Note on terminology

The terms 'reliability' and 'validity' are avoided by many proponents of flexible, qualitative design. Lincoln and Guba (1985, pp. 294–301), for example, prefer the terms *credibility, transferability, dependability* and *confirmability*. This line was followed in the first edition of this text (Robson, 1993, pp. 403–7), where definitions and discussion of these alternative terms was provided. However, this attempt to rename and disclaim the traditional terms continues to provide support for the view that qualitative studies are unreliable and invalid (Kvale 1996, p. 73). As Morse (1999) puts it in a forceful journal editorial entitled 'Myth #93: reliability and validity are not relevant to qualitative inquiry':

> To state that reliability and validity are not pertinent to qualitative inquiry places qualitative research in the realm of being not reliable and not valid. Science is concerned with rigor, and by definition, good rigorous research must be reliable and valid. If qualitative research is unreliable and invalid, then it must not be science. If it is not science, then why should it be funded, published, implemented, or taken seriously? (p. 717)

While this argument goes over the top in apparently denying any value to non-scientific endeavours, it has force when we are seeking to characterize our research as scientific, following the arguments developed in chapter 1.

The problem is not so much with the apple-pie desirability of doing reliable and valid research, but the fact that these terms have been operationalized so rigidly in fixed design quantitative research. An answer is to find alternative ways of operationalizing them that are appropriate to the conditions and circumstances of flexible, qualitative enquiry.

Validity

What do we mean by claiming that a piece of qualitative research is valid, that it has validity? It is something to do with it being accurate, or correct, or true. These are difficult (some would say impossible) things to be sure about. It is possible to recognize situations and circumstances which make validity more likely. These include the features of 'good' flexible design listed in box 6.1. Conversely, it is pretty straightforward to come up with factors likely to lead to invalid research. As with fixed, quantitative designs, they can be thought of as 'threats' to validity, and are discussed below.

An alternative, though related, tack is to focus on the credibility or trustworthiness of the research. The fact that some persons find it credible, or are prepared to trust it, is in itself a pretty weak justification. They may find it

believable because it fits in with their prejudices. However, if the concern is with what might be appropriate bases for judging something to be credible, this returns us to consideration of what constitutes good quality research, and possible threats to validity.

Threats to validity in flexible designs

Maxwell (1992) has presented a useful typology of the kinds of understanding involved in qualitative research. The main types are *description, interpretation* and *theory*. Each of the main types has particular threats to its validity (see also Elsworth, 1994, who presents a similar analysis, also from a realist perspective).

Description   The main threat to providing a valid description of what you have seen or heard lies in the inaccuracy or incompleteness of the data. This suggests that audio- or video-taping should be carried out wherever feasible. Note that the simple fact that you have a tape does not mean that it must be fully transcribed. Where taping is not feasible, the quality of your notes is very important. These issues are discussed in detail in chapter 9, p. 289.

Interpretation   The main threat to providing a valid interpretation is that of imposing a framework or meaning on what is happening rather than this occurring or emerging from what you learn during your involvement with the setting. This does not preclude a style of research where you do start with some kind of prior framework, but this must be subjected to checking on its appropriateness, with possible modification. Mason (1996) shows how you might go about demonstrating the validity of your interpretation:

> In my view, validity of interpretation in any form of qualitative research is contingent upon the 'end product' including a demonstration of how that interpretation was reached. This means that you should be able to, and be prepared to, trace the route by which you came to your interpretation . . . The basic principle here is that you are never taking it as self-evident that a particular interpretation can be made of your data but instead that you are continually and assiduously charting and justifying the steps through which your interpretations were made. (p. 150)

Note that Maxwell's notion of 'interpretation' refers specifically to interpretation of the meaning and perspective of participants, as in 'interpretive' research. He would consider the wider use of interpretation given here as not distinguishable from 'theory'.

Theory   The main threat is in not considering alternative explanations or understandings of the phenomena you are studying. This can be countered by actively seeking data which are not consonant with your theory. See the discussion of 'negative case analysis' below, and in chapter 14, p. 490.

Lincoln and Guba (1985) discuss various possible threats to the validity of flexible design research, dividing them into the three broad headings of *reactivity*, *respondent biases* and *researcher biases*. Reactivity refers to the way in which the researcher's presence may interfere in some way with the setting which forms the focus of the study, and in particular with the behaviour of the people involved. Respondent bias can take various forms, ranging from obstructiveness and withholding information – when, for example, the researcher is seen as a threat – to the 'good bunny' syndrome, when the respondent tries to give the answers or impressions which they judge that the researcher wants. Researcher bias refers to what the researcher brings to the situation in terms of assumptions and preconceptions, which may in some way affect the way in which they behave in the research setting, perhaps in terms of the persons selected for observation or interview, the kinds of questions asked, or the selection of data for reporting and analysis.

Phenomenological approaches to qualitative research (e.g. Crotty, 1998) stress the importance of reflexivity, i.e. an awareness of the ways in which the researcher as an individual with a particular social identity and background has an impact on the research process. They take the view that 'The ability to put aside personal feelings and preconceptions is more a function of how reflexive one is rather than how objective one is because it is not possible for researchers to set aside things about which they are not aware' (Ahern, 1999, p. 408). Ahern has produced a useful set of suggestions to help achieve what she terms 'reflexive bracketing', i.e. using reflexivity to identify areas of potential bias; these are summarized in box 6.4.

These issues are present in all research involving people. However, the nature of much flexible design research is such that they are often particularly problematic. There is typically a close relationship between the researcher and the setting, and between the researcher and respondents. Indeed, the notion of the 'researcher-as-instrument' emphasizes the potential for bias. Padgett (1998) presents a range of commonly used strategies to deal with these threats; they are summarized in box 6.5 and discussed below.

Prolonged involvement   Involvement over a period of years was a defining characteristic of ethnography in its traditional anthropological version. In most current studies following the ethnographic approach fieldwork is much more condensed, but a period of weeks or even months is still usual, much longer than is typical in fixed methods research. This relatively prolonged involvement is also typical of other styles of flexible design research. As indicated in box 6.5, this helps to reduce both reactivity and respondent bias. Researchers who spend a long time in the setting tend to become accepted and any initial

Box 6.4

Using reflexivity to identify areas of potential researcher bias

1   Write down your personal issues in undertaking this research, the taken-for-granted assumptions associated with your gender, race, socio-economic status, and the political milieu of your research. Finally, consider where the power is held in relation to your research project and where you belong in the power hierarchy.

2   Clarify your personal value systems and acknowledge areas in which you know you are subjective.

3   Describe possible areas of potential role conflict. Are there particular types of people and/or situations with or in which you feel anxious, annoyed, at ease? Is the publication of your findings likely to cause problems with a group of people? Consider how this possibly could influence whom you approach or how you approach them.

4   Identify gatekeepers' interests and consider the extent to which they are disposed favourably towards your project. This can help you prevent potential role conflicts.

5   Recognize feelings that could indicate a lack of neutrality. These include avoiding situations in which you might experience negative feelings, seeking out situations in which you will experience positive feelings.

6   Is anything new or surprising in your data collection or analysis? If not, is this cause for concern, or is it an indication of saturation? On occasion, stand back and ask yourself if you are 'going native'.

7   When blocks occur in the research process, re-frame them. For example, is there another group of people who can shed light on this phenomenon? Would an additional form of data collection, such as document analysis or diaries, give a greater insight?

8   Even when you have completed your analysis, reflect on how you write up your account. Are you quoting more from one respondent than another? If you are, ask yourself why.

9   Consider whether the supporting evidence in the literature really is supporting your analysis or if it is just expressing the same cultural background as yourself.

10   A significant aspect of resolving bias is the acknowledgement of its outcomes. Therefore, you might have to re-interview a respondent or reanalyse the transcript once you have recognized that bias in data collection or analysis is a possibility in a specific situation. It is also worth remembering that even if preconceptions and biases are acknowledged, they are not always easily abandoned.

(Abridged from Ahern, 1999, pp. 408–10.)

Box 6.5

Strategies for dealing with threats to validity

Strategy*	Threat to validity Reactivity	Researcher bias	Respondent bias
Prolonged involvement	Reduces threat	Increases threat	Reduces threat
Triangulation	Reduces threat	Reduces threat	Reduces threat
Peer debriefing/ support	No effect	Reduces threat	No effect
Member checking	Reduces threat	Reduces threat	Reduces threat
Negative case analysis	No effect	Reduces threat	No effect
Audit trail	No effect	Reduces threat	No effect

* See text for explanation of the strategies.
(After Padgett, 1998, p. 95.)

reactivity reduces. Similarly, it permits the development of a trusting relationship between the researcher and respondents where the latter are less likely to give biased information.

There can, however, be greater researcher bias with prolonged involvement. A positive or negative bias may build up. It may be difficult to maintain the researcher role over an extended period of time (the 'going native' threat); or developing antipathy might result in a negative bias.

Triangulation This is a valuable and widely used strategy. As discussed in more detail in chapter 12 (p. 371), it involves the use of multiple sources to enhance the rigour of the research. Denzin (1988) has distinguished four types of triangulation:

- *data triangulation*: the use of more than one method of data collection (e.g. observation, interviews, documents);
- *observer triangulation*: using more than one observer in the study;
- *methodological triangulation*: combining quantitative and qualitative approaches;
- *theory triangulation*: using multiple theories or perspectives.

Triangulation can help to counter all of the threats to validity. Note, however, that it opens up possibilities of discrepancies and disagreements among the different sources. Thus interviews and documents may be contradictory; two observers may disagree about what has happened. Bloor (1997, pp. 38–41) argues that while triangulation is relevant to validity, it raises both logical and practical difficulties, e.g. when findings collected by different methods differ to a degree which makes their direct comparison problematic.

Peer debriefing and support   Peer groups (i.e. of researchers or students of similar status who are involved in flexible design research) can have a number of valuable functions. They can contribute to guarding against researcher bias through debriefing sessions after periods in the research setting. Such groups can also fulfil something almost amounting to a therapeutic function. This type of research can be extremely demanding and stressful for the researcher, and the group can help you cope.

Member checking   This involves returning (either physically or through correspondence, phone, e-mail etc.) to respondents and presenting to them material such as transcripts, accounts and interpretations you have made. It can be a very valuable means of guarding against researcher bias. It also demonstrates to them that you value their perceptions and contributions. There are potential problems; perhaps your interpretation is challenged, or a respondent gets cold feet and seeks to suppress some material. It is essential that you have a pre-agreed clear understanding with them about the rules governing such situations, and that you respect both the spirit and the letter of such agreements. However, a supine giving in to any criticism is not called for. Disagreements can usually be negotiated in a way which reflects both respondents' concerns and the needs of the study. Bloor (1997, pp. 41–8) discusses some of the complexities with examples.

Negative case analysis   This is discussed in more detail in chapter 14, p. 490. The search for negative cases is an important means of countering researcher bias. As you develop theories about what is going on, you should devote time and attention to searching for instances which will disconfirm your theory. You may do this using data you already have, or by collecting additional data. It is sometimes referred to as 'playing the devil's advocate', and you have a responsibility to do it thoroughly and honestly. Don't be too concerned that this procedure will lead to your ending up with a set of disconfirmed theories. In practice, it usually amounts to developing a more elaborated version of your theory.

Audit trail   The notion is that you keep a full record of your activities while carrying out the study. This would include your raw data (transcripts

of interviews, field notes, etc.), your research journal (see p. 1), and details of your coding and data analysis (Lincoln and Guba, 1985, appendices A and B).

Maxwell (1996, pp. 92–6) and Miles and Huberman (1994; pp. 262–77) provide alternative, but overlapping, sets of strategies which might be considered. Note, however, that while using such strategies will undoubtedly help in ruling out threats to validity, there is no foolproof way of *guaranteeing* validity. Remember, too, that the strategies only help if you actually use them. Whereas in traditional fixed design research (particularly in true experimentation) threats to validity are essentially dealt with in advance as part of the design process, most threats to validity in flexible design research are dealt with after the research is in progress, using evidence which you collect after you have begun to develop a tentative account.

Reliability in flexible designs

Reliability in fixed design research is associated with the use of standardized research instruments, for example formal tests and scales as discussed in chapter 10. It is also associated with the use of observation, where the human observer is the standardized instrument. The concern is whether the tool or instrument produces consistent results. Thinking in such terms is problematic for most qualitative researchers (Mason, 1996, chs 3 and 4). At a technical level, the general non-standardization of many methods of generating qualitative data precludes formal reliability testing. Nevertheless, there are common pitfalls in data collection and transcription including equipment failure, environmental distractions and interruptions, and transcription errors. Easton et al. (2000) suggest strategies to minimize the risk from these errors.

In a more general sense, researchers using flexible designs do need to concern themselves seriously with the reliability of their methods and research practices. This involves not only being thorough, careful and honest in carrying out the research, but also being able to show others that you have been. One way of achieving this is via the kind of audit trail described above.

Generalizability in flexible designs

Maxwell (1992; 1996, p. 96) makes a useful distinction between *internal* and *external generalizability*. Internal generalizability refers to the generalizability of conclusions within the setting studied. External generalizability is generalizability beyond that setting. The former is an important issue in flexible designs. If you are selective in the people you interview, or the situations that

you observe, in a way which, say, excludes the people or settings which you find threatening or disturbing, this is likely to bias your account.

External generalizability may not be an issue. A case study might just be concerned with explaining and understanding what is going on in a particular business, drop-in centre or whatever is the focus of the study. It very rarely involves the selection of a representative (let alone random) sample of settings from a known population which would permit the kind of statistical generalization typical of survey designs. However, this does not preclude some kind of generalizability beyond the specific setting studied. This may be thought of as the development of a theory which helps in understanding other cases or situations (Ragin, 1987; Yin, 1994), sometimes referred to as *analytic* or *theoretical generalization*: 'Here the data gained from a particular study provide theoretical insights which possess a sufficient degree of generality or universality to allow their projection to other contexts or situations' (Sim, 1998, p. 350).

Research Traditions in Qualitative Research

Box 6.6 provides an overview of the approaches to flexible design research featured in this chapter (see also box 6.1, p. 165).

Case Studies

In case study, the *case* is the situation, individual, group, organization or whatever it is that we are interested in. Case study has been around for a long time (Hamel, 1993, traces its history within social science). To some it will suggest the legal system, to others the medical one. Bromley (1986) points out that case study can be found in areas as disparate as administration, anatomy, anthropology, artificial intelligence, biochemistry, business studies, clinical medicine, counselling, criminology, education, gerontology, history, industrial relations, jurisprudence, management, military studies, personality, politics, psychiatry, social work and sociology. Note that there is some danger in using a well-worn term like 'case study', for all such terms carry 'excess baggage' around with them, surplus meanings and resonances from these previous usages.

The intention here is to provide guidance in carrying out *rigorous* case studies. This involves attention to matters of design, data collection, analysis, interpretation and reporting which form a major part of later chapters. Before getting on with this, however, let us be clear as to what we mean by case study,