University of Southern Denmark

The use of experiments in business research

Haug, Anders

Published in:
Collaborative Research Design

DOI:
10.1007/978-981-10-5008-4_10

Publication date:
2018

Document version:
Accepted manuscript

Citation for pulished version (APA):
Haug, A. (2018). The use of experiments in business research. In P. V. Freytag, & L. Young (Eds.), *Collaborative Research Design: Working with Business for Meaningful Findings* (pp. 223-248). Springer. https://doi.org/10.1007/978-981-10-5008-4_10

Go to publication entry in University of Southern Denmark's Research Portal

Terms of use
This work is brought to you by the University of Southern Denmark.
Unless otherwise specified it has been shared according to the terms for self-archiving.
If no other license is stated, these terms apply:

• You may download this work for personal use only.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying this open access version

If you believe that this document breaches copyright please contact us providing details and we will investigate your claim.
Please direct all enquiries to puresupport@bib.sdu.dk

Download date: 26. Aug. 2022

CHAPTER 10: The Use of Experiments in Business Research

Anders Haug

University of Southern Denmark

Abstract

This chapter provides an understanding of what experiments are and how to use this research method in business research. More specifically, the chapter explains how to (1) conduct experiments, (2) identify the different focuses of experimental studies, (3) position qualitative experiments as a research method, (4) relate experiments to scientific paradigms and (5) discuss ethical aspects of conducting experiments. The chapter may serve as inspiration and guidelines for business students, researchers and managers considering using experimental studies. Furthermore, the chapter's positioning of qualitative experiments as a research approach may increase the appeal of experiments for those not subscribing to the more positivist approaches to research.

10.1 Introduction

When talking about scientific experiments, many may think of natural scientists in white uniforms performing various tests in laboratories. However, experiments are also a commonly used research approach in many other lines of research, including several branches of the social sciences. The *Oxford English Dictionary* defines the term 'experiment' as 'the action of trying anything, or putting it to proof; a test, trial'. However, when using the term in the context of scientific methods, it typically refers to a quantitative research method that involves

1. Manipulation of one or more variables (i.e., independent variables),
2. Measuring the effects of the manipulation on one or more other variables (i.e., dependent variables), and
3. Controlling all other variables (controlled/constant variables).

For example, if testing the performance of a B2B (business-to-business) product prototype as compared to an existing product, the product used in the two experimental setups (i.e., the prototype and the existing product) would be the independent variable, the performance measures would be the dependent variables, while the context in which the test is performed and the tasks the products are used for would be among the controlled variables.

Because of the research area of the author, the empirical examples of experimental studies provided later in this chapter are from the design literature. However, as this chapter shows, experiments are also applicable in many areas of business research, not the least of which are marketing and management. Experiments are, for example, useful for studies of the effects of different types of advertising and approaches for managing employees. On the other hand, experiments, at least in their traditional form, can be difficult to apply outside laboratory settings, e.g., in studies of organisations because of the large number of variables typically involved. Such variables may influence the outcome of the experiment in a significant manner, but be extremely difficult to control or even identify. Thus, traditional experiments are rarely used in organisational studies (Anderson 2004).

This chapter aims to provide business students, researchers and managers with a basis for engaging in experiments. To accomplish this, the remainder of this chapter is structured as follows. First, the most common experimental approaches are explained and discussed. Next, the types of focuses in experimental research are identified and exemplified. Hereafter, qualitative experiments are defined and discussed. Next, action research is presented and discussed in relation to experiments. Then, experiments are discussed in relation to scientific paradigms, after which ethical issues related to experiments are described. Finally, concluding remarks are stated.

10.2 Conducting Experiments

To provide an understanding of experiments, a useful starting point is to distinguish between their different types. A basic distinction can be made between 'traditional experiments' and 'quasi-experiments'. Traditional experiments involve random allocation of participants to treatment and comparison groups, while quasi-experiments lack this random allocation (Cambell and Stanley 1963, Robson 2011). Robson (2011) also differentiates 'single case designs' as a third form of experiment. Here, the average differences of two or more groups are not compared; instead, comparison is of an individual's measurements in different treatments. This type of experiment is used when studies are sensitive to individual differences, for which reason a particular participant acts as both the intervention group and the control group. Most often, a larger number of participants are included in single-case experiments, but each participant still serves as his or her own control group. For example, if a study has the purpose of investigating the efficiency of different learning styles for a group of employees, comparing group means related to the effects of the different learning styles would not show if the efficiency of different learning styles differs among individuals. Thus, in this case, a single-case design would be useful.

Two other types of experiments sometimes mentioned are 'natural experiments' and 'qualitative experiments'. In natural experiments, the researcher does not control the experimental conditions and developments; rather, they are determined by nature or by other factors outside the control of the researcher. For example, if studying the effects of a certain type of organisational change in companies, it may be hard to do other than as a natural experiment, since companies seldom would agree to make particular changes solely for an experimental purpose. The lack of researcher involvement in natural experiments, however, implies that this approach cannot be classified as an experimental design (Johnson 2001, Robson 2011). Thus, this chapter does not deal further with this approach.

Although experiments normally are considered to be a quantitative research method, in fact, the concept of 'qualitative experiments' is also encountered in academic literature (Kleining 1986). Such encounters, at least ones using the term 'qualitative experiment', are, however, extremely rare. On the other hand, since a qualitative approach to experiments may be particularly useful in certain areas of business research, this chapter later elaborates on this approach. First, however, the most common types of experiments are in focus.

Figure 1 provides an overview of the three most common types of experiments together with the types of data collection and types of experiment settings.

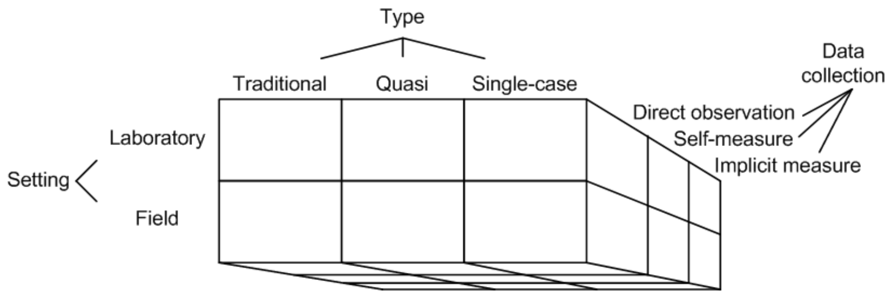

Fig. 10.1. Dimensions of experiments

In subsequent sections, the three experiment types, the two experimental settings and the three data collection techniques in Figure 1 are described, after which their strengths and weaknesses are discussed.

10.2.1 Traditional experiments

As mentioned, traditional experiments involve random allocation of participants to treatment and comparison groups—and a researcher who changes or manipulates the variable that is hypothesized to affect the outcome variable that is studied. Within 'traditional experiments' (or 'true experiments'), Robson (2011) mentions the following types of experiments (subsequently described based on Robson, 2001):

- Two group designs
- Three or more group designs
- Factorial designs
- Parametric designs
- Matched pairs design
- Repeated measures designs

These types of experiments differ in design based on whether all groups are given treatment or not, the number of groups, the number of independent variables, if the independent variable is adjusted to different levels, if participants are paired or not and the number of the conditions participants are tested under.

Two group designs can be done in two principal manners. In one, participants are randomly allocated to an intervention group, i.e., a group exposed to some stimuli during the experiment, and a non-intervention group, i.e., a group not exposed to the stimuli in focus, but a standard treatment or a placebo treatment. In the other, participants are randomly allocated to two different intervention groups, i.e., two groups exposed to different stimuli during the experiment. In both cases, the groups are compared, which can be in the form of either post-test measurements or pre- and post-test measurements. For example, if carrying out an experiment, which compares the reactions to two business marketing portfolios, there is no need for a non-intervention group and pre-tests, but if giving participants medicine, there would often be a need for a non-intervention group for testing placebo effects and pre-tests to determine the effects on the participants. The described types of two group designs can be extended with more groups, which would make them a 'three or more group design'.

A factorial design implies that there is more than one independent variable involved. The participants are randomly allocated to groups covering all possible combinations of different independent variables. This design can be in the form of post-test or pre- and post-test. This type of design is, for example, relevant when investigating effects of office designs, where there would often be a need for multiple combinations of variables (i.e., interior element selection and placement).

A parametric design implies that the independent variable can be adjusted to different levels. The participants are randomly allocated to groups covering all different levels of the independent variable. This design can be in the form of post-test or pre- and post-test. For example, if the goal is to test the effect of lighting intensity in a room in relation to work efficiency, this could involve the lighting being adjusted to several levels during the experiment.

In matched pairs designs, participants are placed in pairs based on having similar scores on a variable that is related to the dependent variable of the experiment. One member of each pair is allocated to the intervention group and one to the comparison group. The basic idea of this setup is to reduce problems related to pre-existing differences between individuals that may obscure effects of the intervention. For example, if making a study of work-motivation factors that are assumed to be related to age, there could be a need to restrict the age of the participants, for example, only including people between 25 and 35 years old, to avoid the age-spread in the groups that might be a significant contributing factor to observed differences. In this situation random assignment to groups will obscure such effects. However, by creating matched age pairs, it is possible to carry out a relatively sensitive test without having to be restricted to a narrow age range.

Repeated measures designs imply that the same participant is tested under two or more conditions, either in multiple intervention conditions or in both intervention and control conditions. In relation to experiments with B2B product prototypes, different products can constitute different interventions, for which reason this approach allows tests of multiple products in the same experiment with the same participant.

10.2.2 Single-case experiments

As mentioned, single-case (or single-subject) experiments imply that a participant is his/her own control group, as opposed to comparing the post-experimental outcomes of an intervention group to a non-intervention group. This type of experiment is used when there is a need for measuring the effect of an experimental treatment on individual participants and when group means conceal patterns that appear in individuals' data. For example, as already noted if a study's goal is to investigate different categories of employees' learning styles, comparing the means of each group (changes in) learning styles would not show if the efficiency of different styles of learning differs among individuals.

As mentioned in the introduction, although this approach is labelled as 'single-case' experiments, typically there is more than one participant included in the study in order to establish the replicability of the findings. Furthermore, it should be noted that while Robson (2011) makes a distinction between traditional and single-case experiments, others consider well-designed single-case experiments as a type of traditional experiment with comparable validity (e.g., Kazdin 1982). Robson (2011) describes the following types of single-case experimental designs:

- A-B designs
- A-B-A designs
- A-B-A-B designs
- Multiple baseline designs

A–B designs imply that a subject's baseline condition (A) is measured, after which an intervention is made and the new condition (B) of the subject is measured. For example, if a study was to investigate the effect of using a mobile device for registering receipt of goods, measures could first be taken so that a group of employees did not have this mobile device; next, measures could be taken so that they *have* this mobile device. By comparing employee efficiency in the two situations, the value of using the mobile device can be established. However, in some cases, it is not only the intervention that is responsible for effects measured. Subjects may also have become better at carrying out the task at hand as a result of participating in the experiment. To investigate this, there is a need to return to the pre-intervention situation and test this again. This is called A–B–A designs, which means that a third phase is introduced, i.e., reverting the situation to the pre-intervention baseline condition (A), e.g., the mobile device would be removed and further measurements taken. If the situation does not return to the baseline condition, the experimenter needs to seek additional/other explanations for the changes that occurred when the intervention was introduced. A–B–A–B designs mean that a second intervention phase (B) is introduced. The argument for adding a second intervention is that the person undergoing the study in this way ends up with the (presumed) beneficial intervention. Multiple baseline designs mean that a dependent variable is measured in two or more settings, behaviours or participants. Multiple baseline designs can be in the form of the three earlier mentioned forms (i.e., A-B, A-B-A, and A-B-A-B). Besides multiple baselines, multiple interventions also can be implemented. This approach could, for example, be relevant in a situation in which the effects of using different types of new software, as compared to an existing system, were to be investigated. Here, the argument for returning to the baseline would be to check if efficiency increases were caused by the software or because of the employee becoming better at the task in focus.

10.2.3 Quasi-experiments

The term 'quasi-experiment' rose to prominence in social experimentation with a chapter by Campbell and Stanley (1963) in the book *Handbook of Research on Teaching* (Robson 2011). Campbell and Stanley (1963) define a quasi-experiment as a research design that involves an experimental approach, but without random assignment to intervention and control groups. Robson (2011) mentions three types of quasi experiments to be considered (subsequently described):

- Non-equivalent groups with pre- and post-test
- Interrupted time series design
- Regression-discontinuity design

Experiments using non-equivalent groups with pre- and post-test imply that two or more groups are created based on a principle other than random assignment, such as people being put in groups based on pre-existing categories or states. This strategy can be chosen to compare different groups' reactions to an intervention. For example, in a study on the effects of using a particular stress-reduction technique for different job types, it would typically not be possible or desirable to randomly assign persons to different job positions, since certain education and experience is needed to carry out these jobs.

In interrupted time series designs, one or more experimental groups, or a control group and one or more experimental groups, are subjected to a series of measurements before and after the intervention. Having multiple measurements before and after the intervention makes a more convincing case for the effect of the intervention and allows for trends to be observed. This approach may, for example, be used to determine the development of positive attitudes towards a company after having been exposed to a company presentation, which could reveal that the presentation, rather than having a great immediate effect, has an increasing effect over time.

Regression-discontinuity designs imply that all participants are pre-tested, after which those scoring below a criterion value are assigned to one group and those above that criterion to another group. These designs are particularly useful in situations where there is a desire to target the participants who most need or deserve the intervention, while placing less-needy participants in the control group—for example, by assigning employees in stressful positions to stress-reduction training programs.

10.2.4 Choosing experimental approach

Having explained the most common types of experiments, the question emerging is, What are their strengths and limitations? Often such decisions are driven by a desire to obtain as solid research data as possible. In this context, certain experimental designs are presented in the literature as being more 'rigorous', denoting that there are fewer potential biases and a higher degree of accuracy than for alternatives. An example is 'randomized controlled trials' (RCT), which its proponents argue is the 'gold standard' for clinical trials (Robson 2011). RCT is a form of 'traditional experiment' in which participants are randomly divided into a group who are exposed to some form of intervention (treatment, instruction, etc.) and a control group who are not. Proponents of RCT consider it to be the best of means in order to assess if an intervention is effective or not. In fact, there is a growing tendency in some research areas to equate RCT with the qualityof science done, and some governmental organizations require RCT (if feasible) in order to provide funding (Robson 2011).

On the other hand, there has been much criticism of the single-minded focus on RCT, not the least from the social sciences. It is argued this focus can become an obstacle for doing meaningful research (Pawson and Tilley 1997). In this vein, Pawson and Tilley (1997) argue that an RCT approach does little or nothing to explain why an intervention has failed or succeeded, which in the context of social science is often much more interesting or important than the effect itself. They also mention that allocating participants to experimental and control groups removes that choice from the participants, which can be problematic, since 'choice is the very condition of social and individual change and not some sort of practical hindrance to understanding that change' (Pawson and Tilley 1997). However, RCT is in some cases the most efficient approach to demonstrate effects. The thing to be aware of is that in many cases it is not.

As mentioned, in some situations, the use of traditional experiments is not a feasible approach, not the least in business research. A typical argument for choosing a quasi-experiment over a traditional experiment would be that it is not possible to randomly allocate participants to treatment conditions, because of either practical or ethical reasons. Practical problems of randomized allocation of participants can occur, for example, if the goal is to study the effects of using a certain software system in relation to sales and engineering. In such a study, it would not be possible to randomly assign participants to the two treatments, but there would be a need to make salespersons carry out the sales part of the experiment and engineers the engineering part of the experiment. Ethical concerns in relation to randomized allocation can occur in cases in which it would imply withholding treatment from someone who needs it or giving someone a harmful treatment. For example, if the goal is to compare the effects of two types of managerial styles, it could be unethical to make a manager use a style that he/she believes to be harming employees. Another example is that if we were to compare work efficiency of smokers and non-smokers, random assignment would imply making non-smokers smoke, which would be considered unethical by most.

As mentioned, choosing single-case experiments is advised when the circumstances are such that group means could conceal patterns that appear in individuals' data. In other words, there is less distortion of effects with single-case experiments, because the participants act as their own control, i.e. each person's "after" score is compared to their own "before" score. In contrast in assessing effects when group means are used, it is the scale and pattern of differences in participants' scores that are then compared. So, the use of single-case experiments could reveal that for a significant proportion of the participants, in fact, another learning style was far more or less efficient but this would be obscured within a group mean.

Single-case experiments can also be argued for when the research involves testing the effectiveness of a treatment, which would benefit a participant. Thus, the ethical problem of placing individuals needing treatment in a control group diminishes. Yet another advantage of single-case experiments is that, as opposed to a traditional experiment where all participants in the intervention groups receive the same experimental treatment, in single-case experiments it is possible to adjust the experiment treatment to the particular individual. On the other hand, single-case experiments are not suited for answering many types of research questions—for example, questions focusing on how large a percentage of the target group would react in a certain way when exposed to certain stimuli (e.g., a particular B2B advertisement) cannot be addressed. Nor can questions involving comparisons of the effects of different stimuli (e.g., different B2B advertisements).

10.2.5 Experiment settings

The types of experiments described in the previous sections can take place in different types of settings—more specifically, in laboratories (i.e., laboratory experiments) or in the real world (i.e., field experiments). There are advantages and disadvantages to both approaches. When conducting experiments in real-life settings, it can be hard to make such experimental designs comply with the sometimes rigid criteria of being able to control variables, for example, demanded by much medical research (Coolican 2013).

On the other hand, field experiments also hold some advantages over laboratory experiments, particularly with respect to validity. First, a laboratory experiment is conducted in an artificial setting, which may make the experiment situation considerably different from the reality that is being studied. The artificiality of the laboratory in combination with the 'unnatural' things that the participants may be asked to do obviously produces some distortion of behaviour. Therefore, a laboratory experiment needs to consider such distortions when generalizing findings. Second, participants in real-world experiments are more inclined to feel relaxed and thus may react more naturally. For example, if studying the effects of a certain office space design, this may be better done in an actual company setting under normal work conditions, rather than simulating parts of this in a 'laboratory'.

In relation to the mentioned potential problems of laboratory experiments, an oft-mentioned type of problem is what is termed 'demand characteristics'. Demand characteristics refer to situations in which the experimenters' expectancies regarding participant behaviour create an implicit demand that makes participants perform according to these expectancies (Orne 1969, Intons-Peterson 1983). The effect of demand characteristics have been demonstrated by Intons-Peterson (1983) through experiments in which she manipulated participants by hinting what she expected the results would be like. These experiments demonstrated that if participants knew that they were expected to perform better on task A than on task B, their actual performance on task A would in fact be comparatively better. Sources of demand characteristic problems include participants being affected by laboratory settings, discussions with former participants and explicit or implicit communication with the experimenter.

Although problems of demand characteristics may also occur in field experiments, they occur to a more limited extent since, as mentioned, the fact that they are conducted in natural settings often implies that participants engage in a more natural behaviour. On the other hand, field experiments can suffer from similar validity problems as do lab experiments, not the least, the problem of participants changing behaviour as a result of being observed. A classic example of this is the studies carried out at the Hawthorne Works (a Western Electric factory in the USA) in the 1920s–1930s (Dickson and Roethlisberger 2003), which later gave name to the so-called Hawthorne effect. These studies aimed to investigate how length of the working day, heating, lighting and more affect productivity. The studies had the surprising 'result' that increases in productivity were virtually unrelated to these specific changes. In the end, it was concluded that the workers in fact were reacting positively to the attention and special treatment given by the experimenters, rather than the intended independent variables. Thus, the Hawthorne effect refers to participants modifying their behaviour, which is being experimentally measured, as a response to knowing that they are being studied, rather than as a response to the experimental manipulation. However, there has been much debate about whether this 'effect' actually exists, and if so, under which conditions (Wickström and Bendix 2000, Verstappen et al. 2004, Kompier 2006, McCarney et al. 2007).

In summary, when choosing between laboratory and field experiments, there is a trade-off between having strong control of variables in laboratories such that effects can be more unambiguously assessed, while field experiments may be a means to overcoming problems of artificiality and complex setups (such as studies of factory processes). In other words, if artificiality and experimental setup issues are not considered to cause significant problems, laboratory experiments are most often preferable, while if such issues seem likely to cause problems, these problems need to be weighted against the problems related to lack of rigorous variable control.

Often, the choice is externally determined. In a business context, it is often not possible to do experiments in laboratory settings, since what is to be tested cannot be simulated in this type of setting. For example, if testing the impact of a particular bonus scheme in an organisation or the effects on different light settings on the efficiency of a production line, such experiments would be almost impossible to conduct in laboratory settings. On the other hand, such field experiments obviously suffer from a lack of variable control, for which reason extra scrutiny is required when interpreting results, with a careful consideration given to factors beyond the experiment contributing to or causing the observed effects.

10.2.6 Data-collection techniques

Besides the type of experiment and its setting, the decision about the design of methods to be used to collect data is critical as this also can have great influence of the quality of the data produced by an experiment. According to Rogers (2003), three main data-collection principles are employed in experimental designs: observational measures, self-report measures and implicit measures. Observational measures refer to direct observation of the behaviour of participants, and they can be in the form of direct measures or classifications of behaviour. An example of a direct measure could be the time it takes for participants to complete a certain task by using different B2B products, and an example of classification of behaviour could be to use a predefined coding frame to register participant behaviour when using a certain B2B product (observations and their measurement are considered in greater detail in Chapter 9).

Self-report measures are common in business research, as they are a particularly straightforward approach to study attitudes towards and opinions of designs. Self-report measures, in general, involve interviewing participants or making them fill out questionnaires about their behaviour and/or behaviour changes. When using self-report methods, questions may be both open-ended and closed-ended. For closed-ended questions, scales are often used—for example, a typical five-level Likert scale including the answer options (1) strongly disagree, (2) disagree, (3) neither agree nor disagree, (4) agree and (5) strongly agree. Open-ended questions typically are used in relation to coding schemes to be able to test hypotheses. Such coding schemes may emerge from the use of pre-existing frameworks (deductive approach) or by counting similar words in participant responses after the data is collected (inductive approach) (interview methods and associated measurement issues are considered in greater detail in Chapter 6).

In much business research, consumer response is measured using self-report survey methods (Schoen and Crilly 2012). However, self-report measures can pose some challenges in relation to reliability and validity, since they rely on participants giving honest answers and not having affected the participants by the way questions were formulated and given. As argued by Schoen and Crilly (2012), in some cases, participants might be motivated to answer dishonestly, because some answers appear to be more socially acceptable or in order to provide answers that they believe would please the researcher. Even if participants respond to carefully worded questions as honestly as possible, studies may still hold a problem because they may not target the same thought processes that would have unfolded in a real-world situation. For example, there is evidence that real-life behaviour often is affected by more spontaneous processes because of being distracted or under time pressure when making decisions (Friese et al. 2009). Furthermore, judgements may occur non-consciously or be the result of subliminal influences, which are factors that may not be activated during self-report measures (Bargh 2002).

Implicit measure is an alternative to self-report measures, where instead of directly asking participants what they think, their attitudes or perceptions are inferred based on their response to instructions, which are not directly addressing the question in focus. In such tests, participants respond to stimulus using input devices such as keyboards, joysticks or microphones, which allow for their reaction times to be directly measured with their engagement and thus attitudes indirectly measured.

Many types of implicit tests exist, such as the Implicit Association Test (IAT), Evaluative (or Affective) Priming Task (EPT), Approach Avoidance Test (AAT), and Extrinsic Affective Simon Task (EAST). Although such tests vary in content, they are all designed to capture attitudes that individuals may be unwilling to report. For example, in the IAT, participants pair words or images representing a target concept and stimuli with known positive/negative valence into two categories. The faster this categorisation occurs, the stronger the association is between words and/or images that are grouped together (e.g., a particular brand placed with positive rather than negative words). In the EAST, participants are given different coloured cards with words. When the words are presented in white, the participant should categorize based on the words perceived as a positive or negative valence. When the words are presented in colour, the participants should categorize based on the colour alone and ignore word meaning. When coloured words are presented, categorization accuracy and speed are higher for words for which the respondent has a positive implicit attitude.

A basic idea of such implicit measures is that participants generally will respond slower to 'incompatible' trial conditions (such as associating a positive stimulus with a negative word or action) than 'compatible' trial conditions (such as associating a positive stimulus with a positive word or action). By comparing reaction times for different trial conditions, it becomes possible for researchers to infer information about the attitudes of participants without them even knowing that their reaction times are of interest to the researcher. Because of its indirectness, this approach also allows for inferring unconscious thoughts and attitudes—because their reaction time may tell something that they themselves are not consciously aware of. Other non-self-report methods are also available, such as eye tracking, brain imaging, heart rate measurement and voice pitch analysis (see, for example, Wang and Minor 2008). Such methods may be harder to employ but can measure variables that are more resistant to participant response bias than reaction times, which participants in some cases may affect (Schoen and Crilly 2012).

Finally, simulation (as described in Chapter 15) needs to be mentioned. In situations in which real-life data cannot be collected through experiments, simulating reality can be an alternative source of data. For example, if one wishes to experiment with how to improve the flow of customer orders, simulations can help predict the impact of how procedure changes can impact costs, lead times and the quantity of transactions at various stages in the process.

As this section has shown, each of the mentioned data-collection methods has both strengths and weaknesses. Therefore, there are good arguments in favour of a combination of different data-collection techniques.

10.2.7 Data analysis

Having collected data from an experiment, the next step is to analyse these data. In most research method books, as mentioned, experiments are almost exclusively described as a quantitative method, for which reason the data produced are subjected to various kinds of statistical analyses. Statistical analyses have the following common structure (Bryman and Bell 2011):

- Setting up a null hypothesis: a hypothesis stating no relation between the variables assumed to be related ('hopefully' to be declared unlikely to be true)
- Defining the acceptable level of statistical significance: the accepted degree of risk that the null hypothesis is erroneously rejected
- Determining the statistical significance of the findings: the use of various statistical tests

A null hypothesis states the commonly believed case, i.e., we believe there is no relationship between things unless this is otherwise proved. For example, if investigation was of whether using a particular tool increases employee efficiency, the null hypothesis would be 'the use of the tool is unrelated to the employee efficiency'. The collected data allows for this null hypothesis to be rejected if there is measurement of improved performance at a certain level of statistical significance; it can be then concluded that the tool has an effect on employee efficiency with the given level of statistical significance.

With regard to the acceptable level of statistical significance, according to Bryman and Bell (2011), the convention in most business research is that the maximum level of statistical significance acceptable is '$p < 0.05$', meaning that there is less than a 5 per cent chance that the sample shows a relationship that is not in the population.

Statistical analyses is not something that is particular to experiments but a kind of analysis used on most kinds of quantitative research data, including those originating from questionnaire surveys, quantitative observation studies, simulations, etc. There is a wide range of statistical methods available that are relevant for analysis of experimental data, involving comparisons of group means, defining mathematical functions that describe correlations between variables, identifying correlations between sets of variables, etc. Because of the extensiveness of this topic, providing useful guidelines in limited text is not possible. Thus, if needing to know more about this area, the reader is encouraged to obtain this knowledge from the numerous books available about statistical analysis.

10.3 Design Experiment Focuses

Having discussed how to conduct experiments, this section turns the focus to the types of knowledge that experiments can produce. In design research, as well as business research, at least three overall types of experiment focus can be identified:

- Human aspects (expertise, preferences, etc.)
- Effects of physical objects (products, spaces, etc.)
- Effects of non-physical objects (techniques, methods, etc.)

To explain these types of focuses, in the following subsections, examples of experimental studies with these three types of focuses are provided. The purpose of these examples is only to illustrate the different types of focuses that experiments can have, not to evaluate the scientific quality of the particular experimental designs. The three examples are Lawson's (1979) investigations of the nature of design expertise, Cupchik's (1996) study of responses to living and dining room appearance, and the study by Shah et al. (2001) of the use of collaborative sketching. At the end of the presentation of each of these experiments, it is described how experiments with such designs may be applied in business research.

10.3.1 A study of design expertise

The study by Lawson (1979) is an example of an experimental study with a focus on effects of different types of education. In this experiment, the nature of design expertise was explored though an experiment in which the solution strategies employed by architecture students were compared to those of science students. This was done in a two-phased experiment involving two matched groups of 18 participants each. In the first phase, the groups were made up of fifth-year architecture students and fifth-year science students, and in the second phase, the groups were first-year architecture students and school pupils who are eligible for university degree courses. In the experiment, the participants were given four pairs of coloured blocks (each side of the block either red or blue) and told to arrange four of these (one from each pair) to cover 12 squares with no blocks projecting, as seen in Figure 2. The participant was asked to maximise the amount of either blue or red showing around the external vertical face. Each of the participants was made to arrange the blocks repeatedly while different rules for which blocks could be used were implemented.

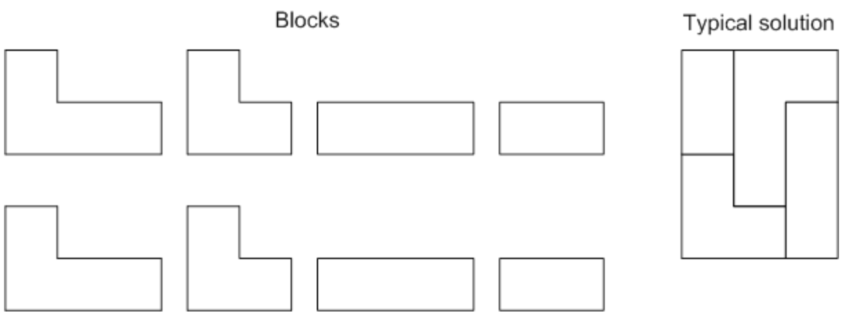

Fig. 10.2. Blocks and typical solution (adapted from Lawson, 1979)

The difference between maximal possible score and achieved score was registered in relation to each of three rule types: (1) one particular block must be present, (2) two particular blocks must both be present and (3) either or both of particular types of two blocks must be present. A computer programme monitored the experiment, and when a participant achieved a less-than-optimal colour score, the programme compared the subject's solution with the programme's optimal solution. This comparison revealed two types of errors: (1) the optimal solution utilised the same four blocks as in the subject's solution, and (2) one or more blocks being incorrectly chosen. Hereafter, statistical analysis of data was carried out. The results of the experiments suggested that while the science students selected blocks in order to discover the structure of the problem (problem-focused strategy), the architects were proceeding by generating a sequence of high-scoring solutions until one proved acceptable (solution-focused strategy).

In business research, similar experiments may be conducted to test the effects of different types of education or training. For example, a study could involve creating two groups of employees in a firm, with different kinds of training or degrees, to solve a particular business problem. In this manner, the effects of the two sets of courses may be compared.

10.3.2 A study of design object effects

The study by Ritterfeld and Cupchik (1996) is an example of an experimental study with a focus on the effects of designs. More specifically, the focus of the study was to examine how people perceive and respond to different living and dining room interiors, and what such rooms said about the assumed inhabitants. The study included two experiments; the first had the purpose of determining if living and dining spaces may be divided into different categories, and the second had the purpose of investigating personal involvement in relation to the different categories of rooms.

In the first experiment, 19 undergraduates from an advanced social psychology course participated. The participants were shown slides of 37 photographs of living and dining rooms chosen from decorating magazines, and were to rate the rooms on nine 7-point scales, where the nine scales were related to cognitive/structural qualities (complexity, order and familiarity), connotative and affective qualities (atmosphere, arousal and general preference), and semantic components (fanciness, traditional/modern and casual/informal). After having completing the rating task, the participants were asked to rank-order the scales in terms of how easy/difficult they were to apply. The data were subjected to statistical analysis, which resulted in the creation of three factors: (1) Decorative (fancy, complex, formal and stimulating, (2) Stylish (orderly, modern and cool) and (3) Familiar (familiar, 'I want to live there', warm, stimulating and orderly).

A second experiment built on the findings of the first. From the initial 37 rooms, two sets of rooms, each with 12 rooms, were created—each set included three subsets of four rooms chosen representing the three identified factors: decorative, stylish and familiar. The 24 rooms were chosen because of being the ones that had a strong match on one of these three factors and a low match on the other two. In this experiment, 24 persons from introductory psychology courses participated. In the first part of the experiment, the participants were told to evaluate (on six 7-point scales) the 12 rooms from one of the sets and to write a brief story about what might be happening in the room. Hereafter, a five-minute distractor task (thus, no data reported) was given, which required them to judge relative similarity between 12 pairs of couches. The second part of the experiment focused on recognition of details from the three types of rooms. The participants were showed details (i.e., parts of the pictures) from the 24 rooms (five seconds for each slide), and they were given five seconds to state if they had seen the room before and how certain they were of this.

The data from the experiments was subjected to a series of statistical analyses. In summary, the study showed that different kinds of cognitive and emotional dynamics shape a subject's response to familiar versus decorative versus stylish rooms. Although in the first part of the experiment, 'familiar' rooms were clearly preferred, the second part of the experiment showed that the participants were less certain about their individual features and generally had difficulty assessing familiarity. The participants distinguished the 'decorative' rooms as being a semantically rich reflection of the (assumed) inhabitant, while the geometric order of the 'stylish' rooms reduced their affective quality, but made it easier to recognize details selected from them.

In business research, similar experiments may be conducted, for example, to investigate effects of office space designs or the attributes of B2B websites.

10.3.3 A study of design technique effects

The study by Shah et al. (2001) is an example of an experimental study with a focus on effects of the use of non-physical objects. The study focuses on a collaborative sketching technique for concept generation in engineering design, named C-Sketch. When using the C-Sketch technique, designers develop graphical representations of solutions to design problems in cycles, and at the end of each cycle, the sketch is passed to the next designer who may then add, modify or delete aspects of the design solution. In this manner, sketches are passed sequentially through the design team with each designer having the possibility to add his/her own contribution to the design sketches.

To evaluate the C-Sketch technique, the first experiment focused on finding evidence of creative cognitive processes when designers used C-Sketch to solve a design problem. The participants were eight graduate students in mechanical engineering. Two solutions to each of two design problems were made in advance, i.e., the explorative cycles of C-Sketch were only simulated. In the experiment, each participant was given fifteen minutes to interpret the sketches and further improve the prepared solutions. The participants were told to think aloud while being recorded on video. Based on observing the video recordings in which the participants were thinking aloud while using the sketching technique, the authors concluded that cognitive mental processes (i.e., generative and explorative processes) could be identified when designers used an idea-generation method to solve an engineering design problem.

The second experiment aimed at investigating if C-Sketch helps designers explore new paths based on the concepts they receive from others or whether they remain fixated on their original ideas. To do so, 16 designers were paired up, and participants generated a solution sketch on their own and then exchanged it with their partner, who was asked to improve the solution received. The analysis of the sketches was done by dividing each sketch into 'units' that consisted of related drawing units (RDU), and three quantities were measured: retention (ratio of the RDUs from the original idea that survived after changes were made by the second designer), modification (ratio of the RDUs added or deleted by the second designer) and fixation (ratio of the RDUs added by the second designer to the sketch received from the first). This analysis showed that, on the average, 69 per cent of the original concept was retained, i.e., only 31 per cent of the solution was modified. Based on these findings, it was concluded that that the designers did not show tendencies to force someone else's idea towards their own first ideas while using C-Sketch, but instead they showed a greater tendency to enhance existing features in the sketch they had received. The paper by Shah et al. (2001) also describes further studies made to evaluate C-Sketch, which are beyond the scope of this chapter.

In business research, similar experiments may be conducted to test the effects of different collaboration strategies, methods and techniques that might concern employee efficiency motivation and wellbeing in team formation and/or new product development.

10.4 A Qualitative Approach to Experimental Studies

Until now, the focus has been on the most common approaches to experiments, which all are quantitative in nature. This is, as mentioned, highlighted by the fact that descriptions of qualitative experiments in research method books are extremely rare. An exception and one of the central proponents of this qualitative variant of experiments is Kleining (1986), who defines a qualitative experiment as (translated from German):

> '…(T)he intervention with relation to a (social) subject that is executed following scientific rules and towards the exploration of the subject's structure. It is the explorative, heuristic form of an experiment.'

Thus, while quantitative experiments are used to test hypotheses, qualitative experiments have a more inductive and interpretative focus. Some of the main differences between quantitative and qualitative experiments are shown in Table X.1 (based on Ravasio et al. 2014).

Experiment aspect	**Quantitative experiments**	**Qualitative experiments**
Purpose	Existing structures and behaviours are tested in order to learn about the strengths of their interdependencies.	Explores presently unknown structures, dependencies and behaviours.
Hypothesis formulation	Expected relationships between known quantifiable variables.	Does not imply quantifiable variables initially.
Variable control	Well controllable but have a tendency to be highly abstracted and thereby artificial in setting.	The results are only quantifiable after additional special treatment.
Random assignment	Subjects are assigned randomly to different conditions.	Subjects may or may not be randomly assigned to the different conditions.
Replicatability	Allow the experiment's repetition under identical conditions with identical results.	Though desired, replicability is not imperative.

Table 1. Quantitative and qualitative experiments

10.4.1 The usefulness of qualitative experiments

Although qualitative experiments, as mentioned, are not given much focus in the experimental literature, they may be useful in many cases. In line with the focus of this book—doing research with business to generate insights—the qualitative experiment has the ability both to generate ideas and to at least somewhat systematically consider conditions under which insights can be generated. This argument is highlighted by considering how qualitative experiments could have contributed in studies similar to the three in the section on design experiment focuses.

In the first of the three experiments described in that section (i.e., comparing solution strategies of architecture students to those of science students), the quantitative approach produced data about how far from the optimal solution the subjects were (i.e., maximal number of colours possible minus the number of generated colours). Although this experiment tells something about how good architects are at solving different types of puzzles compared to scientists, this type of experiment does not provide deep insights into architects' ability to design visually pleasing objects. To do so may require a qualitative approach, for example, to make architects create designs under certain experimental settings that are later analysed, compared and discussed by design experts in order to determine their relative quality—i.e., similar to how projects are evaluated at architecture, design and art schools, where such evaluations to a large extent are based on personal taste and experience, rather than concrete measures.

In the second of the three experiments described in the section on design experiment focuses (i.e., how people perceive and respond to different living and dining room interiors), it was shown that certain rooms were associated with different styles for which the subjects had different attitudes. However, to provide a thorough explanation of why the rooms from the study were considered to be familiar, decorative or stylish, and why the subjects had the particular attitudes to these rooms, it would hardly be possible without qualitative data. To explore such topics, there would be a need to interview the participants about such topics and interpret and analyse their answers

In the third of the three experiments described in the aforementioned section (i.e., a collaborative sketching technique for concept generation in engineering design), it was shown that the use of the sketching technique in focus could have a positive effect on design quality. This experiment focused on engineering design, and quantitative measures of quality were defined (relating to the technical feasibility and potential for fulfilling desired specifications). Had the experiment, on the other hand, been about understanding the aesthetic qualities of solutions produced, it would be extremely difficult to define meaningful quantitative measures to describe this. Instead, such an experiment could, as for the first experiment type, involve comparison of the designs produced by the participants that would be discussed by design experts in order to determine their relative quality.

10.4.2 Analysis Issues in Qualitative Experiments

As previously described, in qualitative experiments the main emphasis is on interpretation of the comparisons of outcomes, as compared to quantitative ones where the focus is on statistical testing such that the amount of impact of an experimental treatment can be assessed. However, there are at least two methods available to use in qualitative experiments that allow conclusions to be drawn similar to those emerging from quantitative experiments. First, there is the possibility of quantifying the qualitative experiment data. More specifically, if analysing qualitative data through the use of coding schemes and categories, such data can be quantified, and on this basis statistics can be applied. This creates problems of validity and reliability as discussed in Chapter 11 on qualitative analysis. As noted there, the translation of rich qualitative data into variables and categories may result in loss of the value and meaning of that data. And, an interpretative qualitative coding that creates those variables and categories is likely to be inconsistently applied. The second type of approach involves improving validity via data saturation. Data saturation occurs 'when researchers sense they have seen or heard something so repeatedly that they can anticipate it' and 'collecting more data is deemed to have no further interpretive value' (Sandelowski 2008). In this manner, a qualitative experiment would have to be repeated until no new information is produced. There are different perspectives on the number of times something needs to be observed sufficiently 'repeated' for saturation to occur (Guest et al. 2006). This to a large extent depends on the homogeneity of the population, where, greater homogeneity, obviously, implies smaller samples can be used. Different minimum numbers of experimental subjects needed in qualitative experiments can be found in the literature, where some set this number as low as six, while others set it at 25 (Kuzel 1992, Morse 1994, Creswell 1998). While this is useful in terms of guiding the generation of data and ensuring there is a better quality of data to analyse, it does not really deal with issues of how to interpret and analyse the data that is generated.

Although qualitative experiments require quite different analysis methods as compared to quantitative ones, the quantitative experiment types described in the Conducting Experiments section may also be relevant in a qualitative setting. This can be envisaged as a hybrid design where the researcher employs a formal experimental design and the procedures of standard quantitative experiments in qualitative experiments.

10.5 Action Research

Another commonly applied approach in business research involving experimentation is action research. Action research involves introducing problems and issues into an environment and observing and interpreting the effects of these interventions. It is most often qualitative but can also be quantitative or include quantitative elements. The term 'action research' was introduced by Kurt Lewin in 1946 to denote a research approach in which the researcher combines theory generation and the changing of social systems. Since the emergence of Lewin's ideas, the action research approach has been adapted by many researchers from different fields—and different perceptions of what exactly action research is and how it should be carried out exist (Waterman et al. 2001). According to Susmann (1978), the definition by Rapoport (1970) of action research is (or at least was) the most frequently quoted in contemporary literature on the subject:

> 'Action research aims to contribute both to the practical concerns of people in an immediate problematic situation and to the goals of social science by joint collaboration within a mutually acceptable ethical framework.'

A widely adapted version of action research views the process as cyclic and involving planning a change, acting and observing what happens, reflection on the processes and consequences, and planning further action and repeating the cycle (Robson 2011).

Action research implies two roles for the researcher, namely as a participant in the change process (often initiator) and as an observer who reports from the study. Action research is, therefore, in great contrast to the positivist science tradition where researchers attempt to disengage themselves from their study subjects. To differentiate and, where appropriate, separate these two roles provides a great challenge for an action researcher. Because this process inevitably involves ambiguity and has only limited transparency, in some parts of academia (in particular those with strong positivistic orientation), action research has been criticised for being unscientific. On the other hand, since the areas typically investigated by using action research are often blurred (complex social systems), the intimacy of action research may be seen as a means of promoting appropriate change and understanding of practice (Waterman 2001). It has obvious applications to designing research that allows researchers and managers to work together to co-create knowledge.

Action research is not an experiment in the same sense as the other designs that are described in this chapter. It has similar characteristics in that there can be an assessment of a system, some form of intervention in that system and observation and assessment of changes that follow that intervention. However, action research focuses on solving a problem in practice, implying that practice defines the nature and design of the study, rather than the researcher doing so, as when conducting experiments.

In other words, action research involves being open to transforming the process of research using both single-loop and double-loop learning cycles; in other words, connecting a strategy and action with a result and modifying the strategy and/or result (single loop) or re-evaluating and reframing one's goals, values and beliefs to improve the process of engagement (double loop) (Argyris and Schön 1978). These adaptive processes are in contrast to experiments where dependent, independent and controlled variables are established beforehand. Also, experiments follow a predefined section of steps, as opposed to action research in which these typically emerge during the study.

In many business research contexts where some form of experimentation is desired, action research may be a better choice than experiments. Action research is particularly useful when the researcher does not have a clear idea of where to start and does not have a lot of time to invest in the study. Thus, action research is useful for exploratory research. Also, action research is valuable when there is a need to be responsive to the changing demands of a situation. For example, if a study focuses on how to use performance measures during a change management project, action research may be appropriate, since this used for evaluation of an on-going program. And, as already stated, action research by its nature involves developing relationships with business stakeholders, which in itself creates insights and opportunities.

10.6 Scientific Paradigms and Experiments

The choice of research methods is to a large extent connected to the scientific position of the researcher. Although experiments mostly are associated with positivistic notions, they can also be of great value for researchers working under other scientific paradigms, such as critical realism, critical theory and social constructivism (Guba and Lincoln 1994, Bhaskar 2008). To some it may appear controversial to claim that quantitative experiments could be relevant from a social constructivist perspective, since social constructivist approaches are often considered to be more or less restricted to qualitative methods. However, if looking at the full spectrum of social constructivist approaches, this is a misconception—which, for example, is highlighted by the fact that one of the most prominent social constructivists, Pierre Bourdieu, to a large extent employed another type of quantitative approach, namely questionnaire surveys with statistical analysis, for example, in his studies of taste (Bourdieu 1984). In this context, it is relevant to note is that there are different perspectives on constructivism, where a central difference relates to the question of whether the physical reality is a construction or it is merely our knowledge about the physical reality that is constructed (Fuglsang og Olsen 2004). Obviously, the latter perspective makes quantitative experiments seem more useful.

Having claimed that experiments, even in the quantitative form, are compatible with critical theory and social constructivist perceptions, it should be noted that the way such research would employ experiments would be likely to differ from more positivistic approaches (see Chapter 13 for further discussion on critical theory). A central difference is that, from a critical theorist or social constructivist perspective, experiments would typically involve qualitative analysis and interpretation of the effects of interventions to a larger extent. For example, if a constructivist or a critical theorist were to employ experiments to investigate how certain media affects people's perceptions of the world, it would be natural to include in-depth interviews or even conduct purely qualitative experiments in order to understand attitude changes caused by the experimental intervention (e.g., analysing their attitudes before and after having red a set of articles). And, constructivists and critical theorists in principle would be particularly supportive of action research as way of deconstructing and reconstructing reality via researcher participation.

In summary, particular scientific paradigms do not necessarily rule out using the common experimental approaches, i.e., the quantitative ones. However, it should be recognised that the design, implementation and analysis would be likely to differ in particular ways—i.e., it is more likely that while positivists would most likely design traditional, purely quantitative experiments, critical realists and constructivist researchers would typically make use of at least some qualitative aspects in their interpretation of findings and might move to less-structured experimental designs.

10.7 Ethical Considerations

As this chapter already has touched a little upon, experiments can imply a need for serious ethical consideration. For example, ethical issues can emerge when using random assignment, where it involves withholding treatment from someone who needs it or giving someone a harmful treatment. In business and social research it is more common to see 'harm' as something that involves deception of participants. However, in some cases this is necessary as providing details about the experiment would make participants act unnaturally. On the other hand, not telling participants about certain experiment details implies that they cannot themselves decide if the experiment crosses their personal boundaries, and it is up to the experimenter to make this judgement. This is doubly a problem in business research where the experimenter is possibly placing himself or herself in a position of power and entirely determining the nature and boundaries of the research rather than collaborating with business to produce mutually beneficial outcomes.

An extreme example of the problem of deceiving participants and potentially harming them is illustrated via the 'Milgram study'. In this study, the participants thought that they had signed up for a fairly benign-sounding memory test, but in fact it was an experiment in which they were led to believe that they, based on the experimenter's orders, were giving electric shocks to a person when he or she answered questions wrong (Milgram 1963, 1974). This was not actually happening, but the ethical concerns raised by this study relate to the emotional stress that some of the participants suffered because of having been pushed to give what they thought were electrical shocks to others (Benjamin and Simpson 2009).

In business research, deception issues in experiments would rarely be as serious as in the Milgram study, but still ethical problems may arise. These kind of problems can emerge if an experiment is presented as having a purpose that the percipient finds interesting or noble, but in fact it has another purpose which the participants do not wish to support—for example, if a marketing study is disguised as a general study of human psychology or vice versa. Another type of problem can emerge if telling participants that an experiment focuses on certain behaviour, which the participant does not mind having observed, but in fact the experiment focuses on behaviour that a participant would mind having observed. This could, for example, be an experiment presented as a study of how people network to form new business relations with others but is in fact a study of how people use manipulation to get others to like them. Even when the deception is not a serious one, it is likely that the fact that participants are being deceived is going to mean that some persons may feel uncomfortable, and that trust between researchers and participants will be eroded and future research may be compromised. Thus, as an overall guideline, the researchers should inform participants about the overall purpose of the experiment while explaining that exactly what is observed cannot be revealed—thereby giving participants a chance to opt out if not accepting these premises.

10.8 Concluding remarks

This chapter explained and discussed experiments as a business research method. As the chapter indicates, there is potential value in all the mentioned experiment types. However, it could be argued that if it is possible to conduct a traditional experiment, this, in many cases, would be the best choice, because this enables meaningful interpretation of cause and effect. Also, it cannot be denied that there are commonly held perceptions of the value of the superior validity of this type of experiment in certain parts of academia and in many funding organizations.

On the other hand, in many cases, traditional experiments are not a feasible approach, in particular in research within the social sciences, including business research. More specifically, in some cases, means for groups of participants conceal data describing individuals, which could necessitate single-case experiments, while in other cases, practical or ethical aspects prevent random assignment, which may necessitate a quasi-experimental approach. Furthermore, the chapter argued that although qualitative experiments hardly are mentioned in business research method books, these could be particularly relevant in business research, since, in many cases, the usual questioning methods on their own do not provide a sufficiently deep understanding of the phenomenon in focus. Instead, quantitative and qualitative ways of observing and recording behaviour responses and systematically comparing and interpreting responses of participants can provide substantial insights to researchers and businesses (see chapter 9 on observation for further discussion of this topic).

The chapter also dealt with action research as an alternative to experiments. More specifically, in some situations where experimentation is desired, but there is a lack of insight about the topic in focus, action research can be a good alternative. Action research allows a more flexible approach in which relevant variables can be discovered during the study, rather than being defined beforehand. It has considerable potential for the business researcher.

In the common approaches to experiments, such data collection is typically of a quantitative nature; however, this chapter highlights the potential value of designing qualitative experiments for including quantitative elements in otherwise qualitative work. This creates the potential for novel, interesting and valuable research outcomes.

10.9 References

Anderson, Valerie (2004), *Research Methods in Human Resource Management*. London: Chartered Institute of Personnel and Development.

Argyris, Chris and Donald Schön (1978) *Organizational Learning: A Theory of Action Perspective.* Reading, MA: Addison Wesley.

Bargh, John A. (2002), 'Losing Consciousness: Automatic Influences on Consumer Judgment, Behavior, and Motivation', *Journal of Consumer Research* 29 (2), 280–285.

Benjamin Ludy T. and Jeffry A. Simpson (2009), 'The Power of the Situation: The Impact of Milgram's Obedience Studies on Personality and Social Psychology'. *American Psychologist*, 64 (1), 12–19.

Bhaskar, Roy (2008), *A Realist Theory of Science* (reprint of original from 1975). London: Taylor Francis Ltd.

Bourdieu, Pierre (1984), *Distinction: A Social Critique of the Judgement of Taste* (translation of La Distinction from 1979). London: Routledge.

Bryman Alan and Emma Bell (2011), *Business Research Methods*, 3rd ed. New York: Oxford University Press.

Campbell, Donald T and Julian C. Stanley(1963), *Experimental and Quasi-Experimental Designs for Research*. Boston: Houghton Mifflin.

Cohen, Jacob (1992), 'A Power Primer', *Psychological Bulletin*, 112 (1), 155–159.

Coolican, Hugh (2013), *Research Methods and Statistics in Psychology*, 5th ed. New York: Routledge.

Creswell, John (1998), *Qualitative Inquiry and Research Design: Choosing Among Five Traditions*. Thousand Oaks, CA: SAGE Publications.

Crilly, Nathan, James Moultrie and P. John Clarkson (2004), 'Seeing Things: Consumer Response to the Visual Domain in Product Design', *Design Studies*, 25 (6), 547–577.

Dickson, William J. and Fritz J. Roethlisberger (2003), *Management and the Worker*. London: Routledge.

Friese, Malte, Wilhelm Hofmann and Michaela Wänke (2009), 'The Impulsive Consumer: Predicting Consumer Behavior with Implicit Reaction Time Measurements', in *Social Psychology of Consumer Behavior*, Michaela Wänke, ed. New York: Psychology Press, 335–364.

Fuglsang, Lars and Poul B. Olsen (2004), *Videnskabsteori i samfundsvidenskaberne*, 2nd ed. Frederiksberg, Denmark: Roskilde Universitetsforlag.

Guba, Egon G. and Yvonna S. Lincoln (1994) 'Competing Paradigms in Qualitative Research', in *Handbook of Qualitative Research*, Norman K. Denzin and Yvonna S. Lincoln, eds. Thousand Oaks, CA: SAGE Publications, 105–117.

Guest, Greg, Arwen Bunce and Laura Johnson (2006), 'How Many Interviews Are Enough? An Experiment with Data Saturation and Variability', *Field Methods*, 18 (1) 59–82.

Intons-Peterson, Margaret J. (1983), 'Imagery Paradigms: How Vulnerable Are They to Experimenters' Expectations?' *Journal of Experimental Psychology - Human Perception and Performance*, 9 (3), 394–412.

Johnson, Burke (2001), 'Toward a New Classification of Nonexperimental Quantitative Research', *Educational Researcher*, 30 (2) 3–13.

Kazdin, Alan E. (1982), *Single-Case Research Designs*. New York: Oxford University Press,.

Kleining G (1986) Das qualitative experiment [The qualitative experiment]. Kölner Zeitschrift für Soziologie und Sozialpsychologie 38(4): 724-750.

Kompier, Michiel A. J. (2006), The 'Hawthorne Effect' Is a Myth, but What Keeps the Story Going?' *Scandinavian Journal of Work and Environmental Health*, 32 (5), 402–12.

Kuzel, Anton (1992), 'Sampling in Qualitative Inquiry', in *Doing Qualitative Research*, Benjamin Crabtree and William Miller, eds. Newbury Park, CA: SAGE Publications, 31–44

Lawson, Bryan R. (1979), 'Cognitive Strategies in Architectural Design', *Ergonomics*, 22 (1), 59–68.

Lewin, Kurt (1946), 'Action Research and Minority Problems', *Journal of Social Issues*, 2 (4), 34–46.

McCarney, Rob, James Warner, Steve Iliffe, Robert van Haselen, Mark Griffin and Peter Fisher (2007), 'The Hawthorne Effect: A Randomised, Controlled Trial', *BMC Medical Research Methodology*, 7 (July), 7–30.

Milgram, Stanley (1963), '*Behavioral Study of Obedience*, *Journal of Abnormal and Social Psychology*', 67 (4) 371–8.

——— (1974), *Obedience to Authority: An Experimental View*. New York: HarperCollins.

Morse, Janice (1994), 'Designing Funded Qualitative Research', in *Handbook of Qualitative Research*, Norman Denzin and Yvonna Lincoln, eds. Thousand Oaks, CA: SAGE Publications 220–235.

Orne, Martin T. (1969), 'Demand Characteristics and the Concept of Quasi-Controls', in *Artifacts in Behavioral Research*, Robert Rosenthal and Ralph L. Rosnow, eds. New York: Academic Press.

Pawson, Ray and Nick Tilley (1997), *Realistic Evaluation*. London: SAGE Publications.

Rapoport, Robert N. (1970), 'Three Dilemmas of Action Research', *Human Relations,* 23 (6), 499–513.

Ravasio, Pamela, Sissel Guttormsen-Schäar and Vincent Tscherter (2014), 'The Qualitative Experiment in HCI: Definition, Occurrences, Value and Use', (*accessed 23 October 2014). [available at* http://*pamela.shirahime.ch/QualExp.pdf].*

Ritterfeld, Ute and Gerald C. Cupchik (1996), 'Perceptions of Interior Spaces', *Journal of Environmental Psychology*, 16 (4), 349–360.

Robson, Colin (2011), *Real World Research: A Resource for Users of Social Research Methods in Applied Settings*, 3rd ed. West Sussex, UK: John Wiley & Sons.

Rogers, Wendy S. (2003), *Social Psychology: Experimental and Critical Approaches*. Maidenhead, UK: *Open University Press.*

Schoen, Katrina and Nathan Crilly (2012), 'Implicit Methods for Testing Product Preference: Exploratory Studies with the Affective Simon Task', in *Proceedings of 8th International Design and Emotion Conference*, Jamie Brasset, Janet McDonnell and Matt Malpass, eds. London: Central Saint Martins College of Art and Design.

Shah, Jami J., Noe Vargas-Hernández, Joshua D. Summers and Santosh Kulkarni (2001), 'Collaborative Sketching (C-Sketch) – An Idea Generation Technique For Engineering Design', *Journal of Creative Behavior*, 35 (3), 168–198.

Susman, Gerald I. and Roger D. Evered (1978), 'An Assessment of the Scientific Merits of Action Research', *Administrative Science Quarterly*, 23 (4), 582–603.

Verstappen, Wim H., Trudy van der Weijden, Gerben ter Riet, Jeremy Grimshaw, Ron Winkens and Richard P. Grol (2004), 'Block Design Allowed for Control of the Hawthorne Effect in a Randomized Controlled Trial of Test Ordering', *Journal of Clinical Epidemiology*, 57 (11), 1119–1123.

Wang, Yong Jiang and Michael S. Minor (2008), ‘Validity, Reliability, and Applicability of Psychophysiological Technique’s in Marketing Research’, *Psychology and Marketing*, 25 (2), 197–232.
Waterman, Heather, D. Tillen, Rumona Dickson and Korrie De Koning (2001), ‘Action Research: A Systematic Review and Guidance for Assessment’, *Health Technology Assessment*, 5 (23), 1–166.
Wickstroöm, Gustav and Tom Bendix (2000), ‘The “Hawthorne Effect” – What Did the Original Hawthorne Studies Actually Show?’ *Scandinavian Journal of Work and Environmental Health*, 26 (4), 363–67.