a list of the points you wish to make, including your thanks to those to whom you speak?
- ✓ Have you considered how you will address likely organisational concerns such as:
 - the amount of time or resources that would be involved on the part of the organisation;
 - the sensitivity of your research topic;
 - the need for confidentiality and anonymity?

6.6 Research ethics and acting ethically

Defining research ethics

Acting ethically throughout the research process is essential. **Research ethics** are the standards of researcher behaviour that guide your conduct in relation to the rights of the subjects of your research and those who are affected by it. The appropriateness or acceptability of your research conduct will be influenced by broader social norms of behaviour. **Social norms** indicate the type of behaviour you should adopt in a particular situation; however, in reality, these allow for a range of behaviours leading to competing and, sometimes, conflicting positions.

Two dominant and conflicting philosophical ethical positions have been identified: deontological and teleological. The **deontological view** is based on the researcher following rules to guide their conduct. According to this view, acting outside the rules can never be justified. Where the rules are inadequate or contested, it is necessary to reappraise and, if required, amend them. In contrast, the **teleological view** is based on conduct being determined by its consequences, not a set of predetermined rules. This means deciding whether the benefits of undertaking an act outweigh its negative consequences. However, it is unlikely that just comparing the benefits to one group with the costs to another will provide you with a clear answer to such an ethical dilemma.

Attempts to overcome ethical dilemmas arising from different social norms and conflicting philosophical approaches have led to the widespread development of **codes of ethics**. These generally contain a list of principles outlining the nature of ethical research and an accompanying statement of ethical standards to guide your research conduct. As a member of a university (and where appropriate a professional association), you will be required to abide by such an ethical code or adhere to its ethical guidelines for research. Codes of ethics (Table 6.2) explicitly or implicitly recognise that ethical dilemmas exist, and it will often be necessary to exercise some choice about conduct. For example, the Statement of Ethical Practice produced by British Sociological Association (2017: 2) recognises expressly it is not possible to produce ‘a set of recipes’ to deal with all ethical dilemmas and that researchers need to exercise choice based on ethical principles and standards. Fortunately, such ethical principles and standards, mean researchers and ethical reviewers have an ethical basis against which to anticipate issues and risk, and exercise choice to avoid conflict and harm.

Ethical review

The conduct of your research will be guided by your university’s or professional association’s code of ethics or ethical guidelines, highlighting what is and what is not considered ethical. This must be followed to ensure that you do not transgress the behavioural norms and ensure your research is acceptable. However, as Bell and Bryman (2007) point out, such codes tend to be written in abstract terms and are designed to prevent misconduct. Table 6.2 provides Internet addresses for a selection of codes of ethics and ethical guidelines.

You should expect to submit your research proposal for ethical review. Like all students, your research will need to comply with your university’s code of ethics or ethical guidelines and the principles and standards that it contains. Ethical review may be conducted by your project tutor or other academic staff. Alternatively, you may be asked to complete and sign an ethical review form. Such ‘light touch’ or ‘fast track’ review, overseen by your school or faculty ethics committee, allows non-controversial research

Chapter 6 Negotiating access and research ethics

Table 6.2 Ethical codes, guidelines and statements of research practice

| Name | Internet address |
|--|---|
| Academy of Management's Code of Ethics | http://aom.org |
| Academy of Social Sciences' Five Ethical Principles for Social Science Research | https://www.acss.org.uk |
| All European Academies (ALLEA) The European Code of Conduct for Research Integrity | https://www.allea.org |
| American Psychological Association's Ethical Principles of Psychologists and Code of Conduct | http://www.apa.org |
| British Academy of Management's Code of Ethics and Best Practice | https://silo.tips |
| British Psychological Society's Code of Ethics and Conduct | https://www.bps.org.uk |
| British Psychological Society's Code of Human Research Ethics | https://www.bps.org.uk |
| British Sociological Association's Statement of Ethical Practice | https://www.britisoc.co.uk |
| Chartered Association of Business Schools' Ethics Guide Advice and Guidance | https://charteredabs.org |
| Economic and Social Research Council's (ESRC) Framework for Research Ethics (FRE) | https://esrc.ukri.org |
| European Union's Respect Code of Practice for Socio-Economic Research (The Respect Project) | http://www.respectproject.org |
| Market Research Society's Code of Conduct | https://www.mrs.org.uk |
| Researcher Development Initiative's Research Ethics Guidebook | http://www.ethicsguidebook.ac.uk |
| Social Research Association's Research Ethics Guidelines | https://the-sra.org.uk |
| UK Data Archive Managing and Sharing Data | http://www.data-archive.ac.uk |
| UK Department for Innovation, Universities and Skills' Universal Ethical Code for Scientists | https://www.gov.uk |
| UK Research and Innovation (UKRI) Policy and Guidelines on Preventing harm in research | https://www.ukri.org |
| UK Research Integrity Office's Code of Practice for Research | http://ukrio.org |
| Universities UK's The Concordat to Support Research Integrity | https://www.universitiesuk.ac.uk |

Please note these links are to the homepages of these organisations, you will need to search the website to find the appropriate document.

proposals posing minimal risk to participants and others to be considered without too much delay. A full ethical review conducted by your school or faculty ethics committee will be required where proposals raise ethical concerns or are considered to have higher levels of risk. You will need to be aware of potential ethical concerns and risks to those involved as you design your research and seek to avoid them. You should not assume using particular procedures will, on their own, reduce the possibility of ethical concerns or risk. While observation or interviews may appear to be more intrusive than a questionnaire all methods can raise ethical concerns. The questions you wish to ask and your intended participants (for example children) may also raise ethical concerns rather than the research method that you intend to use.

Research ethics committees are responsible for all aspects of ethical review and approval. They may have a proactive role in developing an ethical code, alongside disseminating advice about conducting research ethically. However, their primary role will be to review all research conducted by those in the institution, in particular that involving human participants and the collection and storing of personal data. They will examine aspects of research quality that relate to ethics; protecting the rights, dignity and welfare of those who participate as well as others who may be affected by it; and consider the safety of researchers. It is likely to be composed of experienced researchers from a variety of backgrounds, who will draw on their range of experience and knowledge of different ethical perspectives to provide advice. It will be expected to act in an impartial and independent way and its independence is likely to be supported by the inclusion of at least one external member, who otherwise has no connection to the institution.

For some research projects you will also need to satisfy the requirements of a host organisation's ethics committee established as well as your university. This may apply where your research is based in the health service. Many of our students undertaking research within the UK's National Health Service (NHS) have had to meet the requirements of their local NHS Trust's ethics committee (Box 6.11). Such requirements are often very time-consuming to meet.

Consideration of ethical issues is crucial throughout your research project and even beyond it. In Section 6.7 we explore ethical issues that arise at specific stages in the research process. Next consider general ethical principles that form the focus of codes of ethical conduct, followed by general issues associated with Internet-mediated research.

Ethical principles

Codes of ethics are intended to avoid poor practice, malpractice and harm (**non-maleficence**) and promote ethical practice, integrity, respect and fairness, and private or public good (**beneficence**). Rather than write highly detailed and prescriptive regulations to anticipate and deal with ethical issues for each research approach, codes of ethics instead contain a set of principles for researchers to apply to their own research and to that of others (Table 6.2).

To avoid harm, or at the very least to minimise it, it is necessary to evaluate risk in terms of the likelihood of harm occurring and the extent or severity of the harm. Harm may take a number of forms and lead to a range of consequences (Table 6.3). Estimating risk is not straightforward and it may be affected by a number of contextual or cultural factors. However, it is important to anticipate risk to minimise the likelihood of causing harm. Box 6.12 provides a checklist for assessing risk in research.



Box 6.11 **Focus on student research**

Establishing whether research warrants ethical review

Rachel worked for a local hospital. At her first meeting with her project tutor, he had reminded her to check whether she would need to submit her research project to the hospital's research ethics committee (REC) for review. Subsequently, she discussed this with her line manager who suggested she use the UK NHS Health Research Authority's Decision Tool, available online to address this question (Health Research Authority 2020).

The initial web page of this tool told her that its aim is to help users decide whether their 'study is research as defined by the UK Policy Framework for Health and Social Care Research' (Health Research Authority 2020: 1). This tool also allowed Rachel to click through to a web page that provides clear definitions of what is meant by the terms 'research', 'service evaluation', 'clinical audit' and 'usual practice' in the context of the NHS.

Research is 'the attempt to derive generalisable or transferable new knowledge to answer questions with scientifically sound methods including studies that aim to generate hypotheses as well as studies that aim to test them, in addition to simply descriptive studies'.

Service evaluation is 'designed and conducted solely to define or judge current care'.

Clinical audit is 'designed and conducted to produce information to inform delivery of best care'.

Usual practice is 'designed to investigate the health issues in a population in order to improve population health'

(Health Research Authority 2020)

Rachel used the decision-making tool by answering the questions on consecutive web pages. After clicking through these it became evident that her proposed project would be defined as a 'service evaluation'. The policy framework stated that service evaluation 'does not require REC review' (Health Research Authority 2020). After using the online tool she returned to discuss this decision with her line manager and later with her project tutor.

Table 6.3 Ethical principles and their rationale

| Ethical principle(s) | Rationale |
|---|--|
| Integrity, fairness and open-mindedness of the researcher | Research quality depends in part on the integrity, fairness and open-mindedness of the researcher. This means acting openly, being truthful and promoting accuracy. Conversely, it also means avoiding deception, dishonesty, misrepresentation (of data and findings, etc.), partiality, reckless commitments or disingenuous promises. Where appropriate, any conflict of interest or commercial association should be declared. |
| Respect for others | A researcher's position is based on trust and respect. The conduct of research entails social responsibility and obligations to those who participate in or are affected by it. Rights of all persons should be recognised and their dignity respected. |

| Ethical principle(s) | Rationale |
|--|--|
| Avoidance of harm (non-maleficence) | Any harm to participants must be avoided. This may occur to emotional well-being, mental or physical health, or social or group cohesion. It can include embarrassment, stress, discomfort, pain or conflict. Harm may be caused by being intrusive or zealous involving mental or social pressure causing anxiety or stress. It may also be caused by violating assurances of confidentiality and anonymity, or through harassment or discrimination. |
| Privacy of those taking part | Ensuring privacy underpins respect for others, avoidance of harm, voluntary nature of participation, informed consent, ensuring confidentiality and maintaining anonymity, responsibility in the analysis of data and reporting of findings, and compliance in the management and storage of data. |
| Voluntary nature of participation and right to withdraw | The right to not participate in a research project is unchallengeable and is accompanied by the right not to be harassed to participate. It is also unacceptable to attempt to extend the scope of participation beyond that freely given. Participants continue to exercise the right to determine how they will participate throughout the research including modifying the nature of their consent; withdrawing from participation and withdrawing any data they have provided. |
| Informed consent of those taking part | Informed consent necessitates researchers providing sufficient information and assurances about taking part to allow individuals to understand the implications of participation and to reach a fully informed, considered and freely given decision about participating, without pressure or coercion. Participants have the right to expect the researcher to abide by the consent given and prolong the duration of an interview or observation; or widen the scope of the research without first seeking and obtaining permission. |
| Ensuring confidentiality of data and maintenance of anonymity of those taking part | Individuals, organisations and groups have the right to remain anonymous and assurances about anonymity and confidentiality should be maintained. Data provided should be made non-attributable, unless there is an explicit agreement to attribute comments. Harm can result from unauthorised attribution or identification. |
| Responsibility in analysis of data and reporting of findings | Assurances of privacy, anonymity and confidentiality must be upheld when analysing data and reporting findings. Data must not be made up or altered and results should not be falsified. Analysis and findings should be reported fully and accurately, irrespective of whether they contradict expected outcomes. Sources of secondary data should be made clear. |
| Compliance in management of data | Research is likely to involve collection of personal data. Many governments have passed legislation to regulate processing, secure storage and possible sharing of personal data and there is likely to be a statutory requirement to comply in both the country in which you are studying and in which you are collecting data. |
| Ensuring safety of researchers | Researchers' safety is crucial when planning and conducting research. The Social Research Association's Code of Practice for the Safety of Social Researchers identifies possible risks from social interactions including 'risk of physical threat or abuse; risk of psychological trauma . . . ; risk of being in a compromising situation . . . ; increased exposure to risks of everyday life' (Social Research Association 2001: 1). |

Notes and sources: This table synthesises key points from many different approaches to writing ethical principles including those in Table 6.2. It does not provide comprehensive guidance. You will need to consult your university's code of ethics, alongside others appropriate to your research.



Box 6.12 Checklist

Assessing risk in research

- ✓ Is your proposed research likely to harm the well-being of those participating?
- ✓ Will others be harmed by the process or outcomes of your proposed research?
- ✓ How may this harm occur and what characteristics may make this more likely?
- ✓ How likely it is that harm might result?
- ✓ How severe would be any resulting harm?
- ✓ Which features or what aspects of your research may cause harm?
- ✓ How intrusive is your proposed research method or methods?
- ✓ How sensitive are your proposed questions, observations, searches or requests for data?
- ✓ Can you justify your choice of research method or methods and tactics; in particular, can you explain why alternatives that involve fewer potential risks cannot be used?
- ✓ Where anticipated risk cannot be reduced any further during the design of the research and ethical review is favourable, how will the implementation of your research seek to avoid the occurrence of risk in practice, or at the very least seek to minimise it?
- ✓ Does the information you provide to intended participants to facilitate informed consent also allow them to contact you to discuss potential concerns? How have you facilitated this while maintaining your own privacy (e.g. using a university email address, not your personal email or home address)?
- ✓ How will you commence a data collection activity to allow potential concerns to be raised first? How will you make yourself aware of themes that may be sensitive for particular participants?
- ✓ How will you reinforce the voluntary nature of participation to allow participants not to answer a particular question, set of questions or to decline any request for data?
- ✓ Other potential risks are likely to be evident within the context of your particular research project. What might these be and how will you manage them?

Research can benefit the researcher, participants, the group or organisation being researched and the community or society within which it occurs. It is important and ethical to be realistic about the benefits you claim for your research project and to honour any promises made about sharing findings, such as promising to send a summary to an organisation that provides access. Acting ethically is far more than just completing an ethical review form to get your research proposal approved! It means thinking about each aspect and each stage of your research from an ethical perspective and, on the basis of this, acting ethically throughout your research.

Ethical issues associated with Internet-mediated research

While the Internet helps facilitate access to some categories of participants and certain types of data, its use raises a number of issues and even dilemmas about the applicability of the ethical principles. Internet technologies change rapidly, and new ethical issues may well arise (British Psychological Society 2021). Not surprisingly, ethical guidance for Internet-mediated research therefore focuses upon general principles summarised in Table 6.3. In addition, guidance for ethical Internet-mediated research (Table 6.4) highlights specific issues and dilemmas raised by the use of online communications, special interest discussion forums and chat rooms, social networking sites, personal spaces and blogs, virtual worlds and online gaming spaces, commercial websites and databases and repositories. These include deception, lacking respect and causing harm, respecting privacy, nature of

Table 6.4 Internet addresses for ethical guidelines for the conduct of online research

| Name | Internet address |
|--|---|
| Association of Internet Researchers' (AOIR) Internet Research: Ethical Guidelines | https://aoir.org |
| Association of Internet Researchers Ess, C. and AoIR ethics working committee (2002) | http://aoir.org |
| Markham, A., Buchanan, E. and AoIR Ethics Working Committee (2012) Ethical Decision-Making and Internet Research | http://aoir.org |
| British Psychological Society Ethics Guidelines for Internet-mediated Research (2021) | https://www.bps.org.uk |

Please note these links are to the homepages of these organisations, you will need to search the website to find the appropriate document.

participation and scope to withdraw, informed consent, confidentiality and anonymity, analysis and reporting, data management, researcher safety and netiquette.

Deception

Researchers who join online communities with the intention of collecting data rather than participating or seeking consent (known as ‘passive analysis’ or ‘lurking’) are committing a form of deception. Declaring your real intention after a period of ‘lurking’ is seen by many online groups as unethical and will increase the chance of you being asked to leave.

Lacking respect and causing harm

Joining online communities and ‘harvesting’ data from them without their knowledge and permission is considered disrespectful and opposed to the principle of gaining trust. Such deception can cause distrust of you and other researchers and may cause damage to online communities and to their members.

Respecting privacy

Increasingly people have few, if any, expectations of privacy regarding materials they post to online forums that are accessible publicly without barriers. Yet, while it is possible to access online communities operating publicly accessible virtual space, content on these websites should be treated as private conversations, albeit ‘publicly private’ ones.

Nature of participation and scope to withdraw

Mass surveillance and potential (mis)use of personal data scandals have highlighted how potential misuse of personal data, including ‘harvesting’, can be perceived as violating the principles of voluntary participation. They have also drawn to potential participants attention difficulties associated with controlling how their personal data can be used or withdrawn (Box 6.13).

Informed consent

While most people know that their online posts and information can be read by members of the public, this does not mean that they are automatically granting their permission for these to be used as data by researchers (Kozinets 2020). Informed consent in a virtual setting can usually be obtained by contacting the online community’s moderator or



Box 6.13 Focus on research in the news

Online privacy: a fraught philosophical debate

Problems over the societal impact of Big Tech and social media, and their effect on values such as privacy, have long been a source of public concern.

But companies and campaigners find that resolving them involves navigating the ethics of privacy – a concept that countries, nationalities and traditions approach in very different ways.

‘Privacy is a nebulous concept generally and, over the years, we have seen misplaced narratives being framed around it,’ says Vidushi Marda, digital programme officer at Article 19, an international freedom of expression campaign group.

One example is the claim often made by defenders of mass surveillance that it is no threat to privacy if people have ‘nothing to hide’. ‘We see some of that changing,’ explains Marda, as policymakers develop a greater understanding of the importance of privacy. This was especially the case, she says, after whistleblower Edward Snowden revealed the scale of mass surveillance by US and other Western powers in 2013. Other developments that have shifted attitudes include the scandal over personal data and election influencing involving marketing group Cambridge Analytica; and an investigation by the UK’s information regulator into Clearview AI, the controversial facial-recognition system.

At the Big Tech companies, the dominant ideas about what online privacy means tend to be Western-centric – perhaps unsurprisingly, given so many are American.



Source: Extracts from ‘Online privacy a fraught philosophical debate’, Siddharth Venkataramakrishnan (2021) *Financial Times*, 17 May. Copyright © 2021 The Financial Times Limited.

administrator or, in the case of a web questionnaire or online interview, asking participants explicitly. It may include agreed limits about the scope of participation and procedures to allow concerns to be raised or for withdrawal to take place. Such procedures are important as lack of face-to-face contact makes it difficult to anticipate participants’ concerns and attitudes. Signed consent may be facilitated by emailing a consent form or series of questions in an online questionnaire (Box 11.16).

Confidentiality of data and anonymity of participants

Online community members discussions offer a permanent record. Even where this is a public forum and posts may have been made with the belief that no one will be ‘harvesting’ or analysing these materials. Crucially, online community members increasingly expect to be asked for consent before participating, including using their archived discussions. Doing so enables members to control data that is available to researchers (Franzke et al. 2020).

Analysis of data and reporting of findings

Issues of confidentiality, anonymity, privacy and copyright can occur when online data are collected, analysed and reported. Where data are ‘harvested’, the researcher is confronted with the dilemma of whether to use these data openly or anonymously. Although

obtaining ethical approval to harvest data without consent is increasingly unlikely, where this occurs, we recommend using pseudonyms and other changes to disguise the identities of those who created the material. Where you wish to quote from this material, there is the possibility that others could use Internet search engines to identify the author of a quotation. You therefore need to avoid using quotations that would be traceable without first obtaining consent (British Psychological Society 2021). In addition, blogs are protected by copyright laws and those who create them have exclusive rights in relation to their reproduction (Hookway 2008). Web pages and content on social network sites are also protected by copyright laws (British Psychological Society 2021) and, although often considered secondary data, it is important to be careful in using quotations and reference correctly (Appendix 1). Seeking informed consent should help to overcome the dilemmas associated with using materials from the Internet as both primary and secondary data (Section 8.1).

Data management

Data protection legislation has (or is likely to have, depending on country) implications for Internet-mediated research, including in the UK the need for notification and consent if personal data are to be processed and stored. Like all researchers, those using the Internet need to comply with current data protection legislation, as well as with any other legal requirements (Section 6.8). A further set of issues concerns the potential insecurity of data transmission and storage. This may be because of errors. For example, emails containing personal data may be sent to the wrong address. Questionnaire software may contain errors. Insecurity may also occur because others have access to a website and are able to alter data or to copy and direct it elsewhere. As researchers do not control websites or networks, risks associated with data transmission and storage need to be recognised and participants told about these in relation to confidentiality, anonymity and possible ‘data hacking’ or misuse as part of seeking informed consent.

Researcher safety

As a researcher you can help ensure your safety when conducting Internet-mediated research by using your university email address rather than your personal email. Similarly, you should not provide details of your home address. You also need to protect your own privacy when setting access rights to their own personal information on social media sites.

Netiquette

Netiquette refers to user standards to encourage courtesy, focussing principally on the use of email and messaging. Emails and messages may be poorly worded (Box 6.8), appearing unfriendly or unclear and therefore failing to interest those whom you approach. The ease of sending emails and messages may lead to ‘spamming’ potential and actual participants, sending large numbers of unwanted emails. Netiquette also involves respecting the intentions of other users, ensuring private messages are not subsequently be made public. We consider netiquette further in Sections 6.6, 10.10 and 11.8.

This review has highlighted several issues and dilemmas associated with the use of Internet-mediated research, although others will exist in practice. In addition, many aspects associated with these issues will need to be considered during the use of Internet-mediated research. The guidelines from the Association of Internet Researchers advocate, ‘a process [which] is *reflective and dialogical* as it begins with reflection on one’s own research practices and is continually discussed against the accumulated experience and ethical reflections of researchers in the field and ethical studies carried out’ (Franzke et al.

2020: 4). This approach is also recommended by Whiting and Pritchard (2018), who see ethical considerations as a process that need to be anticipated and revisited at each stage of research, discussed in Section 6.6.

6.7 Ethical issues at specific research stages

Ethical issues will be important throughout your research (Figure 6.2). Each stage will require ethical integrity from you in relation to your role as the researcher, any organisational gatekeeper(s) and, where appropriate, your research sponsor. Where you are undertaking research for an organisation, you will need to find the middle ground between the organisation's expectation of useful research and your right not to be coerced into researching a topic in which you are not interested or that does not satisfy your university's assessment requirements.

Research topic formulation and clarification

Ethical issues need to be anticipated and considered from the start of your research project. Each potential research topic will be associated with a number of possible ethical concerns, which will need to be anticipated as you research with integrity. As you focus more attention on your chosen topic, you will start to formulate your overarching research question and objectives that will give direction to your research design. While doing this you should consider potential ethical concerns relating to:

- the research purpose;
- the type(s) of data will you need to collect;
- the methods you will use to collect these data;
- the implications for those whom you will ask to participate;
- how the research might be used and with what consequences.

These and other possible ethical considerations will be important as you generate your research ideas and clarify the topic, designing your research.

Research design and data access

Most ethical issues can be anticipated and considered (at least initially) during the design stage of any research project. You should plan your research in accordance with your university's code of ethics, adapting your research strategy and choice of methods where appropriate. Evidence that ethical issues have been considered and evaluated at this stage is likely to be one of the criteria against which your research proposal is judged.

Gaining access

Ethical issues can arise when you seek access. As noted earlier, you should not attempt to apply any pressure on intended participants to grant access. This is unlikely to be the case where you are approaching a member of an organisation's management to request access. However, where you are undertaking a research project as an internal researcher within your employing organisation (Section 6.3), there may be a temptation to apply pressure on others (colleagues or subordinates) to cooperate. Individuals have a right to privacy and to not participate. By not respecting this, you may well be causing harm.

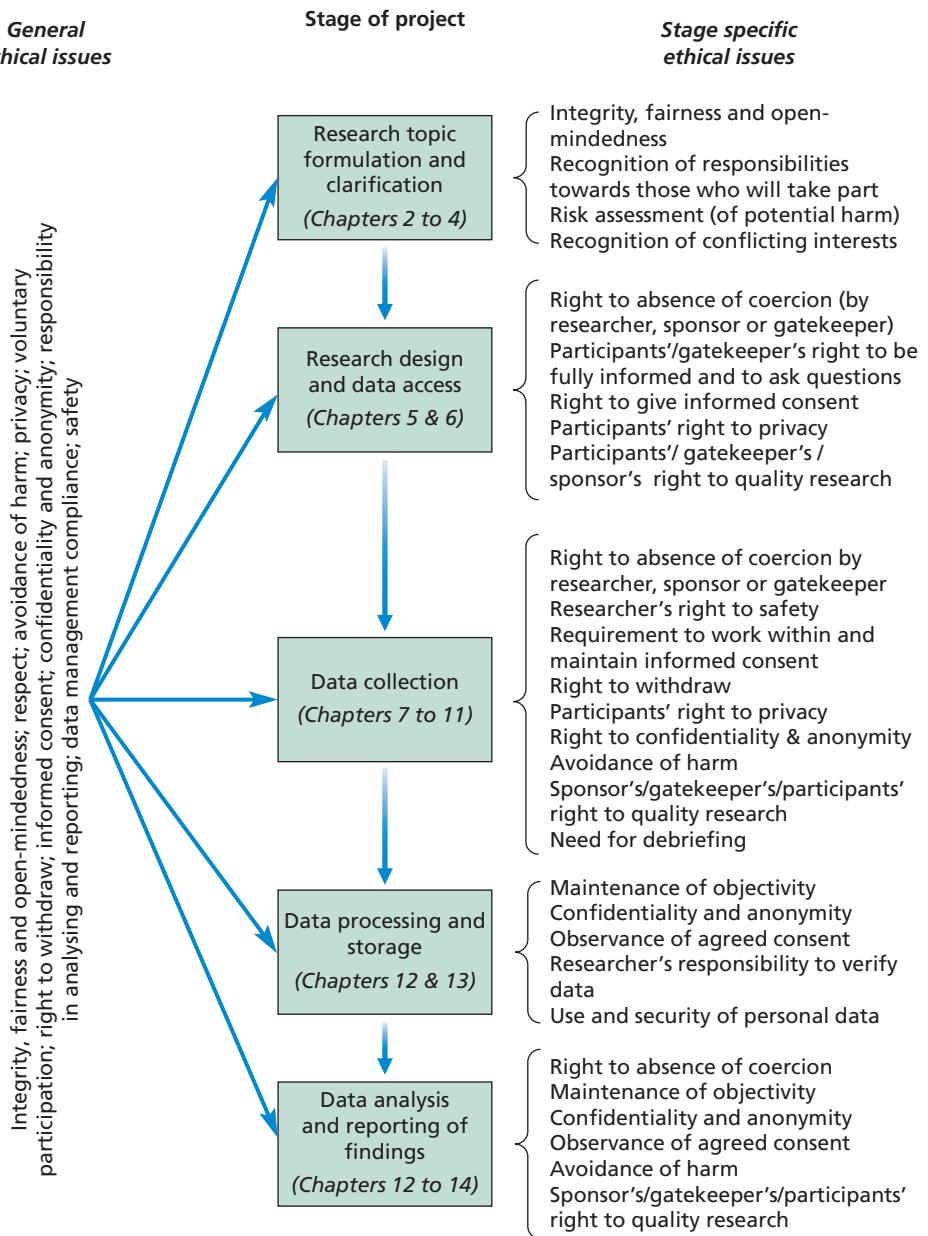


Figure 6.2 Ethical issues at specific research stages

Box 6.14 provides a checklist to help ensure you are not putting pressure on individuals to participate. You may also cause harm by the timing and nature of any approach that you make to intended participants – perhaps by telephoning at ‘unsociable’ times or by ‘confronting’ those from whom you would like to collect data. Access to secondary data may also raise ethical issues. Where you happen to obtain access to personal data about individuals who have not consented to let you have this (such as through personnel or client records), you will be obliged to anonymise these or to seek informed consent from those involved.



Box 6.14 Checklist

To help ensure you do not pressurise individuals to participate

- ✓ Have you ensured participants have not been coerced into participating?

- ✓ Are no inducements (e.g. financial payments), other than reimbursement for travel expenses and in some cases time, offered?
- ✓ Are any risks involved in participation likely to be acceptable to those participating?
- ✓ Are participants free to withdraw from the study at any time and have you informed them of this?

Gaining consent

Gaining participants' consent is not straightforward (Box 6.15). In general terms, a request to a potential participant or respondent to take part is an attempt to gain **consent**. However, this raises a question about the scope of consent given. Where someone agrees to participate in a particular data collection method, this does not necessarily imply consent about the way in which the data provided may be used. Assurances you provide about anonymity or confidentiality, and how the data will be used, can help; but even these may be inadequate to clarify precisely the nature of that consent.

This suggests a continuum that ranges from a lack of consent, involving some form of deception, through **inferred consent**, where taking part leads the researcher to presume data may be analysed, used, stored and reported without clarifying this with the participant, to informed consent (Figure 6.3). **Informed consent** involves ensuring those involved are given sufficient information (discussed next), the opportunity to ask questions and time to consider without any pressure or coercion, and to be able to reach a freely given decision about whether or not to take part (Table 6.3).

Research where people have consented to participate can still involve deception. This may be related to deceit over the real purpose of the research, some undeclared sponsorship or an association with another organisation that will use any data gained for commercial advantage. Such deception can cause embarrassment or harm to those supporting your request for access, as well as to yourself.

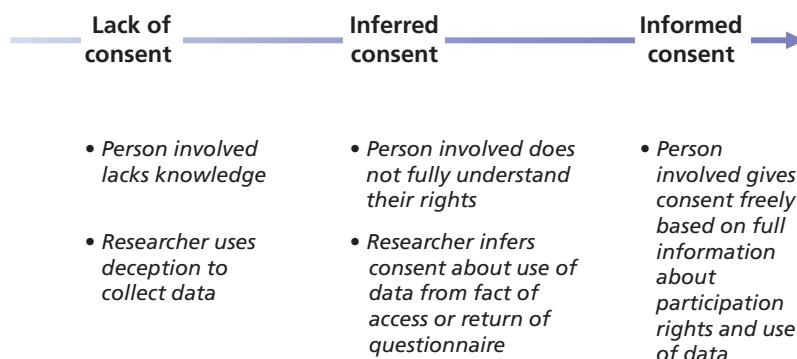


Figure 6.3 The nature of consent



Box 6.15 Focus on management research

Gaining informed consent

Rowlinson et al. (2016) undertook research that explored experiences of using 'payday lending' in the UK. This is a type of short-term but high-cost loan provided by private companies to people who are employed but who have low incomes, so that they can survive financially until their next payday, when the loan is due to be repaid.

Twenty-one in-depth, exploratory interviews were undertaken with participants who had taken loans from payday lending companies over the previous year.

These participants had originally completed a questionnaire administered by a specialist research company in town centres and shopping areas and were recruited to participate in the researchers' interviews.

The researchers state they, 'took ethical concerns seriously', including gaining informed consent (Rowlinson et al. 2016: 533). At the start of each interview they explained the nature of the research they were undertaking and how the data participants provided would be used. They assured participants data and reporting would be anonymised and provided them with a research information sheet.

Rawlinson and colleagues acknowledge their research was intrusive, reporting that each interviewer was trained and experienced in conducting interviews that intruded into sensitive issues. Researcher safety as well as data quality were given as reasons for conducting interviews in pairs.

Participant information sheets

The information that is required for prospective participants or respondents to make informed decisions about whether or not to participate should be provided as a **participant information sheet** or **information sheet**. This may be emailed or made available online to intended participants or respondents or, in the case of face-to-face research, given personally. It should include information about the nature of the research, the requirements and implications of taking part, participants' or respondents' rights, how their data will be analysed, reported and stored and whom to contact in the case of concerns (Box 6.16).



Box 6.16 Checklist

Requirements for a participant information sheet

Organisational gatekeepers and intended participants need to be informed about the following aspects of a research project. This can be drawn together in a research information sheet or participant information sheet.

About the nature of the research

- ✓ What is the title of the research project?
- ✓ What is the purpose of the research?

- ✓ Who is or will be undertaking it – the name(s) of the researcher(s)?
- ✓ Where does the research originate from – the name of the researcher's university or employing organisation?
- ✓ Is the research being funded or sponsored – if so, by whom and why?
- ✓ Who is being asked to participate – i.e. broad details about the sampling frame, sample selection method and size?
- ✓ How far has the research project progressed?

About the requirements of taking part

- ✓ What type of data will be required from those who agree to take part?



Box 6.16 Checklist (continued)

Requirements for a participant information sheet

- ✓ How will these data be collected (e.g. interview, observation or questionnaire)?
- ✓ How much time will be required and on how many occasions?
- ✓ When will data collection take place (dates)?

About the implications of taking part

- ✓ What assurances will be provided about anonymity and confidentiality?
- ✓ What will be the consequences of participating – possible risks, depending on the nature of the approach, and expected benefits?
- ✓ When will any expected benefit, such as the promise of a summary report of the findings, be made available?
- ✓ Depending on the nature of the research, when and how will any debriefing be conducted?

About the rights of those taking part

- ✓ Recognition that participation is voluntary.
- ✓ Recognition that those taking part have the right to decline to answer a question or

set of questions; or observed in particular circumstances.

- ✓ Recognition that those taking part have control over the recording of any of their responses where it is agreed that any type of photographic, video or voice recording may be made.
- ✓ Recognition that those taking part may withdraw at any time.

About the use of the data collected and the way in which it will be reported

- ✓ Who will have access to the data collected?
- ✓ How will the results of the research project be disseminated?
- ✓ How will assurances about anonymity and confidentiality be observed at this stage?
- ✓ What will happen to the data collected after the project is completed?
- ✓ Where data are to be destroyed, what is the date by which this will happen?
- ✓ Where data are to be preserved, where and how will these be stored securely, who might be given access to them and what safeguards will be established to ensure the continuing future confidentiality of these data and anonymity of those taking part?
- ✓ Whom to contact to raise any concerns and questions about the research, including their name, work address, email and contact telephone number.

The precise information required for informed consent will vary according to your research strategy, as will the way in which you seek to establish consent. If you are intending to use a questionnaire where personal data are not collected or where responses are entirely anonymised, its return completed is often taken to imply consent. Yet, as illustrated in Box 11.16, including a question in a questionnaire explicitly to request consent is straightforward. Either approach will require you to include an information sheet detailing how these data will be analysed and reported, for what purpose and what will then happen to them, as well as your identity (Corti et al. 2020). If you are intending to interview a senior manager, correspondence may be exchanged to establish informed consent (Section 6.5). When interviewing individuals, informed consent should be supplemented by a more detailed written agreement, such as a **consent form** (Box 6.17), which is signed by both parties. This helps to clarify the boundaries of consent and should help you to comply with your university's code of ethics as well as data protection legislation (Section 6.8, Corti et al. 2020). Depending on the nature of your research project you may need to seek consent to collect photographic or video-recorded data. As with audio-recording, consent needs to be obtained before the event and recorded on the consent form. Where you have established informed consent through prior written correspondence, it is worthwhile re-establishing

this with each intended participant immediately prior to collecting data (Box 10.9). This is particularly important where you are gaining access on an incremental basis (Section 6.5).

In Section 6.5 we discussed possible strategies to help you to gain access including highlighting possible benefits to an organisation of granting you access. You should be realistic about this and not offer more than is feasible. Offering to supply information arising from your work without intending to do this is unethical, and the effect of such action (or inaction) may result subsequently in a refusal to grant access to other researchers.

Data collection

Data collection is associated with a range of ethical issues (Figure 6.2). Some will apply however data are collected; others are related to particular data collection procedures. Finally, and of paramount importance, there are issues associated with ensuring your own safety while collecting your data.

Consent

After individuals or organisations have completed a consent form (Box 6.17), agreeing to take part in your research, they still maintain their right to withdraw or decline to take part in a particular aspect of your research. You should not ask them to participate in anything that will cause harm or intrude on their privacy. Once access has been granted, you should keep within the consent you agreed or renegotiate access. Not doing so is likely to cause upset and could result in further data collection being refused. There are, perhaps, some situations where deception may be accepted in relation to ‘covert’ research, which we discuss later in this subsection.

Avoiding falsification and fabrication

During the data-collection stage you need to make sure you collect your data accurately and fully, trying to minimise subjective selectivity in what you record. This relates to the validity and reliability, or credibility, transferability and dependability, of your work (Section 5.8). Without this, your ability to analyse and report your work accurately will be impaired. Obviously, **falsification** (distorting or misrepresenting) and **fabrication** (inventing) any data are totally unacceptable and unethical.

Confidentiality and anonymity

Confidentiality and anonymity may be important in gaining access to organisations and individuals (Section 6.5). Once such assurances have been given, they must be kept. Confidentiality is particularly important in relation to personal and sensitive personal data (see Section 6.8). Ways of ensuring anonymity are inevitably research-method specific. While the main concern is likely to be individuals or organisations being able to be identified, it is worth recognising that permission may be given for data to be attributed directly to them.

Anonymising quantitative data by aggregating or removing variables that may identify respondents is relatively straightforward. However, where qualitative data are being reported it may be less straightforward. New points of significance will emerge as the research progresses which you will wish to explore with others. Your key concern is to ensure that you do not cause harm. For example, within interviews, participants can often infer what earlier interviewees might have said from the questions being asked. This may lead to participants indirectly identifying who was responsible for making the point that you now wish to explore with them, with potential repercussions for the person. Where you wish to get others to discuss such an issue you can attempt to steer the discussion to see if they will raise it without in any way making clear that one of the other participants has already referred to it.



Box 6.17 **Focus on student research**

Consent form

Anna's research involved interviewing face-to-face a number of franchisees who had expanded their franchises to run multiple outlets, to understand the competences required to achieve this expansion successfully and how they had developed these. Prior to

commencing each interview, Anna gave each participant an information sheet that summarised her research project, including the possible benefits and disadvantages of taking part. After carefully explaining her research, the reasons why (with the participant's permission) she wished to audio-record or video the interview and emphasising that individuals were not obliged to participate unless they wished, Anna asked them if they wished to participate. Those who did were asked to complete and sign the following consent form:

| | | |
|---|--|------------|
|  Anytown Business School | | |
| CONSENT FORM | | |
| Title of research project: Successful franchise expansion | | |
| Name and position of researcher: Anna Verhoeven, Final year student, Anytown Business School, University of Anytown | | |
| <i>please initial box</i> | | |
| 1. I confirm that I have read and understand the information sheet for the above study and have had the opportunity to ask questions. | <input type="checkbox"/> | |
| 2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving reason. | <input type="checkbox"/> | |
| 3. I agree to take part in the study. | <input type="checkbox"/> | |
| <i>please tick box</i> | | |
| 4. I agree to the interview being audio recorded. | <input type="checkbox"/> Yes <input type="checkbox"/> No | |
| 5. I agree to the interview being video recorded. | <input type="checkbox"/> <input type="checkbox"/> | |
| 6. I agree to the use of anonymised quotes in publications. | <input type="checkbox"/> <input type="checkbox"/> | |
| Name of participant: | Date: | Signature: |
| Anna Verhoeven (researcher) | Date: | Signature: |

Use of the Internet and email during data collection can lead to the possibility of serious ethical and netiquette issues, related to confidentiality and anonymity. For example, it would be technically possible to forward a participant's email (or interview notes) to another participant in order to ask this second person to comment on the issues being raised. Such an action would infringe rights to confidentiality and anonymity, perhaps causing harm. It should definitely be avoided. It may also lead to data protection issues related to the use of personal data (Section 6.8).

Appropriate behaviour

The ability to explore and seek explanations through interview-based techniques offers opportunities for ethical and other issues to arise. Even with online interviews, personal interaction, scope to use non-standardised questions and, with video-conferencing software, to observe on a 'face-to-face' basis, mean you can exercise a greater level of control (Chapter 10) than with a quantitative approach based such as structured interviews or self-completed questionnaires (Chapter 11).

The relatively greater level of research control associated with interview-based techniques means you need to be careful your behaviour remains within appropriate and acceptable parameters. In face-to-face interviews, you should avoid overzealous questioning and pressing your participant for a response as this can be stressful. You should also make clear to your interview participants that they have the right to decline to respond to any question. The nature of questions to be asked also requires consideration. For example, you should avoid asking questions that are in any way demeaning to your participant (Sections 10.5–10.11). For all forms of synchronous (real-time) interviews it will be necessary to arrange a time that is convenient for your participant. For questionnaires (Chapter 11) participants may still be reluctant to answer questions of a personal or sensitive nature even where their responses are anonymous.

Reactivity and covert research

Use of observation techniques (Chapter 9) raises its own ethical concerns. The boundaries of what is permissible to observe need to be defined clearly. You should also avoid attempting to observe behaviour related to private life, such as personal telephone calls and so forth. Without this, the relationship between observer and observed will break down, with the latter finding the process to be an intrusion on their right to privacy. There is, however, a second problem related particularly to the use of this method, the issue of **reactivity**. This is the reaction on the part of those being investigated to the researcher and their research instruments (Bryman 1988: 112).

One solution to reactivity is a **covert** study, so those being observed are not aware. In such a situation, declaring your purpose at the outset of your work might lead to non-participation or to problems related to credibility and dependability (or validity and reliability) if those being observed altered their behaviour. The rationale for being deceitful in a benign way would be related to a question of whether 'the ends justify the means', provided other ethical aspects are considered (Wells 1994: 284). For others, deception should never be used. Fortunately, the problem of reactivity may decline as those being observed adapt to your presence as declared observer. Their adaptation is known as **habituation**.

Where access has been denied, you may consider you have no choice other than covert observation. We strongly advise against this. Covert observation, if discovered after access has been denied, will prove to be a considerable source of irritation. Many universities' ethical codes prohibit any form of research where access has been denied. In such situations, you will need to re-evaluate and revise your design.

If you are an internal or practitioner-researcher, you may consider initially adopting a covert approach. You may decide to interview subordinate colleagues, organise focus groups through your managerial status or observe interactions during meetings without declaring your research interest. However, we would urge considerable caution and to think carefully about the longer term impact. While this approach may be more likely to yield trustworthy data than declaring your real purpose and acting overtly, the impact on your work relationships may well be detrimental.

Irrespective of the reason why deception is used, it is widely accepted that after covert observation has taken place, those affected should be informed. This process is known as **debriefing** and is used to inform participants about the nature of the research, its outcomes and ascertain if there have been any adverse consequences from taking part; if so, to talk to the participant affected and arrange for assistance as required (British Psychological Society 2018). Debriefing also occurs after agreed participation in strategies such as a research experiment.

Personal safety

When thinking about avoiding harm, many researchers forget about themselves! The possibility of harm to you as the researcher is an important ethical issue that you should not ignore. You should not reveal personal information about yourself such as your home address or telephone number. Careful consideration needs to be given to a range of risk factors including the research purpose, location and timing of data collection and health and safety considerations. You need to consider risks to your safety and minimise these through strategies such as meeting participants in safe spaces, collecting data collection during the daytime and letting others know your arrangements, including where you will be.

In discussing researcher safety with our students, we have found the guidance sheets on working alone and dealing with aggression provided by the Suzy Lamplugh Trust (<http://www.suzylamplugh.org/>) extremely helpful. As the Trust's guidance sheets emphasise, you should never allow your working practices (research design and conduct) to put your own safety in danger. This, and 'Code of Practice for the Safety of Social Researchers' (Social Research Association 2001), offer a range of strategies to promote personal safety (Box 6.18).



Box 6.18 Checklist

Personal safety when collecting primary data

- ✓ Plan your meeting with a person in a busy public place or office where other people work nearby if at all possible.
- ✓ Consider carefully the location you are travelling to and your travel plans: what risks might you encounter; whether you will use public transport, a reputable taxi firm or a private car (if you use a private car ensure there is a safe place to leave it).
- ✓ Carry sufficient money to cover your expenses and any unexpected ones.
- ✓ Carry a mobile phone and make sure it is fully charged.
- ✓ Make a mental note of a safe way to leave the building or place where you meet.
- ✓ Make a telephone call to a friend before a particular meeting to tell them who you are meeting, where and how long you expect the meeting to last; call them again to tell them you have left and about your subsequent meeting plans and/or travel arrangements.
- ✓ Set up a system where you contact someone each day with a full list of whom you are meeting, where and at what times.
- ✓ In a meeting be aware of the use of body language, appearance, cultural norms, social distance and the gender dynamics of interactions.

- ✓ The considerable majority of meetings are helpful and non-threatening but in very rare cases someone may become aggressive or angry: be aware of any changes in behaviour; consider what questions you are asking and how you are asking them; remain calm; where necessary be assertive but not aggressive; if necessary end the meeting politely and leave quickly.
- ✓ Carry a personal (screech) alarm in case of an emergency.
- ✓ Consider carefully your safety if the location of your research means that you will be in a lone working situation; some researchers work in pairs in such situations to reduce safety risks.
- ✓ Always consider your safety and any risks to yourself and avoid any situation that might be difficult or dangerous.

Data processing and storage

Confidentiality and anonymity issues are crucial while processing and storing data. In particular, the use and security of personal data that either directly identifies individuals or makes them identifiable when used in combination must be managed in accordance with data protection legislation (Section 6.8). The ethical and legal management of data is often formalised in a **data management plan**. This outlines how your research data, both raw and processed, will be collected, organised, stored securely (including backup copies) and, where appropriate, shared. It also considers issues of ethics and legal compliance and, as a living document, should be kept up to date (Box 6.19).



Box 6.19 Checklist

Requirements for a data management plan

Data collection and documentation

- ✓ In overview, what data are being collected or sourced?
- ✓ How are these data being collected (e.g. 'observation', 'interview', 'questionnaire') or sourced and, for longitudinal research, over what time period?
- ✓ In what format are the data being stored (e.g. 'Word document', 'SPSS data file') and likely file size?
- ✓ How are details of data coding being noted (e.g. 'separate document')?

Data storage and backup

- ✓ How and where are the data being stored and password protected (e.g. 'university files server')?
- ✓ How are the data being backed up, how frequently and how are different versions be labelled (e.g. 'v1', 'v2')?

- ✓ How are non-digital data being stored securely (e.g. 'locked filing cabinet')?

Ethical and legal compliance

- ✓ How are the identities of participants being protected, where required (e.g. 'anonymised')?
- ✓ Does the processing and storing of personal data comply fully with data protection legislation?
- ✓ Who owns these data? (e.g. 'researcher')
- ✓ (If appropriate) Are there restrictions on the re-use of secondary data?

Data archiving

- ✓ What will happen to the data after the research project is completed (e.g. 'destroyed', 'stored on university files server')?
- ✓ How long will the data be kept after the research project (e.g. '10 years')?
- ✓ Who will have access to these data (e.g. 'researcher')?

Source: Developed from Digital Curation Centre (2013).

Files containing confidential or personal data need to be properly labelled and kept securely. This refers not only to your original notes or recordings but also to any subsequent drafts, transcriptions, re-recordings, backup and anonymised versions. For example, notes, recordings or questionnaires are likely to include personal identifiers such as names, job titles and workplace locations that identify the participant. Anonymised versions of data will have used pseudonyms, tactics such as aggregating data and higher levels of generalisation to remove personal identifiers. Nevertheless, if personal identifiers still exist, even in another document, they may be used to reveal the identities of participants or respondents. Particular care therefore needs to be exercised when storing both original versions of data that include personal identifiers and personal identifiers that provide the key to revealing the identities of anonymised individuals. Data that contain personal identifiers therefore need to be held securely and separately to anonymised versions of data to which they relate (Corti et al. 2020) to protect them from unauthorised access.

Non-digital data such as paper copies of interview or observation notes, signed consent forms, structured observation forms, questionnaires and other documents containing confidential or personal data need to be stored in a restricted, secure and safe place. Data held externally, such as on USB mass storage devices, also needs to be stored under the same conditions, encrypted and password protected. Although online file sharing and storage services enable you to keep an online copy of your data files, beware. Many universities explicitly prohibit the use of third-party cloud services.

When data are destroyed, paper documents should be shredded, not just placed in a bin, and digital materials permanently deleted (Corti et al. 2020). Ethical concerns of confidentiality and anonymity, participants' privacy and the need to ensure that harm is not caused to those who participated are likely to remain beyond the end of your research project.

Data analysis and reporting findings

Trying to maintain objectivity will be vital during the analysis stage to ensure you do not misrepresent the data collected. This will include not being selective about which data to report or, where appropriate, misrepresenting its statistical accuracy. A great deal of trust is placed in each researcher's integrity, and it would clearly be a major ethical issue were this to be open to question. Lack of honesty when analysing data and reporting findings will distort your conclusions and associated recommendations. Distorting or misrepresenting data, findings and conclusions are all examples of falsification, which is a totally unacceptable and unethical.

The ethical issues of confidentiality and anonymity also come to the fore during the reporting stage of your research. Wells (1994) recognises that it may be difficult to maintain the assurances that have been given. Allowing a participating organisation to be identified by those who can 'piece together' the characteristics that you reveal may result in embarrassment and access being refused to those who seek this subsequently. Great care therefore needs to be exercised to avoid this situation. You also have the option of requesting permission from the organisation to use their name. To gain this, you will almost certainly need to let them read your work to understand the context within which they would be named.

The same care needs to be exercised in maintaining individuals' anonymity. Embarrassment and harm might result from reporting data that are clearly attributable to an individual. Think carefully about collecting data that identify individuals, such as full names, unless necessary to do so. Always seek to anonymise the identities of those who take part by using a level of generalisation that ensures that others are not able to identify them.

For example, do not refer to specific ages, dates, locations, countries, real names, actual organisations' names or job positions or include photographs that will make it easy to identify participants or respondents, participating organisations, groups or communities (UK Data Archive 2017), unless there is express permission to identify any of these (Box 6.20).

A further ethical concern stems from the use made by others of your conclusions and any course of action that is explicitly referred to or implicitly suggested. Some argue it would be unethical to use the research data collected from a group of people to disadvantage. Others consider that while the identity of those taking part should not be revealed, they cannot be exempt from the way in which research findings are used subsequently to make decisions. This is clearly an ethical issue, requiring very careful evaluation.

Where you are aware that your findings could adversely affect the collective interests of those who took part, it would be ethical to refer to this possibility, even if it reduces the level of access you achieve. An alternative position is to develop your research question and objectives to avoid this possibility or so that decisions taken as a result of your research should have only positive consequences collectively for those who participate. You may find that this alternative is not open to you, perhaps because you are a part-time student in employment and your employing organisation directs your choice of research topic. In such situations it is more honest to explain that you are, in effect, acting as an internal consultant rather than a (dispassionate) researcher.

We now consider ethical issues that arise in relation to the analysis of secondary data derived from questionnaires. Where questionnaire data are subsequently used as secondary data, the original assurances provided to those who participated in the research can be inadvertently set aside, and participants may be disadvantaged through this use of data (Dale et al. 1988). Use of data for secondary purposes can therefore also lead to ethical concerns that need to be considered. Secondary analysis of qualitative data can also raise ethical concerns, although, as with quantitative data, these can be anticipated during data collection (Bishop and Kuula-Luumi 2017). In particular, they can be reduced by



Box 6.20 Focus on student research

Inadvertently revealing participants' identities

Over the years we have read a large number of student research projects. The following examples, drawn from some of these, highlight how easy it is to inadvertently reveal the identities of research participants in your research project:

- attributing a comment to a female accounts manager when there is only one such person;
- referring to a comment made by a member of the sales team, when only one salesperson would

have had access to the information referred to in the comment;

- reporting data and comments related to a small section of staff, where you state the name or job title of the one person interviewed from that section elsewhere in your research report;
- referring to an 'anonymous' organisation by name on the copy of the questionnaire placed in an appendix;
- attributing comments to named employees;
- thanking those who participated in the research by name;
- using pseudonyms where the initials of the pseudonym are the same as those of the actual person interviewed or where the name is similar, e.g. using Emilia Morris for Emily Morrison;
- including a photograph of the interview site or interviewee in your project report.



Box 6.21 Checklist

To help anticipate and deal with ethical issues

- ✓ Recognise potential ethical issues that will affect your proposed research.
 - ✓ Treat consideration of ethical issues as an active, continuous and reflexive process occurring throughout your research, from topic formulation to reporting, rather than just something you consider at the start of your project.
 - ✓ Use your university's code of research ethics to guide the topic choice, design and conduct of your research.
 - ✓ Anticipate potential ethical issues at topic formulation and clarification and anticipate how you would address these. Use this to help evaluate your choice of potential topics and decide which to research.
 - ✓ Anticipate ethical issues at the research design and data access stage and outline how you will seek to address these in your research proposal.
 - ✓ Seek informed consent being open and honest, rather than using deception.
 - ✓ Do not exaggerate the likely benefits of your research for those participating.
 - ✓ Respect others' rights to privacy at all stages of your research.
 - ✓ Maintain integrity and quality in data collection.
 - ✓ Recognise more intrusive research approaches have greater scope for ethical issues to arise and seek to avoid the particular problems related to interviews and observation.
- ✓ Avoid referring to data from a particular participant when talking to others, where this would allow the individual to be identified.
 - ✓ Be wary of researching covertly even where reactivity is likely to be a significant issue and a covert presence is practicable.
 - ✓ Maintain your objectivity during data analysis and reporting findings.
 - ✓ Maintain assurances of confidentiality and anonymity in reporting findings.
 - ✓ When using the Internet, recognise this may raise particular ethical issues and dilemmas. Anticipate these and determine how you will conduct your Internet-mediated research ethically.
 - ✓ Avoid using the Internet or email to share data with others taking part.
 - ✓ Consider how the collective interests of those involved may be affected adversely by the data collected and amend your research question and objectives where possible. Alternatively, declare this possibility to those whom you wish to participate in your proposed research.
 - ✓ If using secondary data, consider how you will protect the identities of those who contributed to its collection or who are named within it.
 - ✓ Unless necessary, base your research on genuinely anonymised data. Where it is necessary to process personal data, ensure that you comply fully with all current data protection legislation and your university's code of ethics.
 - ✓ Protect the identities of those involved, ensuring their anonymity in your project report unless you have their explicit permission to do otherwise.

anonymising data as they are collected and recorded, ensuring real names and organisations are not part of the data set.

A final checklist to help you anticipate and deal with ethical issues is given in Box 6.21.

6.8 Data protection principles

This section offers an overview of the principles of data protection, which are crucial to managing your data ethically and lawfully. We first consider the use and protection of personal data, and then the use of anonymised data.

Use and protection of personal data

Data protection legislation refers specifically to protecting **personal data**. These are data that either directly identify individuals by, for example, naming them, showing their image or their IP address; or make individuals identifiable when used in combination with other information. Personal data are therefore different to truly **anonymised data**, from which individuals cannot be identified.

Data protection in the European Union (EU) has assumed even greater importance since the implementation of the General Data Protection Regulation EU 2016/679 (GDPR). This Regulation repealed and replaced Directive 95/46/EC on 25 May 2018 (Box 6.19). As a Regulation of the European Parliament and European Council, it is directly applicable and legally binding in all EU member states. The EU GDPR provides protection for living individuals in relation to the processing of personal data applying in full to the UK, along with the UK's Data Protection Act 2018 and UK GDPR until 31 December 2020. On 1 January 2021, when the UK left the EU, the EU GDPR ceased to apply to the UK. However, it effectively became part of UK domestic law, the European Union (Withdrawal) Act 2019, amending UK GDPR so it is directly applicable to the UK on a standalone basis.

Both EU and UK GDPR set out clear rules regarding the processing and free movement of personal data and rules regarding to the free movement of personal data (Information Commissioner's Office 2018; Official Journal of the European Union 2016: L119/32 EN). Personal data are defined as data about a living individual that allow them to be identified, perhaps in combination with other information known to the controller of the data. These data include a person's name, identification number, location, online presence or some other attribute. Anything you do with these data is termed 'processing', whether automated or manual, including collecting, recording, organising and storing. The 'controller' is the person (or legal entity) who decides how and why to collect, record, organise and store the personal data; while the 'processor' processes these data on behalf of the controller.

Both GDPRs establish seven principles regarding personal data, that it must be:

- 1 processed lawfully, fairly and transparently;
- 2 collected for specified, explicit and lawful purposes and not processed further in a manner incompatible with those purposes, while allowing data to be processed further for scientific, historical and statistical research purposes where this is not incompatible with the initial purposes;
- 3 adequate, relevant and limited to the purpose for which they are processed;
- 4 accurate and, where necessary, kept up to date;
- 5 kept in a form that allows identification of data subjects for no longer than is necessary in relation to the purpose for which they are processed, while allowing personal data to be stored for longer periods where this is solely for scientific, historical and statistical research purposes and subject to measures to safeguard the rights and freedoms of data subjects;
- 6 kept securely and protected from wrongful processing and accidental loss or damage;
- 7 held responsibly by the controller in compliance with the points listed above.

The GDPRs emphasise these data can only be processed if it is necessary for a specific purpose and providing the data subject has consented and that the data 'controller' will be able to demonstrate that this has been given. The processing of sensitive and personal data is prohibited, unless one of a number of conditions applies, including explicit consent given by the data subject. Such personal data includes racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, genetic or biometric data, health and sex. Effective explicit consent is likely to mean clear and unambiguous written consent in this context.

EU and UK GDPRs deal with the rights of data subjects in relation to personal data collected from data subjects and obtained from other sources. Data subjects have rights to

access data held about them, to rectify errors, to be forgotten (the erasure of personal data) and to restrict processing under certain conditions. The GDPRs also regulate the roles of data controllers and processors; transfers of personal data to third countries or international organisations; mechanisms to supervise the implementation of this regulation; remedies, liabilities and penalties; and provision relating to specific processing situations. These include safeguards and derogations (exemptions) relating to scientific, historical and statistical research purposes that are designed to protect data subjects during the processing of personal data. Safeguarding measures include the use of pseudonyms where appropriate and other ways to process personal data that prevent the identification of data subjects.

Our brief summary of selected aspects of this legislation should only be treated as an introductory outline and not as providing any type of advice or guidance. Neither should this brief summary be interpreted as suggesting whether or not this or any other legislation is applicable to your research project. Where your research is covered by the scope of this or other legislation, you should seek advice that is appropriate to the particular circumstances of your research project where this involves the processing of personal data. Data protection legislation is likely to exist in countries outside the EU and UK, and you will need to be familiar with legislative requirements where you undertake your research project to understand how these may affect your research and the legal obligations that this places on you. Whether or not your research is affected by data protection legislation, you will also be aware of the need to conduct your research ethically and ensure you abide by your university's ethical code (Sections 6.6 and 6.7).

Use of anonymised data

The discussion of the legal protection of personal data has hopefully focused your mind on the implications of processing personal data. Unless there is a clear reason for processing these data, the best course of action is likely to be to ensure that your data are completely and genuinely anonymised and that any 'key' to identify data subjects is not retained by those who control these data. The EU and UK GDPRs do not apply to personal data that have been effectively anonymised so that data subjects are not, or no longer, identifiable. It therefore is not concerned with the processing of such anonymous information, including for research purposes.

There are various techniques to anonymise personal data. In relation to qualitative data these include removing data subjects' names and other personal identifiers from documents and records; using pseudonyms, especially in reporting; obscuring faces and other identifiers in visual images; blurring facial images and other identifiers in video recordings; and electronically altering voices in audio recordings. In relation to quantitative data these include data masking, where personal identifiers are removed; using pseudonyms, especially in reporting; data aggregation and reducing precision; and restricting upper and lower ranges to hide outliers (UK Data Service 2021).

6.9 Summary

- Access and ethics are critical aspects for the conduct of research.
- Access has three overlapping sets of characteristics:
 - type: traditional, Internet-mediated, intranet-mediated and hybrid access;
 - nature: single organisation, multiple organisation, individual person and elite person access;
 - level: physical, cognitive and continuing access.
- Suitability of access can be evaluated in terms of feasibility and sufficiency.
- Gaining access will depend to some extent on your status as either an external researcher or an internal or participant researcher.

- Internet-mediated access necessitates identifying an appropriate sample and negotiating virtual access. In doing this you should respect rights to privacy and copyright.
- There are a range of strategies that can be used in combination to support gaining access to organisations and intended participants or respondents dependent upon your research design.
- Research ethics refers to the standards of behaviour that guide your conduct in relation to the rights of those who become the subject of your research or are affected by it.
- There are two dominant but conflicting philosophical ethical positions: the deontological and the teleological.
- The conduct of your research will be guided by your university's or professional association's code of ethics or ethical guidelines that must be followed.
- Your research will almost certainly be subject to ethical review and you should not collect data until ethical approval has been given.
- Potential ethical issues should be recognised and considered from the outset of your research and are one of the criteria against which your research is judged. Issues may be anticipated by using codes of ethics, ethical guidelines and ethical principles.
- Ethical concerns can occur at all stages of your research project; during research topic formulation and clarification, research design and data access, data collection, data processing and storage, and data analysis and reporting of findings.
- Qualitative research is likely to lead to a greater range of ethical concerns in comparison with quantitative research, although all research methods have ethical issues associated with them.
- Gaining participants' or respondents' consent, although not always straightforward, is crucial. While covert research does not require consent, it is only likely to receive ethical approval in very specific cases.
- Assurances of confidentiality and anonymity are important in gaining access and, if given, must be kept.
- You need to consider your own safety very carefully when planning and conducting research.
- Further ethical and legal concerns are associated with data protection and data management, affecting the collection, processing, storage and use of personal and confidential data. You must comply carefully with data protection legislation when using personal data, to protect the privacy of their data subjects and to avoid the risk of any harm occurring.

Self-check questions

Help with these questions is available at the end of the chapter.

- 6.1** What are the characteristics of access and why is it important to differentiate between them?
- 6.2** What do you understand by the use of the terms 'feasibility' and 'sufficiency' when applied to the question of access?
- 6.3** Which strategies to help to gain access are likely to apply to the following scenarios:
 - a** an 'external' researcher seeking direct access to managers who will be the research participants;
 - b** an 'external' researcher seeking access through an organisational gatekeeper/broker to their intended participants or respondents;
 - c** an internal researcher planning to undertake a research project within their employing organisation?
- 6.4** What are the principal ethical issues you will need to consider irrespective of the particular research methods that you use?
- 6.5** What problems might you encounter in attempting to protect the interests of participating organisations and individuals despite the assurances that you provide?

Review and discussion questions

- 6.6** In relation to your proposed research project, evaluate your scope to use:
- a** traditional access;
 - b** Internet- or intranet-mediated access;
 - c** hybrid access
- to gain access to those you wish to take part. Make notes about the advantages and disadvantages of each access strategy.
- 6.7** With a friend, discuss your answers to Question 6.6. Now discuss how you intend to gain access to the data you need for your research project. In your discussion make a list of possible barriers to your gaining access and how these might be overcome. Make sure that the ways you consider for overcoming these barriers are ethical!
- 6.8** Agree with a friend to each obtain a copy of your university's or your own professional association's ethical code. Each of you should make a set of notes regarding those aspects in the ethical code that you feel are relevant to your own research proposal and a second set of notes of those aspects you feel are relevant to your friend's research proposal. Discuss your findings.
- 6.9** Visit the Suzy Lamplugh Trust website at <http://www.suzylamplugh.org> and the Social Research Association at <https://the-sra.org.uk/common/Uploaded%20files/SRA-safety-code-of-practice.pdf>. Browse the guidance leaflets/web pages and code of practice located at these websites. Make a list of the actions you should take to help ensure your own personal safety when undertaking your research project. Make sure you actually put these into practice.
- 6.10** Visit the Research Ethics Guidebook at www.ethicsguidebook.ac.uk and browse through the sections of this guide. In relation to the context of your proposed research project, make a note of points that provide additional guidance to help you to anticipate and deal with potential ethical concerns.



Progressing your research project

Negotiating access and addressing ethical issues

Consider the following aspects:

- Which types of data will you require in order to be able to answer your overarching research question and address your research objectives?
- Which research methods do you intend to use to obtain these data (including secondary data as appropriate)?
- What are the characteristics of the access you require in order to be able to collect data?

- What problems are you likely to encounter in gaining access?
- Which strategies to gain access will be useful to help you to overcome these problems?
- Depending on the type of access envisaged and your status as an external researcher or internal/practitioner researcher, draft requests for organisational access and/or requests to individuals for their cooperation along with associated information sheets.
- Describe the ethical issues that are likely to affect your proposed research project, including your own personal safety. Discuss how you will seek to overcome or control these. This should be undertaken in relation to the various stages of your research project.
- Note down your answers. Use the questions in Box 1.4 to guide your reflective diary entry.

References

- Althammer, S.E., Reis, D., van der Beck, S., Beck, S. and Michel, A. (2021) 'A mindfulness intervention promoting work-life balance: How segmentation preferences affects changes in detachment, well-being, and work-life balance', *Journal of Occupational and Organizational Psychology*, Vol. 94, pp. 282–308.
- Bell, E. and Bryman, A. (2007) 'The ethics of management research: An exploratory content analysis', *British Journal of Management*, Vol. 18, No. 1, pp. 63–77.
- Bishop, L. and Kuula-Luumi, A. (2017) 'Revisiting qualitative data reuse: A decade on', *Sage Open*, January–March, pp. 1–15.
- British Psychological Society (2018) *Code of Ethics and Conduct*. Leicester: British Psychological Society. Available at <https://www.bps.org.uk/sites/www.bps.org.uk/files/Policy/Policy%20-%20Files/BPS%20Code%20of%20Ethics%20and%20Conduct%20%28Updated%20July%202018%29.pdf> [Accessed 4 July 2021].
- British Psychological Society (2021) *Ethics Guidelines for Internet mediated Research*. Leicester: British Psychological Society. Available at <https://www.bps.org.uk/news-and-policy/ethics-guidelines-internet-mediated-research> [Accessed 20 July 2021].
- British Sociological Association (2017) *Statement of Ethical Practice*. Available at https://www.britsoc.co.uk/media/24310/bsa_statement_of_ethical_practice.pdf [Accessed 4 July 2021].
- Bryman, A. (1988) *Quantity and Quality in Social Research*. London: Unwin Hyman.
- Buchanan, D., Boddy, D. and McCalman, J. (2013) 'Getting in, getting on, getting out and getting back', in A. Bryman (ed.) *Doing Research in Organisations*. London: Routledge Library Edition, pp. 53–67.
- Call for Participants (2021) *Call for Participants*. Available at <https://www.callforparticipants.com/> [Accessed 2 July 2021].
- Corti, L., Van den Eynden, V., Bishop, L. and Woppard, M. (2020) *Managing and Sharing Research Data: A Guide to Good Practice*. London: Sage.
- Dale, A., Arber, S. and Procter, M. (1988) *Doing Secondary Research*. London: Unwin Hyman.
- Digital Curation Centre (2013) *Checklist for Data Management Plan v.4.0*. Available at https://www.dcc.ac.uk/sites/default/files/documents/resource/DMP/DMP_Checklist_2013.pdf [Accessed 30 July 2021].
- Franzke, A.S., Bechmann, A., Zimmer, M., Ess, C.M. and the Association of Internet Researchers (2020) *Internet Research: Ethical Guidelines 3.0*. Available at <https://aoir.org/reports/ethics3.pdf> [Accessed 19 July 2021].
- Groves, R.M., Presser, S. and Dipko, S. (2004) 'The role of topic interest in survey participation decisions', *Public Opinion Quarterly*, Vol. 68, No. 1, pp. 2–31.
- Groves, R.M., Singer, E. and Corning, A. (2000) 'Leverage-saliency theory of survey participation', *Public Opinion Quarterly*, Vol. 64, No. 3, pp. 299–308.
- Gummesson, E. (2000) *Qualitative Methods in Management Research* (2nd edn). Thousand Oaks, CA: Sage.
- Health Research Authority (2020) *Is My Study Research?* Available at <http://www.hra-decisiontools.org.uk/research/> [Accessed 2 July 2021].
- Hookway, N. (2008) 'Entering the blogosphere: Some strategies for using blogs in social research', *Qualitative Research*, Vol. 8, No. 1, pp. 91–113.
- Information Commissioner's Office (2018) *About the DPA 2018*. Available at <https://ico.org.uk/for-organisations/guide-to-data-protection/introduction-to-data-protection/about-the-dpa-2018/> [Accessed 30 July 2021].
- Jiang, Q., Chan, C.S., Eichelberger, S., Ma, H. and Pikkemaat, B. (2021) (online first). 'Sentiment analysis of online destination image of Hong Kong held by mainland Chinese tourists', *Current Issues in Tourism*, pp. 1–22.

- Kozinets, R.V. (2020) *Netnography: The Essential Guide to Qualitative Social Media Research* (3rd edn). London: Sage.
- Marshall, C. and Rossman, G.B. (2021) *Designing Qualitative Research* (7th edn). London: Sage Publications.
- Official Journal of the European Union (2016) Regulation (EU) 2016/679 of The European Parliament and of The Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC. Vol 59, L119, pp. 1–88. Available at <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679> [Accessed 4 July 2021].
- Peticca-Harris, A., deGama, N., and Elias, S.R.S.T.A. (2016) 'A dynamic process model for finding informants and gaining access in qualitative research' *Organizational Research Methods*, Vol. 19, No. 3, pp. 376–401
- Rowlinson, K., Appleyard, L. and Gardner, J. (2016) 'Payday lending in the UK: The regul(aris)ation of a necessary evil?', *Journal of Social Policy*, Vol. 45, No. 3, pp. 527–43.
- Saunders, M.N.K., Gray, D.E. and Bristow, A. (2017) 'Beyond the single organization: Inside insights from gaining access for large multiorganization survey HRD research', *Human Resource Development Quarterly*, Vol. 28, No. 3, pp. 401–25.
- Social Research Association (2001) *A Code of Practice for the Safety of Social Researchers*. Available at <https://the-sra.org.uk/common/Uploaded%20files/SRA-safety-code-of-practice.pdf> [Accessed 4 July 2021].
- Stationery Office, The (2018) *Data Protection Act 2018*. London: The Stationery Office. Available at: https://www.legislation.gov.uk/ukpga/2018/12/pdfs/ukpga_20180012_en.pdf [Accessed 4 July 2021].
- Trussell, N. and Lavrakas, P.J. (2004) 'The influence of incremental increases in token cash incentives on mail survey response', *Public Opinion Quarterly*, Vol. 68, No. 3, pp. 349–67.
- UK Data Service (2021) *Anonymisation*. Available at <https://www.ukdataservice.ac.uk/manage-data/legal-ethical/anonymisation.aspx> [Accessed 4 July 2021].
- Wells, P. (1994) 'Ethics in business and management research', in V.J. Wass and P.E. Wells (eds) *Principles and Practice in Business and Management Research*. Aldershot: Dartmouth, pp. 277–97.
- Whiting, R. and Pritchard, K. (2018) 'Digital ethics', in C. Cassell, A.L. Cunliffe and G. Grandy (eds) *The Sage Handbook of Qualitative Business and Management Research Methods*. Sage: London, pp. 562–79.

Further reading

- Buchanan, D., Boddy, D. and McCalman, J. (2013) 'Getting in, getting on, getting out and getting back', in A. Bryman (ed.) *Doing Research in Organisations*. London: Routledge Library Edition, pp. 53–67. This continues to provide a highly readable, relevant and very useful account of the negotiation of access.
- Corti, L., Van den Eynden, V., Bishop, L. and Woppard, M. (2020) *Managing and Sharing Research Data: A Guide to Good Practice*. London: Sage. This good practice guide deals with both primary and secondary data and contains numerous checklists and templates.
- Kozinets, R.V. (2020) *Netnography: The Essential Guide to Qualitative Social Media Research* (3rd edn). London: Sage. Chapter 6 provides a useful insight into the notions of ethical territory, research ethics and when data and people can be considered public or private. Issues of informed consent and harm are also discussed, along with concealment and fabrication.
- Suzy Lamplugh Trust website at <http://www.suzylamplugh.org> and the Social Research Association website at <http://the-sra.org.uk>. These give useful tips, information and a code of practice to help improve your personal safety.