

CHAPTER 4

General design issues

This chapter:

- develops a framework for designing a real world study linking purpose, conceptual framework, research questions, methods and sampling strategy;
- sensitizes the reader to the issues involved in selecting a research strategy;
- introduces experimental and non-experimental fixed design strategies;
- suggests that flexible design strategies particularly appropriate for real world studies include case studies, ethnographic studies and grounded theory studies;
- covers a range of multi-strategy (mixed-method) designs;
- emphasizes that it is advisable to read the other chapters in Part II before making decisions about strategy; and
- concludes by considering the trustworthiness of research findings, and its relationship to research design.

Introduction

Design is concerned with turning research questions into projects. This is a crucial part of any research project, but it is often slid over quickly without any real consideration of the issues and possibilities. There is a strong tendency, both for those carrying out projects and those who want them carried out, to assume that there is no alternative to their favoured approach. Comments have already been made on the assumption by many psychologists that an experimental design is inevitably called for. For other social scientists, and for quite a few clients when commissioning studies, designs involving the statistical analysis of sample survey data are seen as the only possible approach.

As stressed in the previous chapter, the strategies and tactics you select in carrying out a piece of research depend very much on the type of research question you are trying to answer. Hakim (2000), in one of the few books which focuses on design issues across a range of social science disciplines, makes a comparison between designers of research projects and architects, and then goes on to extend this to suggest that those who actually carry out projects are like builders. For her:

Design deals primarily with aims, purposes, intentions and plans within the practical constraints of location, time, money and availability of staff. It is also very much about *style*, the architect's own preferences and ideas (whether innovative or solidly traditional) and the stylistic preferences of those who pay for the work and have to live with the final result (p. 1, emphasis in original).

In small-scale research, the architect-designer and builder-researcher are typically one and the same person. Hence the need for sensitivity to design issues, to avoid the research equivalent of the many awful houses put up by speculative builders without benefit of architectural expertise.

Such muddling through should be distinguished from the opportunity to develop and revise the original plan, which is easier in a small-scale project than in one requiring the coordination of many persons' efforts. Design modification is more feasible with some research strategies than with others – it is an integral part of what are referred to in this text as flexible designs. However, this kind of flexibility calls for a concern for design *throughout* the project, rather than providing an excuse for not considering design at all.

A framework for research design

Design, in the sense discussed above, concerns the various things which should be thought about and kept in mind when carrying out a research project. Many models have been put forward and Figure 4.1 is my attempt. The components are:

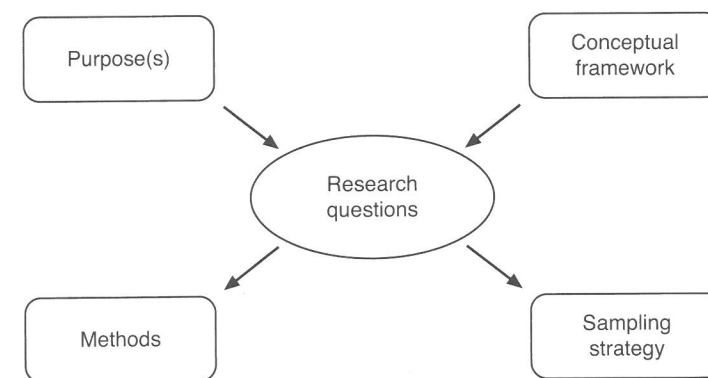


Figure 4.1: Framework for research design.

- *Purpose(s)*. What is this study trying to achieve? Why is it being done? Are you seeking to describe something, or to explain or understand something? Are you trying to assess the effectiveness of something? Is it in response to some problem or issue for which solutions are sought? Is it hoped to change something as a result of the study?
- *Conceptual framework*. Your theory about what is going on, of what is happening and why. What are the various aspects or features involved, and how might they be related to each other?
- *Research questions*. To what questions is the research geared to providing answers? What do you need to know to achieve the purpose(s) of the study? What is it feasible to ask given the time and resources that you have available?
- *Methods*. What specific techniques (e.g. semi-structured interviews, participant observation) will you use to collect data? How will the data be analysed? How do you show that the data are trustworthy?
- *Sampling procedures*. Who will you seek data from? Where and when? How do you balance the need to be selective with that of collecting the data needed?

Ethical considerations, though not included in the design framework, inevitably arise when carrying out research involving people and should be taken into account both in the planning and carrying out of your project (see Chapter 9).

All these aspects need to be interrelated and kept in balance. The diagram suggests that there is some directionality about the whole process. Both your purposes and the conceptual framework feed in to, and help you specify, the research questions. When you know something about the research questions you want to be answered, then you are able to make decisions about the methods and the procedures to be used when sampling. However, unless you are dealing with a fixed design which is tightly pre-specified, this should not be taken to imply a once only consideration of the different aspects.

In flexible designs there should be a repeated revisiting of *all* of the aspects as the research takes place. In other words, the detailed framework of the design *emerges* during the study. The various activities of collecting and analysing data, of refining and modifying the set of research questions, of developing theory; of changing the intended sample to follow up interesting lines or to seek answers to rather different questions and perhaps even reviewing the purposes of the study in the light of a changed context arising from the way in which the other aspects are developing – are likely to be going on together.

This might suggest that a better representation of the relationship between these aspects in flexible designs would show two-way arrows between each of the components in the figure. Maxwell (2005, p. 5) approximates to this in a very similar diagram which he refers to as an 'Interactive' model of research design. Or even that one might revert to what Martin (1981) has called the 'garbage can' model of research design where such components are 'swirling around in the garbage can or decision space of the particular research project' (Grady and Wallston, 1988, p. 12). However, providing the interactive nature of what goes on in this kind of project is understood, Figure 4.1 has the advantage of presenting a simple and logical structure.

The design framework should have high compatibility between purposes, research questions, conceptual framework and sampling strategy. Some mismatches call for serious attention. For example:

- If the only research questions to which you can think of ways to get answers to are not directly relevant to the purposes of the study, then something has to change. Probably the research questions.
- If the methods and/or the sampling strategy are not providing answers to the research questions, something should change. Collect additional data and/or change the data collection method(s), extend the sampling or cut down on or modify the research questions.
- If there are research questions which do not link to the conceptual framework, or parts of the conceptual framework which are not represented in the set of research questions, then one or other (or both) needs changing.

This is something of a counsel of perfection. Don't let it block any further progress if you can't get it quite right. You may not get an ideal solution with the time and resources you have available. Go for a practical solution that seems reasonably adequate (an example of the strategy of *satisficing* as advocated by Simon, 1979).

In fixed research designs you should get as much of this right as you can before embarking on the major phase of data collection. Hence the importance of pilot work, where you have the opportunity of testing out the feasibility of what you propose. In flexible research designs you have to get all of this sorted out by the *end* of the study. As Brewer and Hunter (2005, p. 45) put it 'Once a study is published, it is in many ways irrelevant whether the research problem prompted the study or instead emerged from it'. This is not a licence to rewrite history. In many qualitative research traditions there is an expectation that you provide an account of your journey, documenting the various changes made along the way. However, you are definitely not bound to some form of 'honour code' where, say, you declare your initial set of research questions and then stick to them through thick and thin. Your aim is to come up with a final set of research questions, which are relevant to the purposes of the study (which may, or may not, have been renegotiated along the way); and which show clear linkage to the conceptual structure (from whatever source it has obtained) and for which the sampling has been such that the data you have collected and analysed provides answers to those questions.

In the real world, of course, it won't be as neat and tidy as this. Some research questions may remain stubbornly unanswerable given the amount of sampling and data collection your resources permit. This is not a capital offence. Providing you have answers to some of the questions which remain on your agenda, then you have made worthwhile progress. And the experience will no doubt be salutary in helping you to carry out more realistically designed projects in the future. You could even claim this for a study where you ended up with no answers to relevant research questions, but this is not going to further your career as a researcher.

It may also be that you come up with unexpected findings which appear interesting and illuminative. These findings may well be assimilable into your framework by appropriate extension or modification of your research questions. There is nothing wrong with adding a further question providing it is relevant to your purposes and it can

be incorporated within a (possibly modified) theoretical framework. If your ingenuity fails, and you can't link it in, then simply regard this as a bonus to be reported under the heading of 'an interesting avenue for further research'.

Getting a feel for design issues

The shaded pages below give an overview of what is involved in choosing a research strategy, including a short description of the strategies you might consider.

This might be a good time for you to get hold of the reports of a range of published studies (journal articles, research reports, dissertations, etc.) and to read them through to get a feel for different designs. Try not to get bogged down in the details, and don't be put off by complex analyses. When you get on to the detailed design of your own study and its analysis, you can seek assistance on such matters. The obvious sources are academic and professional journals close to your own concerns but, as previously suggested, there is a lot to be said for 'spending some time in the next village'. If your field is, say, social work, browse through a few health-related, educational or management journals. The purpose here is not so much to build up knowledge of directly relevant literature, or to find something you can replicate, although both of these are reputable aims in their own right. It's the overall design that you are after. The website gives details of a mixed bag of studies with fixed, flexible and multi-strategy designs worth chasing up and looking through. Note that they won't necessarily use the terminology adopted here of research questions, purposes, etc. (it is instructive to try to work these out as an exercise).

◆ The website gives references to a selection of examples of research using fixed, flexible and multi-strategy designs.

If you follow up these examples you will notice that several of them involve evaluating some practice, intervention or programme, or have an action perspective where they are concerned with change of some kind taking place. Chapter 8 covers the additional features to be considered in studies which have these purposes.

Choosing a Research Design Strategy

This section seeks to sensitize you to the issues involved in choosing a research design strategy.

- A. *Is a FIXED, FLEXIBLE or MULTI-STRATEGY design strategy appropriate?*
- A *fixed design* calls for a tight pre-specification before you reach the main data collection stage. If you can't pre-specify the design, don't use the fixed approach. Data are almost always in the form of numbers; hence this type is commonly referred to as a *quantitative* strategy. See Chapter 5 for details.

- A *flexible design* evolves during data collection. Data are typically non-numerical (usually in the form of words); hence this type is often referred to as a *qualitative* strategy. See Chapter 6 for details.
- A *multi-strategy design* combines substantial elements of both fixed and flexible design. A common type has a flexible phase followed by a fixed phase (the reverse sequence is more rare). See Chapter 7 for details.

Note: Flexible designs can include the collection of small amounts of quantitative data (Chapter 6, p. 135) Similarly, fixed designs can include the collection of small amounts of qualitative data (Chapter 5, p. 81).

- B. *Is your proposed study an EVALUATION?*
Are you trying to establish the worth or value of something such as an intervention, innovation or service? This could be approached using either a fixed, flexible or multi-strategy design strategy depending on the specific purpose of the evaluation. If the focus is on *outcomes*, a *fixed design* is probably indicated, if it is on *processes*, a *flexible design* is probably preferred. Many evaluations have an interest in *both outcomes and processes* and use a *multi-strategy design*. See Chapter 8, p. 176, for details.
- C. *Do you wish to carry out ACTION RESEARCH?*
Is an action agenda central to your concerns? This typically involves direct participation in the research by others likely to be involved, coupled with an intention to initiate change. A *flexible design* is almost always used. See Chapter 8, p. 188, for details.
- D. *If you opt for a FIXED design strategy, which type is most appropriate?*
Two broad traditions are widely recognized; *experimental* and *non-experimental designs*. Box 4.1 on p. 78 summarizes their characteristics.
- E. *If you opt for a FLEXIBLE design strategy, which type is most appropriate?*
Flexible designs have developed from a wide range of very different traditions. Three of these are widely used in real world studies. These are *case studies*, *ethnographic studies* and *grounded theory* studies. Box 4.2 on p. 79 summarizes their characteristics.
- F. *If you are considering a MULTI-STRATEGY design strategy, which type is most appropriate?*

It may well be that a strategy which combines fixed and flexible design elements seems to be appropriate for the study with which you are involved. One or more case studies might be linked to an experiment. Alternatively, a small experiment might be incorporated actually within a case study. Issues involved in the carrying out of multi-strategy designs are discussed in Chapter 7.

Note: The research strategies discussed above by no means cover all possible real world research designs. They are more of a recognition of the camps into which researchers have tended to put themselves, signalling their preferences for certain ways of working. Such camps have the virtue of providing secure bases within which fledgling researchers can be inculcated in the ways of the tribe, and, more generally, high professional standards can be maintained. They carry the danger of research being 'strategy driven' in the

sense that someone skilled in, say, doing experiments assumes automatically that every problem has to be attacked through that strategy.

G. *The purpose(s) helps in selecting the strategy*

The strategies discussed above represent different ways of collecting and analysing empirical evidence. Each has its particular strengths and weaknesses. It is also commonly suggested that there is a hierarchical relationship between the different strategies, related to the purpose of the research; that

- flexible (qualitative) strategies are appropriate for exploratory work;
- non-experimental fixed strategies are appropriate for descriptive studies;
- experiments are appropriate for explanatory studies.

There is some truth in this assertion – certainly as a description of how the strategies have tended to be used in the past. There is a further sense in which a flexible strategy lends itself particularly well to exploration, a sense in which certain kinds of description can be readily achieved using non-experimental (typically survey approaches) and a traditional view that the experiment is a particularly appropriate tool for getting at cause and effect relationships (although see the discussion in Chapter 2, p. 32). However, these are not necessary or immutable linkages. *Each strategy (fixed, flexible or multi-strategy) can be used for any or all of the purposes.* For example, grounded theory studies aim to be explanatory through the development of theory; also there can be, and have been, exploratory, descriptive and explanatory case studies (Yin, 2003, 2009).

Real world studies are very commonly evaluations, i.e. their purpose is to assess the worth or value of something. A fixed, flexible or multi-strategy design may be appropriate depending on the specific focus of the evaluation (see B above).

If a purpose is to initiate change and/or to involve others, then an action research strategy may be appropriate. A flexible design is probably called for (see C above).

H. *The research questions have a strong influence on the strategy to be chosen*

While purpose is of help in selecting the research design strategy, the type of research questions you are asking is important. For example, questions asking ‘how many?’ or ‘how much?’ or ‘who’ or ‘where’ suggest the use of a non-experimental fixed strategy such as a survey. ‘What’ questions concerned with ‘what is going on here?’ lend themselves to some form of flexible design study. ‘How?’ and ‘why?’ questions are more difficult to pin down. They often indicate a flexible design. However, if the research can have control over events and if there is substantial prior knowledge about the problem and the likely mechanisms involved, then an experiment might be indicated.

Box 4.3 on p. 80 considers the research questions set out in Box 3.3, p. 60, and discusses research strategies that might be appropriate.

I. *Specific methods of investigation need not be tied to particular research strategies*

The methods or techniques used to collect information, what might be called the *tactics of enquiry*, such as questionnaires or various kinds of observation, are sometimes regarded as necessarily linked to particular research strategies.

Thus, in fixed non-experimental designs, surveys may be seen as being carried out by structured questionnaire and experiments through specialized forms of observation, often requiring the use of measuring instruments of some sophistication. In flexible designs, grounded theory studies were often viewed as interview-based and ethnographic studies seen as entirely based on participant observation.

However, this is not a tight or necessary linkage. For example, while participant observation is a central feature of the ethnographic approach, it can be augmented by interviews and documentary analysis. Similarly, there is no reason in principle for particular fixed design studies to be linked to specific data collection techniques. Non-experimental surveys could well be carried out using observation, the effect of an experiment assessed through questionnaire responses.

You should now have some appreciation of what is involved in selecting an appropriate research strategy. Before plunging in and making a decision, you need to know more about the issues involved in working within these strategies to help you get a feel for what might be involved. The rest of the chapters in Part II cover them in some detail.

Establishing trustworthiness

How do you persuade your audiences, including yourself, that the findings of your research are worth taking account of? What is it that makes the study believable and trustworthy? What are the kinds of argument that you can use? What questions should you ask? What criteria are involved?

In this connection *validity* and *generalizability* are central concepts. Validity is concerned with whether the findings are ‘really’ about what they appear to be about. Generalizability refers to the extent to which the findings of the enquiry are more generally applicable outside the specifics of the situation studied. These issues, together with the related one of *reliability* (the consistency or stability of a measure; for example, if it were to be repeated would the same result be obtained), were initially developed in the context of traditional fixed designs and there is considerable debate about their applicability to flexible designs. Hence trustworthiness is considered separately in each of the following chapters covering different types of research designs.

Further reading

◆ The website gives annotated references to further reading for Chapter 4.

BOX 4.1

Experimental and non-experimental fixed design research strategies**Experimental strategy**

The central feature is that the researcher actively and deliberately introduces some form of change in the situation, circumstances or experience of participants with a view to producing a resultant change in their behaviour.

In 'experiment-speak' this is referred to as measuring the effects of manipulating one variable on another variable. The details of the design are fully pre-specified before the main data collection begins (there is typically a 'pilot' phase before this when the feasibility of the design is checked and changes made if needed).

Typical features:

- selection of samples of individuals from known populations;
- allocation of samples to different experimental conditions;
- introduction of planned change on one or more variables;
- measurement on very small number of variables;
- control of other variables; and
- testing of formal hypotheses.

Non-experimental strategy

The overall approach is the same as in the experimental strategy but the researcher does not attempt to change the situation, circumstances or experience of the participants.

The details of the design are fully pre-specified before the main data collection begins (there is typically a 'pilot' phase before this when the feasibility of the design is checked and changes made if needed).

Typical features:

- selection of samples of individuals from known populations;
- allocation of samples to different experimental conditions;
- measurement on relatively small number of variables;
- control of other variables; and
- may or may not involve hypothesis testing.

BOX 4.2

Three widely used flexible design research strategies**Case study**

Development of detailed, intensive knowledge about a single 'case', or of a small number of related 'cases'.

The details of the design typically 'emerge' during data collection and analysis.

Typical features:

- selection of a single case (or a small number of related cases) of a situation, individual or group of interest or concern;
- study of the case in its context; and
- collection of information via a range of data collection techniques including observation, interview and documentary analysis (typically, though not necessarily exclusively, producing qualitative data).

Ethnographic study

Seeks to capture, interpret and explain how a group, organization or community live, experience and make sense of their lives and their world.

It typically tries to answer questions about specific groups of people, or about specific aspects of the life of a particular group.

Typical features:

- selection of a group, organization or community of interest or concern;
- immersion of the researcher in that setting; and
- use of participant observation.

Grounded theory study

The central aim is to generate theory from data collected during the study.

Particularly useful in new, applied areas where there is a lack of theory and concepts to describe and explain what is going on. Data collection, analysis and theory development and testing interspersed throughout the study.

Typical features:

- applicable to a wide variety of phenomena;
- commonly interview-based; and is
- a systematic but flexible research strategy which provides detailed prescriptions for data analysis and theory generation.

Notes: There are many other types of flexible design, some of which are summarized in Chapter 6. Many studies involving flexible designs focus on a particular 'case' in its context and can be conceptualized as case studies. Case studies can follow an ethnographic or grounded theory approach, but don't have to.

BOX 4.3

Linking research questions to research strategy

Consider the research questions discussed in Box 3.3 (p. 60):

1. Do the children read better as a result of this programme?
or
2. Do the children read better in this programme compared with the standard programme?
or
3. For what type of special need, ability level, class organization or school is the programme effective?

If the interest is in quantitative outcome measures, and it is feasible to exert some degree of control over the situation (e.g. setting up different groups of children for the innovatory and standard programmes), these questions could be approached using an *experimental strategy*. If random allocation is used, this becomes a *true experiment*; if not, a *quasi-experiment*.

If this control were not feasible, or not desired, but quantitative data were still sought, a *non-experimental fixed design* is possible.

If there is a broader notion of what is meant by 'reading better' or of an 'effective' programme than that captured by a small number of quantitative variables, some type of *flexible strategy* is called for. This is likely to be a multimethod *case study*, and could also be *ethnographic* or *grounded theory* in style.

A *multi-strategy* approach where the *case study* could incorporate, say, an *experimental* component, could be considered.

4. What is the experience of children following the programme?
5. What are teachers' views about the programme?
and/or
6. To what extent are parents involved in and supportive of the programme?

These questions could be approached using any of the flexible strategies; though (4) might particularly indicate an *ethnographic* approach.

Questions (5) and (6) could, alternatively or additionally, follow a *non-experimental fixed design* if quantitative data are sought.

The overall message is that, while the research questions help in deciding research strategy, much is still dependent on your own preferences and on the type of design and data which are going to speak most strongly to the stakeholders.

CHAPTER 5

Fixed designs

This chapter:

- covers general features of fixed design research, typically involving the collection of quantitative data;
- discusses how the trustworthiness (including reliability, validity and generalizability) of findings from this style of research can be established;
- explores the attractions and problems of doing experiments in real world research;
- gives particular attention to the randomized controlled trial (RCT) and whether it can be legitimately viewed as the 'gold standard' of research designs;
- attempts to provide a balanced view of the ubiquitous evidence-based movement;
- differentiates between true experimental, quasi-experimental and single case experimental designs;
- considers non-experimental fixed designs; and
- concludes by discussing how to decide on sample sizes in fixed design research.

Introduction

This chapter deals with approaches to social research where the design of the study is fixed before the main stage of data collection takes place. In these approaches the phenomena of interest are typically quantified. This is not a necessary feature. As pointed out by Oakley (2000, p. 306) there is nothing intrinsic to such designs which rules out qualitative methods or data (see Murphy *et al.*, 1998, for examples of purely qualitative fixed design studies, and of others using both qualitative and quantitative methods, in the field of health promotion evaluation).

It has already been argued in Chapter 3 that there can be considerable advantage in linking research to theory. With fixed designs, that link is straightforward: *fixed designs are theory-driven*. The only way in which we can, as a fixed design requires, specify in advance the variables to be included in our study and the exact procedures to be followed, is by having a reasonably well-articulated theory of the phenomenon we are researching. Put in other terms, we must already have a substantial amount of conceptual understanding about a phenomenon before it is worthwhile following the risky strategy of investing precious time and resources into such designs. This may be in the form of a model, perhaps represented pictorially as a *conceptual framework* as discussed in Chapter 3.

Such models help to make clear the multiple and complex causality of most things studied in social research. Hard thinking to establish this kind of model before data collection is invaluable. It suggests the variables we should target: those to be manipulated or controlled in an experiment and those to be included in non-experimental studies.

In realist terms, this means that you have a pretty clear idea about the mechanisms likely to be in operation and the specific contexts in which they will, or will not, operate. You should also know what kind of results you are going to get, and how you will analyse them, before you collect the data. If the study does deliver the expected relationships, it provides support for the existence of these mechanisms and their actual operation in this study. This does not preclude your following up interesting or unexpected patterns in the data. They may suggest the existence of other mechanisms which you had not thought of.

Large-scale studies can afford to draw the net relatively wide. Large numbers of participants can be involved: several subgroups established, perhaps a range of different contexts covered, more possible mechanisms tested out. For the small-scale studies on which this text focuses, and in real world settings where relevant previous work may be sparse or non-existent, there is much to be said for a *multi-strategy design* (see Chapter 7) with an initial flexible design stage which is primarily exploratory in purpose. This seeks to establish, both from discussions with professionals, participants and others involved in the initial phase, and from the empirical data gathered, likely 'bankers' for mechanisms operating in the situation, contexts where they are likely to operate and the characteristics of participants best targeted. The second fixed design phase then incorporates a highly focused survey, experiment or other fixed design study.

Even with a preceding exploratory phase, *fixed designs should always be piloted*. You carry out a mini-version of the study before committing yourself to the big one. This is, in part, so you can sort out technical matters to do with methods of data collection to ensure that, say, the questions in a questionnaire are understandable and unambiguous. Just as importantly, it gives you a chance to ensure you are on the right lines conceptually. Have you 'captured' the phenomenon sufficiently well for meaningful data to be collected? Do you really have a good grasp of the relevant mechanisms and contexts? This is an opportunity to revise the design: to sharpen up the theoretical framework; develop the research questions; rethink the sampling strategy. And perhaps to do a further pilot.

Also, while the central part of what you are going to do with your data should be thought through in advance, i.e. you are primarily engaged in a *confirmatory* task in

fixed designs, there is nothing to stop you also carrying out *exploratory data analysis* (see Chapter 16, p. 419). It may be that there are unexpected patterns or relationships which reveal inadequacies in your initial understanding of the phenomenon. You cannot expect to confirm these revised understandings in the same study but they may well provide an important breakthrough suggesting a basis for further research.

This chapter seeks to provide a realist-influenced view of fixed design research. There is coverage of *true experimental*, *single-case experimental*, *quasi-experimental* and *non-experimental fixed designs*. The differences between these types of design are brought out and some examples given. In the 'true' experiment, two or more groups are set up, with random allocation of people to the groups. The experimenter then actively manipulates the situation so that different groups get different treatments. Single-case design,¹ as the name suggests, focuses on individuals rather than groups and effectively seeks to use persons as their own control, with their being subjected to different experimentally manipulated conditions at different times. Quasi-experiments lack the random allocation to different conditions found in true experiments. Non-experimental fixed designs do not involve active manipulation of the situation by the researcher. However, the different fixed designs are similar in many respects, as discussed in the following section.

General features of fixed designs

Fixed designs are usually concerned with aggregates: with group properties and with general tendencies. In traditional experiments, results are reported in terms of group averages rather than what individuals have done. Because of this, there is a danger of the *ecological fallacy* – that is of assuming that inferences can be made about individuals from such aggregate data (Connolly, 2006). Single-case experimental designs are an interesting exception to this rule. Most non-experimental fixed research also deals with averages and proportions. The relative weakness of fixed designs is that they cannot capture the subtleties and complexities of individual human behaviour. Even single-case designs are limited to quantitative measures of a single simple behaviour or, at most, a small number of such behaviours. The advantage of fixed designs is in being able to transcend individual differences and identify patterns and processes which can be linked to social structures and group or organizational features.

Fixed designs traditionally assume a 'detached' researcher to guard against the researcher having an effect on the findings of the research. Researchers typically remain

¹Single-case fixed designs are, typically, very different from *case studies*. The latter are almost always of flexible design using several data collection methods (see Chapter 6, p. 135). However, it would be feasible to have a multi-strategy design which incorporated a single-case fixed design element within a case study.

at a greater physical and emotional distance from the study than those using flexible designs. In experimental research, the *experimenter effect* is well known. It is now widely acknowledged that the beliefs, values and expectations of the researcher can influence the research process at virtually all of its stages (Rosenthal, 1976, 2003; Rosnow and Rosenthal, 1997). Hence the stance now taken is that all potential biases should be brought out into the open by the researcher and every effort made to counter them.

There are often long periods of preparation and design preliminaries before data collection and a substantial period of analysis after data collection. This does not, of course, in any way absolve the researcher from familiarity with the topic of the research, which is typically acquired vicariously from others, or from a familiarity with the literature, or from an earlier, possibly qualitative, study. There will be involvement during the data collection phase, but with some studies such as postal surveys this may be minimal. Your personal preference for a relatively detached, or a more involved, style of carrying out research is a factor to take into account when deciding the focus of your research project and the selection of a fixed or flexible design.

It has been fashionable in some academic and professional circles to denigrate the contribution of quantitative social research. As Bentz and Shapiro (1998) comment, in a text primarily covering qualitative approaches:

There is currently an antiquantitative vogue in some quarters, asserting or implying that quantitative research is necessarily alienating, positivistic, dehumanizing, and not 'spiritual'. In fact, it is clear that using quantitative methods to identify causes of human and social problems and suffering can be of immense practical, human, and emancipatory significance, and they are not necessarily positivistic in orientation. For example, quantitative methods are currently being used in the analysis of statistics to help identify the principal causes of rape. Baron and Straus have analyzed police records on rape quantitatively to look at the relative roles of gender inequality, pornography, gender cultural norms about violence, and social disorganization in causing rape (1989). Clearly knowing the relative contribution of these factors in causing rape would be of great significance for social policy, economic policy, the law, socialization, and the criminal justice system, and it is difficult to see how one would arrive at compelling conclusions about this without quantitative analysis (p. 124).

They also point out that quantitative and experimental methods have been used to understand social problems and criticize prevailing ideologies in a way which contributes to social change and the alleviation of human suffering (i.e. for emancipatory purposes as discussed in Chapter 2, p. 39).

Oakley (2000) suggests that this antipathy to quantitative, and in particular experimental, research derives in part from the influence of feminist methodologists who have viewed quantitative research as a masculine enterprise, contrasting it with qualitative research which is seen as embodying feminine values. She rejects this stereotyping and in her own work has made the transition from being a qualitative researcher to a staunch advocate of true randomized experiments.

Establishing trustworthiness in fixed design research

This is to a considerable extent a matter of common sense. Have you done a good, thorough and honest job? Have you tried to explore, describe or explain in an open and unbiased way? Or are you more concerned with delivering the required answer or selecting the evidence to support a case? If you can't answer these questions with yes, yes and no, respectively, then your findings are essentially worthless in research terms. However, pure intentions do not guarantee trustworthy findings. You persuade others by clear, well-written and presented, logically argued accounts which address the questions that concern them. These are all issues to which we will return in Chapter 18 on reporting.

This is not simply a presentational matter, however. Fundamental issues about the research itself are involved. Two key ones are *validity* and *generalizability*. *Validity*, from a realist perspective, refers to the accuracy of a result. Does it capture the real state of affairs? Are any relationships established in the findings true, or due to the effect of something else? *Generalizability* refers to the extent to which the findings of the research are more generally applicable, for example in other contexts, situations or times, or to persons other than those directly involved.

Validity

Suppose that we have been asked to carry out some form of research study to address the research question:

Is educational achievement in primary schools improved by the introduction of standard assessment tests at the age of seven?

Leave on one side issues about whether or not this is a sensible question and about the most appropriate way to approach it. Suppose that the findings of the research indicated a 'yes' answer – possibly qualified in various ways. In other words, we measure educational achievement, and it appears to increase following the introduction of the tests. Is this relationship what it appears to be – is there a real, direct, link between the two things?

Central to the scientific approach is a degree of scepticism about our findings and their meaning (and even greater scepticism about other people's). Can we have been fooled so that we are mistaken about them? Unfortunately, yes – there is a wide range of possibilities for confusion and error.

Reliability

Some problems come under the heading of reliability. This is the stability or consistency with which we measure something. For example, consider how we are going to assess

educational achievement. This is no easy task. Possible contenders, each with their own problems, might include:

- a formal 'achievement test' administered at the end of the primary stage of schooling;
- teachers' ratings, also at the end of the primary stage; or
- the number, level and standard of qualifications gained throughout life.

Let's say we go for the first. It is not difficult to devise something which will generate a score for each pupil. However, this might be unreliable in the sense that if a pupil had, say, taken it on a Monday rather than a Wednesday, she would have got a somewhat different score. There are logical problems in assessing this, which can be attacked in various ways (e.g. by having parallel forms of the test which can be taken at different times, and their results compared). These are important considerations in test construction – see Chapter 12 for further details.

Unless a measure is reliable, it cannot be valid. However, while reliability is necessary, it is not sufficient. A test for which all pupils always got full marks would be totally consistent but would be useless as a way of discriminating between the achievements of different pupils (there could of course be good educational reasons for such a test if what was important was mastery of some material).

Unreliability may have various causes, including:

Participant error

In our example the pupil's performance might fluctuate widely from occasion to occasion on a more or less random basis. Tiredness due to late nights could produce changes for different times of the day, pre-menstrual tension monthly effects or hay fever seasonal ones. There are tactics which can be used to ensure that these kinds of fluctuations do not bias the findings, particularly when specific sources of error can be anticipated (e.g. keep testing away from the hay fever season).

Participant bias

This is more problematic from a validity point of view. It could be that pupils might seek to please or help their teacher, knowing the importance of 'good results' for the teacher and for the school, by making a particularly strong effort at the test. Or for disaffected pupils to do the reverse. Here it would be very difficult to disentangle whether this was simply a short-term effect which had artificially affected the test scores, or a more long-lasting side-effect of a testing-oriented primary school educational system. Consideration of potential errors of these kinds is part of the standard approach to experimental design.

Observer error

This would be most obvious if the second approach, making use of teachers' ratings as the measure of pupil achievement, had been selected. These could also lead to more or

less random errors if, for example, teachers made the ratings at a time when they were tired or overstretched and did the task in a cursory way. Again, there are pretty obvious remedies (perhaps involving the provision of additional resources).

Observer bias

This is also possible and, like participant bias, causes problems in interpretation. It could be that teachers in making the ratings were, consciously or unconsciously, biasing the ratings they gave in line with their ideological commitment either in favour of or against the use of standard assessment tests. This is also a well-worked area methodologically, with procedures including 'blind' assessment (the ratings being made by someone in ignorance of whether the pupil had been involved in standard assessment tests) and the use of two independent assessors (so that inter-observer agreements could be computed). Further details are given in Chapter 13, p. 341.

Types of validity

If you have made a serious attempt to get rid of participant and observer biases and have demonstrated the reliability of whatever measure you have decided on, you will be making a pretty good job of measuring something. The issue then becomes – does it measure what you think it measures? In the jargon – does it have *construct validity*?

There is no easy, single, way of determining construct validity. At its simplest, one might look for what seems reasonable, sometimes referred to as *face validity*. An alternative looks at possible links between scores on a test and the third suggested measure – the pupils' actual educational achievement in their later life (i.e. how well does it predict performance on the criterion in question, or *predictive criterion validity*). These and other aspects of construct validity are central to the methodology of testing.

The complexities of determining construct validity can lead to an unhealthy concentration on this aspect of carrying out a research project. For many studies there is an intuitive reasonableness to assertions that a certain approach provides an appropriate measure. Any one way of measuring or gathering data is likely to have its shortcomings, which suggests the use of multiple methods of data collection. One could use all three of the approaches to assessing educational achievement discussed above (achievement tests, teachers' ratings and 'certificate counting') rather than relying on any one measure. This is one form of *triangulation* – see Chapter 6, p. 158. Similar patterns of findings from very different methods of gathering data increase confidence in the validity of the findings. Discrepancies between them can be revealing in their own right. It is important to realize, however, that multiple methods do not constitute a panacea for all methodological ills. They raise their own theoretical problems; and they may in many cases be so resource-hungry as to be impracticable (see Chapter 14, p. 385).

Let us say that we have jumped the preceding hurdle and have demonstrated satisfactorily that we have a valid measure of educational achievement. However, a finding that achievement increases *after* the introduction of the tests does not necessarily mean that it increased *because* of the tests. This gets us back to the consideration of causation which occupied us in Chapter 2 (see p. 32).

What we would like to do is to find out whether the treatment (introduction of the tests) actually caused the outcome (the increase in achievement). If a study can plausibly demonstrate this causal relationship between treatment and outcome, it is referred to as having *internal validity*. This term was introduced by Campbell and Stanley (1963), who provided an influential and widely used analysis of possible 'threats' to internal validity.

These threats are other things that might happen which confuse the issue and make us mistakenly conclude that the treatment caused the outcome (or obscure possible relationships between them). Suppose, for example, that the teachers of the primary school children involved in the study are in an industrial dispute with their employers at the same time that testing is introduced. One might well find, in those circumstances, a decrease in achievement related to the disaffection and disruption caused by the dispute, which might be mistakenly ascribed to the introduction of tests per se. This particular threat is labelled as 'history' by Campbell and Stanley – something which happens at the same time as the treatment. There is the complicating factor here that a case might be made for negative effects on teaching being an integral part of the introduction of formal testing into a child-centred primary school culture, i.e. that they are part of the treatment rather than an extraneous factor. However, for simplicity's sake, let's say that the industrial dispute was an entirely separate matter.

Campbell and Stanley (1963) suggested eight possible threats to internal validity which might be posed by other extraneous variables. Cook and Campbell (1979) have developed and extended this analysis, adding a further four threats. All 12 are listed in Box 5.1 (Onwuegbuzie and McLean, 2003, expand this list to 22 threats at the research design and data collection stage, with additional threats present at the data analysis and interpretation stages). The labels used for the threats are not to be interpreted too literally – mortality doesn't necessarily refer to the death of a participant during the study (though it might). Not all threats are present for all designs. For example, the 'testing' threat is only there if a pre-test is given, and in some cases, its likelihood, or perhaps evidence that you had gained from pilot work that a 'testing' effect was present, would cause you to avoid a design involving this feature.

BOX 5.1

Threats to internal validity

1. *History*. Things that have changed in the participants' environments other than those forming a direct part of the enquiry (e.g. occurrence of major air disaster during study of effectiveness of desensitization programme on persons with fear of air travel).
2. *Testing*. Changes occurring as a result of practice and experience gained by participants on any pre-tests (e.g. asking opinions about factory farming of animals before some intervention may lead respondents to think about the issues and develop more negative attitudes).

3. *Instrumentation*. Some aspect(s) of the way participants were measured changed between pre-test and post-test (e.g. raters in observational study using a wider or narrower definition of a particular behaviour as they get more familiar with the situation).
4. *Regression*. If participants are chosen because they are unusual or atypical (e.g. high scorers), later testing will tend to give less unusual scores ('regression to the mean'); e.g. an intervention programme with pupils with learning difficulties where 10 highest-scoring pupils in a special unit are matched with 10 of the lowest-scoring pupils in a mainstream school – regression effects will tend to show the former performing relatively worse on a subsequent test; see further details on p. 113.
5. *Mortality*. Participants dropping out of the study (e.g. in a study of an adult literacy programme – selective drop-out of those who are making little progress).
6. *Maturation*. Growth, change or development in participants unrelated to the treatment in the enquiry (e.g. evaluating extended athletics training programme with teenagers – intervening changes in height, weight and general maturity).
7. *Selection*. Initial differences between groups prior to involvement in the enquiry (e.g. through use of arbitrary non-random rule to produce two groups: ensures they differ in one respect which may correlate with others).
8. *Selection by maturation interaction*. Predisposition of groups to grow apart (or together if initially different); e.g. use of groups of boys and girls initially matched on physical strength in a study of a fitness programme.
9. *Ambiguity about causal direction*. Does A cause B, or B cause A? (e.g. in any correlational study, unless it is known that A precedes B, or vice versa – or some other logical analysis is possible).
10. *Diffusion of treatments*. When one group learns information or otherwise inadvertently receives aspects of a treatment intended only for a second group (e.g. in a quasi-experimental study of two classes in the same school).
11. *Compensatory equalization of treatments*. If one group receives 'special' treatment there will be organizational and other pressures for a control group to receive it (e.g. nurses in a hospital study may improve the treatment of a control group on grounds of fairness).
12. *Compensatory rivalry*. As above but an effect on the participants themselves (referred to as the 'John Henry' effect after the steel worker who killed himself through over-exertion to prove his superiority to the new steam drill); e.g. when a group in an organization sees itself under threat from a planned change in another part of the organization and improves performance.

(after Cook and Campbell, 1979, pp. 51–5)

In general design terms, there are two strategies to deal with these threats. If you know what the threat is, you can take specific steps to deal with it. For example, the use of comparison groups who have the treatment at different times or places will help to neutralize the 'history' threat. This approach of designing to deal with specific threats calls for a lot of forethought and is helped by knowledge and experience of the situation that you are dealing with. However, you can only hope to deal with a fairly small number of predefined and articulated threats in this way. In flexible design research it is feasible to address such threats to validity after the research has begun, as discussed in the following chapter.

The alternative strategy, central to the design philosophy of true experiments as developed by Fisher (1935, 1960), is to use randomization, which helps offset the effect of a myriad of unforeseen factors.

While true experiments are therefore effective at dealing with these threats, they are by no means totally immune to them. The threats have to be taken very seriously with quasi-experimental designs, and non-experimental fixed designs, and a study of the plausibility of the existence of various threats provides a very useful tool in interpretation. The interpretability of designs in the face of these threats depends not only on the design itself but also on the specific pattern of results obtained.

If you rule out these threats, you have established internal validity. You will have shown (or, more strictly, demonstrated the plausibility) that a particular treatment caused a certain outcome. Note however, that while an experiment can be effective in doing this, it tells you nothing about the actual mechanisms by which it did so, except insofar as you have anticipated possible alternative mechanisms and controlled for them in your design. As Shadish, Cook and Campbell (2002) put it:

The unique strength of experimentation is in describing the consequences attributable to deliberately varying a treatment. We call this *causal description*. In contrast, experiments do less well in clarifying the mechanisms through which and the conditions under which that causal relationship holds – what we call causal explanation (p. 9, emphases in original).

This limitation of experiments is central to Pawson and Tilley's (1997) critique of randomized controlled trials (RCTs) discussed later in the chapter (p. 100).

It is important to appreciate that 'validity threats are made implausible by *evidence*, not methods; methods are only a way of getting evidence that can help you rule out these threats' (Maxwell, 2005, p. 105, emphasis in original). The view that methods themselves can guarantee validity is characteristic of the discredited positivist approach and is itself untenable. Whatever method is adopted there is no such guarantee. The realist assumption is that all methods are fallible: 'a realist conception of validity . . . sees the validity of an account as inherent, not in the procedures used to produce and validate it, but in its relationship to those things that it is intended to be an account of' (Maxwell, 1992, p. 281, emphasis in original). See also House (1991).

The whole 'threat' approach sits well with a realist analysis, which is not surprising as Campbell was an avowed realist (see, however, House, Mathison and McTaggart, 1989, which makes a case for his approach, particularly in Cook and Campbell, 1979, as being essentially eclectic, taking aspects from a whole range of theoretical positions).

These threats tend to be only discussed in relation to experimental and quasi-experimental designs. However, validity is an important issue for all types of fixed designs and Onwuegbuzie and McLean (2003) have expanded Campbell and Stanley's framework for use with non-experimental fixed designs.

Generalizability

Sometimes one is interested in a specific finding in its own right. You may have shown, say, that a new group workshop approach leads, via a mechanism of increases in self-esteem, to subsequent maintained weight loss in obese teenagers at a residential unit. This may be the main thing that you are after if you are only concerned with whether or not the approach works with that specific group of individuals at the unit.

If, however, you are interested in what would happen with other client groups or in other settings, or with these teenagers when they return home, then you need to concern yourself with the *generalizability* of the study. Campbell and Stanley (1963) used the alternative term 'external validity'. Both this and generalizability are in common use. Internal and external validity tend to be inversely related in the sense that the various controls imposed in order to bolster internal validity often fight against generalizability. In particular, the fact that the laboratory is the controlled environment *par excellence* makes results obtained there very difficult to generalize to any settings other than close approximations to laboratory conditions. This aspect is sometimes referred to as a lack of *ecological validity*, i.e. findings from laboratory research may not be relevant to real world situations.

If your teenagers are a representative sample from a known population, then the generalization to that population can be done according to rules of statistical inference (note, however, that experimenters rarely take this requirement seriously). Generalizability to other settings or to other client groups has to be done on other, non-statistical, bases. LeCompte and Goetz (1982) have provided a classification of threats to external validity similar to that given for internal validity, which is listed in Box 5.2.

BOX 5.2

Threats to generalizability (external validity)

1. *Selection*. Findings being specific to the group studied.
2. *Setting*. Findings being specific to, or dependent on, the particular context in which the study took place.
3. *History*. Specific and unique historical experiences may determine or affect the findings.
4. *Construct effects*. The particular constructs studied may be specific to the group studied.

(after LeCompte and Goetz, 1982)

There are two general strategies for showing that these potential threats are discountable: *direct demonstration* and *making a case*. Direct demonstration involves you, or someone else who wishes to apply or extend your results, carrying out a further study involving some other type of participant, or in a different setting, etc. Making a case is more concerned with persuading that it is reasonable for the results to generalize, with arguments that the group studied, or setting, or period is representative (i.e. it shares certain essential characteristics with other groups, settings or periods and hence that the same mechanism is likely to apply in those also). This sorting out of the wheat of what is central to your findings from the chaff of specific irrelevancies can be otherwise expressed as having a theory or conceptual framework to explain what is going on.

Such a theory or conceptual framework may be expressed in formal and explicit terms by the presenter of the findings as discussed in Chapter 3 (p. 67). A study may be repeated with a different target group or in a deliberately different setting to assess the generalizability of its findings. There is a strong case, particularly with important or controversial findings, for attempting a replication of the original study. While in practice no replication is ever exact, an attempt to repeat the study as closely as possible which reproduces the main findings of the first study is the practical test of the reliability of your findings. Whether it is worthwhile to devote scarce resources to replication depends on circumstances. Replication is nowhere near as common as it should be in social research. In consequence, we may well be seeking to build on very shaky foundations. The argument is sometimes put that as validity depends on reliability then we should simply worry about the validity; if we can show that validity is acceptable then, necessarily, so is reliability. The problem here is that it becomes more difficult to disentangle what lies behind poor validity. It might have been that the findings were not reliable in the first place.

It is easy to guarantee unreliability. Carelessness, casualness and lack of commitment on the part of the researcher help, as does a corresponding lack of involvement by participants. Reliability is essentially a quality control issue. Punctilious attention to detail, perseverance and pride in doing a good job are all very important, but organization is the key.

While validity and generalizability are probably the central elements in establishing the value and trustworthiness of a fixed design enquiry, there are other aspects to which attention should be given. They include, in particular, objectivity and credibility.

Objectivity

The traditional, scientific approach to the problem of establishing objectivity is exemplified by the experimental approach. The solution here is seen to be to distance the experimenter from the experimental participant, so that any interaction that takes place between the two is formalized – indeed, some experimenters go so far as not only to have a standardized verbatim script but even to have it delivered via a tape-recorder.

To some, this artificiality is lethal for any real understanding of phenomena involving people in social settings. An alternative is to erect an objective/subjective contrast. ‘Objective’ is taken to refer to what multiple observers agree to as a phenomenon, in contrast to the subjective experience of the single individual. In other words, the criterion for objectivity is intersubjective agreement. This stance tends to go along with an

involved rather than a detached investigator, and notions of ‘triangulation’ (see Chapter 6, p. 158) where the various accounts of participants with different roles in the situation are obtained by investigators who, by combining them with their own perceptions and understandings, reach an agreed and negotiated account.

Formulated in terms of threats, objectivity can be seen to be at risk from a methodology where the values, interests and prejudices of the enquirer distort the response (experiment being for some the answer, and for others an extreme version of the problem). Relying exclusively on data from a single individual can similarly threaten objectivity. And again, a project carried out for an ideological purpose other than that of research itself clearly threatens objectivity.

Credibility

Shipman (1997) has suggested that we should go beyond the traditional concerns for reliability, validity and generalizability when considering the trustworthiness of research and also ask whether there is sufficient detail on the way the evidence is produced for the credibility of the research to be assessed. We cannot satisfy ourselves about the other concerns unless the researcher provides detailed information on the methods used and the justification for their use. This is a responsibility which has always been accepted by those using experimentation. The report of an experiment in a journal article carries an explicit requirement that sufficient detail must be given about procedures, equipment, etc. for the reader to be able to carry out an exact replication of the study.

This kind of requirement may be rejected as scientific by some practitioners using flexible designs, relying largely on qualitative data. However, it could be argued that there is a strong case for such research calling for an even greater emphasis on explaining the methods used and the warrant for the conclusions reached, because of the lack of codification of the methods of data collection or of approaches to analysis. This need is increasingly recognized in the design of qualitative research (e.g. Marshall and Rossman, 2006). However, there is considerable debate about the applicability of concepts such as reliability and validity, and the possibility and appropriateness of objectivity, when assessing the trustworthiness of flexible qualitative research. The following chapter pays considerable attention to this issue.

Experimental fixed designs

If, following your reading of the previous chapter, it appears possible that an experimental fixed design may be appropriate for your project and its research questions, then perusal of this section should help in choosing a specific experimental design. However, before confirming that choice, it will be necessary to read the chapters in Part III of this book to help select appropriate methods of collecting data, and Chapter 16 to establish how you will analyse the data after it has been collected.

To 'experiment', or to 'carry out an experiment' can mean many things. In very general terms, to be experimental is simply to be concerned with trying new things – and seeing what happens, what the reception is. Think of 'experimental' theatre, or an 'experimental' car, or an 'experimental' introduction of a mini-roundabout at a road junction. There is a change in something, and a concern for the effects that this change might have on something else.

However, when experimentation is contrasted with the other research designs, a stricter definition is employed, usually involving the control and active manipulation of variables by the experimenter.

Experimentation is a research strategy involving:

- the assignment of participants to different conditions;
- manipulation of one or more variables (called 'independent variables', IVs) by the experimenter;
- the measurement of the effects of this manipulation on one or more other variables (called 'dependent variable's, DVs); and
- the control of all other variables.

Note the use of the term *variable*. This is widespread within the experimental strategy and simply denotes something which can vary. However, it carries within it the notion that there are certain specific aspects which can be isolated and which retain the same meaning throughout the study.

The experimental strategy is a prime example of a fixed research design. You need to know exactly what you are going to do before you do it. It is a precise tool that can only map a very restricted range. A great deal of preparatory work is needed (either by you or someone else) if it is going to be useful. An experiment is an extremely *focused* study. You can only handle a very few variables, often only a single independent variable and a single dependent variable. These variables have to be selected with extreme care. You need to have a well-developed theory or conceptual framework. The major problem in doing experiments in the real world is that you often only have, at best, a pretty shaky and undeveloped theory; you don't know enough about the thing you are studying for this selectivity of focus to be a sensible strategy. This need to know what you are doing before you do it is a general characteristic of fixed research designs, but experiments are most demanding in this respect because of their extreme selectivity.

Laboratory experiments

Real world research seeks to address social problems and issues of current concern and to find ways of addressing such problems. Experiments typically take place in special places known as laboratories. In principle, just as certain kinds of academic research can be carried out in real world settings, which anthropologists and other social scientists refer to as 'field' settings, so research with a real world problem-solving focus might be carried out in a laboratory.

However, the necessary artificiality of laboratories can limit their value. Aronson, Brewer and Carlsmith (1985) have distinguished two senses in which laboratory

experimentation may lack realism (incidentally, nothing to do with realist philosophy). One is *experimental realism*. In this sense an experiment is realistic if the situation which it presents to the participant is realistic, if it really involves the participants (then referred to as 'subjects'), and has impact upon them. In the well-known Asch (1956) experiment on conformity, subjects made what seemed to them to be straightforward judgements about the relative length of lines. These judgements were contradicted by others in the room whom they took also to be subjects in the experiment. This study showed experimental realism in the sense that subjects were undergoing an experience which caused them to show strong signs of tension and anxiety. They appeared to be reacting to the situation in the same realistic kind of way that they would outside the laboratory.

However, it might be argued that the Asch study lacks what Aronson *et al.* term *mundane realism* (see also Aronson, Wilson and Akert, 2007). That is, the subjects were encountering events in the laboratory setting which were very unlikely to occur in the real world. Asch, following a common strategy in laboratory experimentation, had set up a very clearly and simply structured situation to observe the effects of group pressure on individuals. The real life counterpart, if one could be found, would be more complex and ambiguous, and in all probability would result in findings which were less conclusive. (The ethics of Asch's study are a different matter – see Chapter 9.)

Notwithstanding worries about the realism of laboratory-based studies, they remain popular with researchers, including those with real world concerns. After a review of the two approaches, Levitt and List (2006) conclude that 'the sharp dichotomy sometimes drawn between lab experiments and data generated in natural settings is a false one. Each approach has strengths and weaknesses, and a combination of the two is likely to provide deeper insights than either in isolation' (p. i).

Bias in experiments

Simplification of the situation, which is central to the experimental approach, may lead to clear results, but it does not protect against bias in them. The effects of two types of bias have been investigated in some detail. These are the *demand characteristics* of the experimental situation, and *experimenter expectancy* effects. In a very general sense, these are the consequences of the participants and the experimenters being human beings.

Bias due to demand characteristics occurs because participants know that they are in an experimental situation, know that they are being observed, know that certain things are expected or demanded of them (Orne, 1962; Strohmets, 2008). Hence the way in which they respond is some complex amalgam of the experimental manipulation and their interpretation of what effect the manipulation is supposed to have on them. Their action based on that interpretation is likely to be cooperative but could well be obstructive. Even in situations where participants are explicitly told that there are no right or wrong answers, that one response is as valued as another, participants are likely to feel that certain responses show themselves in a better light than others. There is evidence that persons who volunteer for experiments are more sensitive to these effects than those who are required to be involved (Rosenthal and Rosnow, 1975; Rosnow, 1993). However, Berkowitz and Troccoli (1986) are not persuaded of the widespread existence of biasing effects from demand characteristics.

The classic ploy to counteract this type of bias is deception by the experimenter. Participants are told that the experiment is about X when it is really about Y. X is made to appear plausible and is such that if the participants modify their responses in line with, or antagonistically to, what the experimenter appears to be after, there is no systematic effect on the experimenter's real area of interest. As discussed in Chapter 9 (p. 205) increasing sensitivity to the ethical issues raised by deceiving participants means that this ploy, previously common in some areas of social psychology, is now looked on with increasing suspicion.

Experimenter expectancy effects are reactive effects produced by the experimenters who have been shown, in a wide variety of studies, to bias findings (usually unwittingly) to provide support for the experimental hypothesis. Rosenthal and Rubin (1980) discuss the first 345 such studies! The effects can be minimized by decreasing the amount of interaction between participant and experimenter: using taped instructions, automated presentation of materials, etc. However, for many topics (apart from studies in areas such as human-computer interaction) this further attenuates any real world links that the laboratory experiment might possess. Double-blind procedures can also be used, where data collection is subcontracted so that neither the person working directly with the participants, nor the participants themselves, are aware of the hypothesis being tested.

Knowledge about determinants of laboratory behaviour (demand characteristics, etc.) can be of value in real life settings. For example, police identity parades can be thought of as experiments, and suggestions for improving them have been based on this knowledge and on general principles of experimental design (Wells *et al.*, 1998).

Experiments in natural settings

The laboratory is essentially a place for maximizing control over extraneous variables. Move outside the laboratory door and such tight and comprehensive control becomes impossible. The problems of experimentation discussed in the previous section remain. Any special conditions marking out what is happening as 'an experiment' can lead to reactive effects. The classic demonstration of such effects comes from the well-known series of experiments carried out at the Hawthorne works of the Western Electric Company in the USA in the 1920s and 1930s (Dickson and Roethlisberger, 2003), and hence called the 'Hawthorne effect'. Their studies investigating changes in length of working day, heating, lighting and other variables, found increases in productivity during the study which were virtually irrespective of the specific changes. The workers were in effect reacting positively to the attention and special treatment given by the experimenters. Re-evaluations of the original study have cast serious doubt on the existence of the effect (Kompier, 2006) and the interpretation of the original study (Wickström and Bendix, 2000). However, new, more strictly controlled, studies have demonstrated the existence of (relatively small) Hawthorne effects (McCarney *et al.*, 2007; Verstappen *et al.*, 2004).

Problems in carrying out experiments in natural settings are listed in Box 5.3. There are gains, of course. Notwithstanding some degree of artificiality, and related reactivity, generalizability to the 'real world' is almost self-evidently easier to achieve when the

BOX 5.3

Problems in carrying out experiments in natural settings

Moving outside the safe confines of the laboratory may well be traumatic. Particular practical difficulties include:

1. *Random assignment.* There are practical and ethical problems of achieving random assignment to different experimental treatments or conditions (e.g. in withholding the treatment from a no-treatment control group). Random assignment is also often only feasible in atypical circumstances or with selected respondents, leading to questionable generalizability. Faulty randomization procedures are not uncommon (e.g. when procedures are subverted through ignorance, kindness, etc.). For small samples of the units being randomly assigned, sampling variability is a problem. Treatment-related refusal to participate or continue can bias sampling.
2. *Validity.* The actual treatment may be an imperfect realization of the variable(s) of interest, or a restricted range of outcomes may be insensitively or imperfectly measured, resulting in questionable validity. A supposed no-treatment control group may receive some form of compensatory treatment, or be otherwise influenced (e.g. through deprivation effects).
3. *Ethical issues.* There are grey areas in relation to restricting the involvement to volunteers, the need for informed consent and the debriefing of participants after the experiment. Strict adherence to ethical guidelines is advocated, but this may lead to losing some of the advantages of moving outside the laboratory (e.g. leading to unnecessary 'obtrusiveness', and hence reactivity, of the treatment). Common sense is needed. If you are studying a natural experiment where some innovation would have taken place whether or not you were involved, then it may simply be the ethical considerations relating to the innovation which apply (fluoridation of water supplies raises more ethical implications for users than an altered design of a road junction). See Chapter 9 for further discussion.
4. *Control.* Lack of control over extraneous variables may mask the effects of treatment variables, or bias their assessment. Interaction between participants may vitiate random assignment and violate their assumed independence.

study takes place outside the laboratory in a setting which is almost real 'real life'. Note, however, that there are claims of good generalization of some findings from laboratory to field settings (Locke, 1986). Other advantages are covered in Box 5.4.

Experimental designs as such are equally applicable both inside and outside laboratories. The crucial feature of so-called 'true' experiments (distinguishing them from 'quasi-experiments' discussed below) is *random allocation of participants to experimental conditions*. If you can find a feasible and ethical means of doing this

BOX 5.4

Advantages in carrying out experiments in natural settings

Compared to a laboratory, natural settings have several advantages:

1. *Generalizability.* The laboratory is necessarily and deliberately an artificial setting where the degree of control and isolation sets it apart from real life. If we are concerned with generalizing results to the real world, the task is easier if experimentation is in a natural setting. Much laboratory experimentation is based on student participants, making generalization to the wider population hazardous. Although this is not a necessary feature of laboratory work, there is less temptation to stick to student groups when experiments take place in natural settings.
2. *Validity.* The *demand characteristics* of laboratory experiments, where participants tend to do what they think you want them to do, are heightened by the artificiality and isolation of the laboratory situation. Real tasks in a real world setting are less prone to this kind of game playing. So you are more likely to be measuring what you think you are measuring.
3. *Participant availability.* It is no easy task to get non-student participants to come into the laboratory (although the development of pools of volunteers is valuable). You have to rely on them turning up. Although it depends on the type of study, many real life experiments in natural settings have participants in abundance, limited only by your energy and staying power – and possibly your charm.

when planning a field experiment, then you should seriously consider carrying out a true experiment.

The advantage of random allocation or assignment is that it allows you to proceed on the assumption that you have equivalent groups under the two (or more) experimental conditions. This is a probabilistic truth, which allows you, among other things, to employ a wide battery of statistical tests of inference legitimately. It does not guarantee that in any particular experiment the two groups will in fact be equivalent. No such guarantee is ever possible, although the greater the number of persons being allocated, the more confidence you can have that the groups do not differ widely.

An alternative way of expressing this advantage is to say that randomization gets rid (probabilistically at least) of the *selection* threat to internal validity (see Box 5.1, p. 88). That is, it provides a defence against the possibility that any change in a dependent variable is caused not by the independent variable but by differences in the characteristics of the two groups. Other potential threats to internal validity remain and the discussion of some of the designs that follows is largely couched in terms of their adequacy, or otherwise, in dealing with these threats.

Randomized controlled trials and the ‘gold standard’

A randomized controlled trial (RCT) is a type of experiment where participants are randomly allocated, either to a group who receive some form of intervention or treatment, or to a control group who don't. Use of RCTs is a central feature of the evidence-based movement currently highly influential in many fields of social research. Proponents argue that it is the ‘gold standard’ – the scientific method of choice, primarily because they consider it to be the best means of assessing whether or not the intervention is effective.

There is a growing tendency in some circles to equate the doing of science with the carrying out of RCTs. For example, the US Department of Education's Institute of Education Sciences in 2003 (www.eval.org/doi.fedreg.htm) made it clear that it would privilege applications for funding of applied research and evaluation which used RCTs (with a grudging acceptance of other experimental approaches when RCTs were not feasible) in the interests of using ‘rigorous scientifically based research methods’. This has sparked a heated debate among applied social researchers on ‘what counts as credible evidence in applied research and evaluation practice’ (Donaldson, Christie and Mark, 2009).

Privileging RCTs in this way is a serious distortion of the nature of scientific activity. It is historically inaccurate and carries with it an inappropriately narrow view of what constitutes evidence. A review of practices in the natural sciences reveals the minor role played by RCTs. Phillips (2005) concludes that:

One cannot help but be struck by the huge range of activities engaged in by researchers in the natural sciences, and the variety of types of evidence that have been appealed to: establishing what causal factors are operating in a given situation; distinguishing genuine from spurious effects; determining function; determining structure; careful description and delineation of phenomena; accurate measurement; development and testing of theories, hypotheses, and causal models; elucidation of the mechanisms that link cause with effect; testing of received wisdom; elucidating unexpected phenomena; production of practically important techniques and artefacts (p. 593).

He gives a wide range of illustrative examples. In his view, any attempt to give a simple, single account of the nature of science appears quite arbitrary. Relying on RCTs, or any other specific methodology, as the criterion of scientific rigour ‘detracts from the main question at hand when one is assessing an inquiry, which is this: *Has the overall case made by the investigator been established to a degree that warrants tentative acceptance of the theoretical or empirical claims that were made?*’ (original emphasis).

While the methodology used in a particular study is an important consideration it is the convincingness of the argument; how well the evidence is woven into the structure of the argument; how rigorously this evidence was gathered; and how well counter-arguments and counter-claims are themselves countered or confronted with recalcitrant facts or data.

The message here is not that RCTs should be avoided – they have an important role and many practising real world researchers will be expected to be able to carry them out competently. However, they are by no means the only show in town.

✦ *The website provides further discussion on the issues involved in social experimentation and the use of RCTs.*

Realist critique of RCTs

Pawson and Tilley (1997, especially Chapter 2) elaborate the view that the methodology of the RCT is inappropriate for dealing with complex social issues. They consider that, as well as generating inconsistent findings, the concentration on outcomes does little or nothing to explain why an intervention has failed (or, in relatively rare cases, succeeded). Hence, there is not the cumulation of findings which would help to build up understanding.

Experimentalists acknowledge the practical and ethical problems of achieving randomization of allocation to experimental and control groups in applied field experiments. Pawson and Tilley add causal problems to these perils. Allocation of participants to experimental or control groups by the experimenter removes that choice from the participants but *'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, p. 36; emphasis in original).

In their discussion of correctional programmes for prison inmates, they make the undeniable point that it is not the programmes themselves which work, but people cooperating and choosing to make them work. The traditional solution to this problem is to run volunteer-only experiments. Volunteers are called for, then assigned randomly to one of the two groups. The assumption is that motivation and cooperation will be the same in each of the groups. The reasonableness of this assumption will depend on the specific circumstances of the experiment. Pawson and Tilley illustrate, through an example of what they consider to be high-quality experimental evaluation research (Porperino and Robinson, 1995), the way in which participants' choice-making capacity cuts across and undermines a volunteer/non-volunteer distinction:

The act of volunteering merely marks a moment in a whole evolving pattern of choice. Potential subjects will consider a program (or not), volunteer for it (or not), co-operate closely (or not), stay the course (or not), learn lessons (or not), retain the lessons (or not), apply the lessons (or not). Each one of these decisions will be internally complex and take its meaning according to the chooser's circumstances. Thus the act of volunteering for a program such as 'Cog Skills' might represent an interest in rehabilitation, a desire for improvement in thinking skills, an opportunity for a good skive, a respite from the terror or boredom of the wings, an opening to display a talent in those reputedly hilarious role-plays, a chance to ogle a glamorous trainer, a way of playing the system to fast-track for early parole, and so on (p. 38).

They back up this intuitive understanding of how prisoners find their way on to programmes by a detailed re-analysis of the findings. Their overall conclusion is that

such volunteer-only experiments encourage us to make a pronouncement on whether a programme works without knowledge of the make-up of the volunteers. The crucial point is the 'programs tend to work for some groups more than others, but the methodology then directs attention away from an investigation of these characteristics and towards . . . the battle to maintain the equivalence of the two subsets of this self-selected group' (p. 40).

However, the messages that come from RCTs are undoubtedly invested with considerable value by many audiences. As we appear to be approaching a situation where governments and other decision-making bodies are more receptive to evidence from research findings, why not use RCTs? Unfortunately, the track record for RCTs in social research is very poor. Even Oakley (2000) in arguing for their use accepts the continuing equivocal nature of their findings, while putting some proposals forward for improving them. We are in danger of repeating the cycle of enthusiasm–disillusion found by educational experimenters in the 1920s and 1930s and evaluation research in the US in the 1960s and 1970s.

A possible way forward is via the realist mantra of establishing 'what works, for whom, and in which contexts', rather than looking for overall effects of social programmes, interventions, etc. By establishing the likely operative mechanisms for different groups or types of participants in particular situations and settings, it becomes feasible to set up circumstances where large effects are obtained. In other words, the experiment is retained as the tool for obtaining quantitative confirmation of something that we already know to exist (or have a strong intuition or hunch as to its existence).

How is this actually done? Pawson and Tilley (1997), discussing these matters largely in the context of large-scale evaluative research, advocate the use of subgroup analysis. With large numbers of participants, it becomes feasible to set up contrasts between subgroups illustrating and substantiating the differential effects of mechanisms on different subgroups. For small-scale studies, and in real world settings where relevant previous work may be sparse or non-existent, there is much to be said for a *multi-strategy design* with an initial flexible design stage primarily exploratory in purpose. This seeks to establish, both from discussions with professionals, participants and others involved in the initial phase, and from the empirical data gathered, likely 'bankers' for mechanisms operating in the situation, contexts where they are likely to operate, and the characteristics of participants best targeted. A second fixed design phase then incorporates a highly focused experiment or other fixed design study. An RCT may be the design of choice for this second phase if:

- the sponsor of the research, and/or important decision-makers consider the evidence from an RCT to be required (either for their own purposes or to help in making a case to others); *and*
- the establishment of randomized experimental and control groups is feasible practically and ethically; *and*
- it appears unlikely that there will be differential effects on the experimental or control groups unconnected to the intervention itself (e.g. persons within the control group become disaffected or disgruntled because of their non-selection for the experimental group).

Note that the subgroup for whom a particular mechanism is considered to be likely to be operative (as established in the initial phase) should form the basis for the pool of volunteers from whom the experimental and control groups are randomly formed. Where feasible, similar restrictions can be placed on the contexts so that they are equivalent for the two groups.

If one or other of the three circumstances listed above does not obtain, then other designs can be considered. If randomization can be achieved then a true experiment involving two or more comparison groups (rather than an experimental and control group) has attractions. For example when the initial work indicates different contexts or settings where a particular enabling mechanism is likely to operate in one context but not in the second or a disabling mechanism operates in the second. This avoids problems in establishing 'non-intervention' control groups.

Where there are problems, of whatever kind, in achieving randomization, quasi-experimental designs remain feasible. A control group design might be used, with efforts being made to ensure as far as possible that the experimental and control groups are closely equivalent (particularly in aspects identified during the initial phase as being of relevance to the operation of the mechanisms involved; e.g. by using selected participants for the two groups for whom a particular mechanism appears salient). Quasi-experimental designs (see p. 109) can be used in situations where a mechanism is considered to be likely to be operative with one set or subgroup of participants but not with a second subgroup. Or where an additional disabling mechanism is thought to be operative in the second subgroup. The initial exploratory phase is used not only to build up a picture of the likely enabling and disabling mechanisms, but also to find a way of typifying or categorizing the best ways in which participants might be grouped to illustrate the operation of these mechanisms. Randomized allocation of participants is, of course, not possible when a comparison between different subgroups is being made.

Single-case designs (p. 118) lend themselves well to a realist reconceptualization. The strategy of thoroughly analysing and understanding the situation so that reliable and reproducible effects can be achieved bears a striking resemblance to the methodology developed by the experimental psychologist B. F. Skinner (Sidman, 1960), even though the terminology and underlying philosophy are very different. Similarly the various non-experimental fixed designs can be viewed through realist eyes. In particular, they lend themselves to the type of subgroup analyses advocated by Pawson and Tilley (1997).

The designs discussed in the following sections of this chapter can be looked at using the realist perspective considered here. While they bear more than a passing resemblance to traditional positivist-based experimental and non-experimental designs (and can be used in this traditional manner by those who have not yet seen the realist light), there are major hidden differences. As discussed above, the participants involved in the different groups, the situations, circumstances and contexts and the aspects of an intervention or programme that are targeted, are all carefully selected and refined in the interests of obtaining substantial clear differential effects. This is simply a rephrasing of the injunction that fixed designs are theory-driven which opened this chapter. By an initial exploratory phase where hunches and hypotheses about the likely mechanisms and contexts and those participants for whom the mechanisms will operate (or by some other means such as modelling your approach on earlier work, or by yourself having an

intimate experience of the working of a programme or intervention; or talking to those who have that experience and understanding), you set up a highly focused study. It is worth noting that this is the common approach taken in the natural sciences. The actual experiment to test a theory is the culmination of much prior thought and exploration, hidden in the textbook rationalizations of the scientific method and the conventions of experimental report writing.

A successful experiment with clear differential outcomes is supporting evidence for the causal mechanisms we proposed when designing the study and a contribution to understanding where, how and with whom, they operate.


Systematic reviews and the evidence-based movement

Systematic reviews are a specific way of identifying and synthesizing research evidence. They are closely linked to the evidence-based movement which includes some of the strongest advocates of using RCTs in social research. They differ from the traditional literature reviews (discussed in Chapter 3, p. 51) by their emphasis on:

- providing a comprehensive coverage of the available literature in the field of interest;
- the quality of the evidence reviewed;
- following a detailed and explicit approach to the synthesis of the data; and
- the use of transparent and rigorous processes throughout.

Their justification is in the reliability and validity of the findings from following this process. Petticrew and Roberts (2005) provide a useful practical guide to carrying out a systematic review. The amount of time, effort and resources required to do this would preclude their use in preparation for a small-scale real world research project, although if a recent review relevant to your research questions exists, by all means make use of it.

Systematic reviews, and the high status that they are given, have their critics. See, in particular, Pawson (2006b) who provides a scathing critique of evidence-based systematic reviews from a realist perspective.

 *The website gives further discussion on the development of systematic reviews.*

Sources of systematic reviews

A register of reports and papers, including RCTs on social, behavioural and educational interventions known as the Social, Psychological, Educational and Criminological Trials Register (SPECTR) has been developed by an international organization known as the Campbell Collaboration (www.campbellcollaboration.org) (Boruch, Soydan and de Moya, 2004). It is modelled on the approach taken by the Cochrane Centres in the medical and health fields. They take their title from Donald Campbell, the methodologist who did substantial work on assessing the validity of causal influences about the effects of interventions (see p. 88). See also the UK's Evidence Network (www.kcl.ac.uk/schools/sspp/interdisciplinary/evidence/) which, as well as providing access to systematic reviews, provides links to other centres focusing on the use of evidence in policy making and practice, and the London University Institute of Education's Evidence for

Policy and Practice Information and Co-ordinating Centre (<http://eppi.ioe.ac.uk/cms/>) (EPPI Centre, 2007). The What Works Clearinghouse (WWC) (<http://ies.ed.gov/ncee/wwc/>), established by the US Department of Education's Institute of Education Sciences, focuses on evidence-based practices in school education.

True experiments

A small number of simple designs are presented here. Texts on experimental design give a range of alternatives, and of more complex designs (e.g. Maxwell and Delaney, 2003; Shadish *et al.*, 2002).

Often those involved in real world experimentation restrict themselves to the very simplest designs, commonly the 'two group' design given below. However, the main hurdle in carrying out true experiments outside the laboratory is in achieving the principle of random allocation, and once this is achieved there may be merit in considering a somewhat more complex design.

Box 5.5 provides an overview of the some commonly used true experimental designs.

BOX 5.5

Overview of simple true experimental designs

Note: The defining characteristic of a true experimental design is random allocation of participants to the two (or more) groups of the design.

1. Two-group designs

- (a) *Post-test-only randomized controlled trial.* Random allocation of participants to an experimental group (given the experimental 'treatment') and a 'no-treatment' control group. Post-tests of the two groups compared.
- (b) *Post-test-only two treatment comparison.* Random allocation of participants to experimental group 1 (given experimental 'treatment' 1), or to experimental group 2 (given experimental 'treatment' 2). Post-tests of the two groups compared.
- (c) *Pre-test post-test randomized controlled trial.* Random allocation of participants to an experimental group (given the experimental 'treatment') and a 'no-treatment' control group. Pre-test to post-test changes of individuals in the two groups compared.
- (d) *Pre-test post-test two treatment comparison.* Random allocation of participants to experimental group 1 (given experimental 'treatment' 1), or to experimental

group 2 (given experimental 'treatment' 2). Pre-test to post-test changes of individuals in the two groups compared.

2. Three- (or more) group simple designs

It is possible to extend any of the above two group designs by including additional experimental groups (given different experimental 'treatments'). The RCTs retain a 'no-treatment' control group.

3. Factorial designs

Two (or more) independent variables (IVs) involved (e.g. 'type of music' and 'number of decibels'). Each IV studied at two (or more) 'levels'. Random allocation of participants to groups covering all possible combinations of levels of the different IVs. Can be post-test only or pre-test post-test.

4. Parametric designs

Several 'levels' of an IV covered with random allocation of participants to groups to get a view of the effect of the IV over a range of values. Can be post-test only or pre-test post-test.

5. Matched pairs designs

Establishing pairs of participants with similar scores on a variable known to be related to the dependent variable (DV) of the experiment. Random allocation of members of pairs to different experimental groups (or to an experimental and control group). This approach can be used in several two group designs. Attractive, but can introduce complexities both in setting up and in interpretation.

6. Repeated measures designs

Designs where the same participant is tested under two or more experimental treatments or conditions (or in both an experimental and control condition). Can be thought of as the extreme example of a matched pairs design.

Designs involving matching


In its simplest form, the *matched pairs design*, matching involves testing participants on some variable which is known to be related to the dependent variable on which observations are being collected in the experiment. The results of this test are then used to create 'matched pairs' of participants, that is, participants giving identical or very similar scores on the related variable.

Random assignment is then used to allocate one member of each pair to the treatment group and one to the comparison group. In this simplest form, the design can be considered as an extension of the simple two-group design, but with randomization being carried out on a pair basis rather than on a group basis. The principle can be easily extended to other designs, although of course if there are, say, four groups in the design then 'matched fours' have to be created and individuals randomly assigned from them to the four groups.

While the selection and choice of a good matching variable may pose difficult problems in a field experiment, it is an attractive strategy because it helps to reduce the problem of differences between individuals obscuring the effects of the treatment in which you are interested. Generally we need all the help we can get to detect treatment effects in the poorly controlled field situation, and matching can help without setting strong restrictions on important variables (which could have the effect of limiting the generalizability of your findings). To take a simple example, suppose that age is a variable known to be strongly related to the dependent variable in which you are interested. It would be possible to control for age as a variable by, say, only working with people between 25 and 30 years old. However, creating matched age pairs allows us to carry out a relatively sensitive test without the conclusions being restricted to a particular and narrow age range.

Designs involving repeated measures

The ultimate in matching is achieved when an individual's performance is compared under two or more conditions. Designs with this feature are known as *repeated measures* designs. We have come across this already in one sense in the 'before and after' design – although the emphasis there is not on the before and after scores per se, but on the relative difference between them in the treatment and comparison groups as a measure of the treatment effect.

 The website discusses some methodological problems with designs using matching or repeated measures.

Choosing among true experimental designs

Box 5.6 gives suggestions for the conditions under which particular experimental designs might be used when working outside the laboratory. Cook and Campbell (1979) have discussed some of the real world situations which are conducive to carrying out randomized experiments. Box 5.7 is based on their suggestions.

BOX 5.6

Considerations in choosing among true experimental designs

1. To do any form of true experimental design you need to be able to carry out random assignment to the different treatments. This is normally random assignment of persons to treatments (or of persons to the order in which they receive different treatments, in repeated measures designs). Note, however, that the unit which is randomly assigned need not be the person; it could be a group (e.g. a school class), in which case the experiment, and its analysis, is on classes, not individuals.

2. Use a *matched design* when:
 - (a) you have a matching variable which correlates highly with the dependent variable;
 - (b) obtaining the scores on the matching variable is unlikely to influence the treatment effects; and
 - (c) individual differences between participants are likely to mask treatment effects.
3. Use a *repeated measures design* when:
 - (a) order effects appear unlikely;
 - (b) the independent variable(s) of interest lend themselves to repeated measurement (participant variables such as sex, ethnic background or class don't – it is not easy to test the same person as a man and as a woman);
 - (c) in real life, persons would be likely to be exposed to the different treatments; and
 - (d) individual differences between participants are likely to mask treatment effects.
4. Use a *simple two-group design* when:
 - (a) order effects are likely;
 - (b) the independent variable(s) of interest don't lend themselves to repeated measurement;
 - (c) in real life, persons would tend not to receive more than one treatment; and
 - (d) persons might be expected to be sensitized by pre-testing or being tested on a matching variable.
5. Use a *before-after design* when:
 - (a) pre-testing appears to be unlikely to influence the effect of the treatment;
 - (b) there are concerns about whether random assignment has produced equivalent groups (e.g. when there are small numbers in the groups); and
 - (c) individual differences between participants are likely to mask treatment effects.
6. Use a *factorial design* when:
 - (a) you are interested in more than one independent variable; and
 - (b) interactions between independent variables may be of concern.
7. Use a *parametric design* when:
 - (a) the independent variable(s) have a range of values or levels of interest; and
 - (b) you wish to investigate the form or nature of the relationship between independent variable and dependent variable.

There are occasions when one starts out with a true experiment but along the way problems occur, perhaps in relation to assignment to conditions, or to mortality (loss of participants) from one or other group, or where you don't have the time or resources to carry out what you originally intended. Such situations may be rescuable by reconceptualizing what you are proposing as one of the quasi-experiments discussed below.

BOX 5.7

Real life situations conducive to randomized experiments

1. *When lotteries are expected.* Lotteries are sometimes, though not commonly, regarded as a socially acceptable way of deciding who gets scarce resources. When done for essentially ethical reasons it provides a good opportunity to use this natural randomization for research purposes.
2. *When demand outstrips supply.* This sets up a situation where randomized allocation may be seen as a fair and equitable solution (however, using randomization to allocate places at oversubscribed schools in England has proved highly controversial). There are practical problems. Do you set up waiting lists? Or allow reapplication? Cook and Campbell (1979) advocate using the initial randomization to create two equivalent no-treatment groups, as well as the treatment group. One no-treatment group is told that their application is unsuccessful, and that they cannot reapply. This group acts as the control group. The second no-treatment group is permitted to go on a waiting list, they are accepted for the treatment if a vacancy occurs, but data from them are not used.
3. *When an innovation cannot be introduced in units simultaneously.* Many innovations have to be introduced gradually, because of resource or other limitations. This provides the opportunity for randomization of the order of involvement. Substantial ingenuity may be called for procedurally to balance service and research needs, particularly when opportunities for involvement arise irregularly.
4. *When experimental units are isolated from each other.* Such isolation could be temporal or spatial – or simply because it is known that they do not communicate. Randomization principles can then be used to determine where or when particular treatments are scheduled.
5. *When it is agreed that change should take place but there is no consensus about solutions.* In these situations decision-makers may be more susceptible to arguments in favour of a system of planned variation associated with random allocation.
6. *When a tie can be broken.* In situations where access to a particular treatment is based upon performance on a task (e.g. for entry to a degree or other course) there will be a borderline. It may be that several persons are on that border (given the less than perfect reliability of any such task, this is more accurately a border region than a line). Randomization can be used to select from those at the border who then form the treatment and no-treatment control groups.
7. *When persons express no preference among alternatives.* In situations where individuals indicate that they have no preference among alternative treatments, their random assignment to the alternatives is feasible. Note that you will be comparing the performance on the treatments of those without strong preferences, who may not be typical.

8. *When you are involved in setting up an organization, innovation, etc.* Many opportunities for randomization present themselves if you as researcher can get in on the early stages of a programme, organization or whatever. It would also help if guidelines for local and national initiatives were imbued with a research ethos, which would be likely to foster the use of randomization.

(after Cook and Campbell, 1979, pp. 371–86)

Quasi-experiments

The term *quasi-experiment* has been used in various ways, but its rise to prominence in social experimentation originates with a very influential chapter by Campbell and Stanley in Gage's *Handbook of Research on Teaching*. This was republished as a separate slim volume (Campbell and Stanley, 1963). For them, a quasi-experiment is:

A research design involving an experimental approach but where random assignment to treatment and comparison groups has not been used.

Campbell and Stanley's main contribution was to show the value and usefulness of several such designs. More generally, they have encouraged a flexible approach to design and interpretation, where the particular pattern of results and circumstances under which the study took place interact with the design to determine what inferences can be made. Their concern is very much with the *threats to validity* present in such studies (see Box 5.1, p. 88), and with the extent to which particular threats can be plausibly discounted in particular studies. Quasi-experimental approaches have considerable attraction for those seeking to maintain a basic experimental stance in work outside the laboratory.

Quasi-experiments are often viewed as a second-best choice, a fall-back to consider when it is not possible to randomize allocation. Cook and Campbell (1979), however, prefer to stress the relative advantages and disadvantages of true and quasi-experiments, and are cautious about always advocating randomized experiments even when they are feasible. They recommend considering all possible design options without necessarily assuming the superiority of a randomized design – and with the proviso that if a randomized design is chosen then it should be planned in such a way as to be interpretable as a quasi-experimental design, just in case something goes wrong with the randomized design, as it may well do in the real world.

Box 5.8 provides an overview of the main types of quasi-experimental designs covered in the following section.

BOX 5.8

Overview of a range of quasi-experimental designs

Note: A quasi-experimental design follows the experimental approach to design but does not involve random allocation of participants to different groups. The following list outlines a few commonly used designs.

1. *Pre-experimental designs*
 - (a) *Single-group post-test-only.*
 - (b) *Post-test only non-equivalent groups* i.e. use of groups established by some procedure other than randomization (e.g. two pre-existing groups).
 - (c) *Pre-test post-test single group design.*

These designs should be avoided owing to difficulties in interpreting their results (though they may be of value as part of a wider study, or as a pilot phase for later experimentation).
2. *Pre-test post-test non-equivalent group designs*
Two (or more) groups established on some basis other than random assignment. One of these might be a control group. Interpretation of findings more complex than with equivalent true experimental designs.
3. *Interrupted time series designs*
In its simplest (and most common) form, involves a single experimental group on which a series of measurements or observations are made before and after some form of experimental intervention. Requires a dependent variable on which repeated measures can be taken and an extended series of measurements.
4. *Regression-discontinuity designs*
All participants are pre-tested and those scoring below a criterion value are assigned to one group (say an experimental group); all those above that criterion are assigned to a second group (say a control group). The pattern of scores after the experimental intervention provides evidence for its effectiveness.

Quasi-experimental designs to avoid – the ‘pre-experiments’

Quasi-experimental designs are essentially defined negatively – they are not true experimental designs. They include several which are definitely to be avoided, although these so-called ‘pre-experimental’ designs (listed in Box 5.8) continue to get used, and even published. Details are presented here (in Boxes 5.9, 5.10 and 5.11) to enable you to recognize and avoid them, and also because the reasons why they are problematic present useful methodological points.

The ‘pre-test post-test single-group’ design is commonly found and it is important to stress that the deficiencies covered here concern its nature as an *experimental design* where

BOX 5.9

Designs to avoid, no.1: the one-group post-test only design

Scenario: A single experimental group is involved in the treatment and then given a post-test.

Reasons to avoid: As an experiment, where the only information that you have is about the outcome measure, this is a waste of time and effort. Without pre-treatment measures on this group or measures from a second no-treatment control group, it is virtually impossible to infer any kind of effect.

Improvements: Either improve the experimental design or adopt a case study methodology.

Note: This is not the same thing as a case study. Typically the case study has multiple *sources* of data (usually qualitative, but some may be quantitative) extending over time, and there is also information about the context.

BOX 5.10

Designs to avoid, no. 2: the post-test only non-equivalent groups design

Scenario: As no. 1 but with the addition of a second non-equivalent (not determined by random assignment) group that does not receive the treatment, i.e.:

- a. Set up an experimental and a comparison group on some basis other than random assignment.
- b. The experimental group gets the treatment, the comparison group doesn't.
- c. Do post-tests on both groups.

Reasons to avoid: It is not possible to determine whether any difference in outcome for the two groups is due to the treatment, or to other differences between the groups.

Improvements: Strengthen the experimental design by incorporating a pre-test or by using random assignment to the two groups; or use case study methodology.

you are trying to decide whether the experimental treatment was responsible for the effects found. If the concern is simply to determine whether there is an increase of performance after a treatment, or even to assess its *statistical* significance (see the discussion in Chapter 16, p. 446), there are no particular problems. The difficulty is in possible validity threats. They may also be useful as pilot studies, to determine whether it is worthwhile to commit resources to carry out a more adequate experiment.

BOX 5.11

Designs to avoid, no. 3: the pre-test post-test single-group design

Scenario: As no. 1, but with the addition of measurement on the same variable before the treatment as well as after it; i.e. the single experimental group is pre-tested, gets the treatment, and is tested again.

Reasons to avoid: Although widely used, it is subject to lots of problems. It is vulnerable to many threats to validity – including *history* (other events apart from the treatment occurring between measures), *maturational* (developments in the group between measures), *statistical regression* (e.g. choice of a group ‘in need’ in the sense of performing poorly on the measure used, or some other measure which correlates with it, will tend to show an improvement for random statistical reasons unconnected with the treatment – see p. 113).

Improvements: Strengthen the experimental design, e.g. by adding a second pre-tested no-treatment control group.

Note: It may be possible on particular occasions to show that this design is interpretable. This could be because the potential threats to validity have not occurred in practice. For example, if you can isolate the group so that other effects do not influence it; or if you have information that there are no pre-treatment trends in the measures you are taking – although strictly that type of information turns this into a kind of time-series design (see p. 114).

Quasi-experimental designs to consider

It is possible to get at a feasible quasi-experimental design by considering the main problems with the previous two designs – the ‘post-test only non-equivalent groups’ design, and the ‘pre-test post-test single-group’ design. With the former, we do not know whether or not the two groups differ before the treatment. With the latter, we do not know how much the group would have changed from pre-test to post-test in the absence of the treatment.

One tactic used to strengthen the design is, effectively, to combine the two designs into a ‘pre-test post-test non-equivalent groups’ design. A second tactic is to make additional observations:

- over time with a particular group, leading to the ‘interrupted time-series’ design (p. 110); and/or
- over groups at the same time, leading to the ‘regression-discontinuity’ design (p. 110).

In quasi-experiments the pattern of pre-test and post-test results has to be investigated to assess the effectiveness of the treatment. It is a general rule of quasi-experimental designs that it is necessary to consider not only the *design* of a study, but also the *context* in

which it occurs, and the particular *pattern of results* obtained, when trying to decide whether a treatment has been effective. Note the similarities with the realist approach, e.g. in the emphasis on context and the importance of detailed analysis of what actually happens in a study, the stress on ‘what works, for whom and in what circumstances’. This is not surprising as Cook and Campbell (1979, pp. 28–36) endorse a realist approach and seek to move beyond positivist notions of causation in their analysis.

Pre-test post-test non-equivalent groups design

The interpretability of the findings depends on the pattern of results obtained. If, for example, the experimental group starts lower and the outcome is an increase in the experimental group taking it above the comparison group at post-test, while there is no change in comparison group, the switching of the two groups from pre- to post-test permits many threats to validity to be ruled out. Other patterns can be more difficult to interpret.

📖 *The website gives an analysis of the interpretability of different patterns of outcomes.*

A common strategy in this type of design is to use one or more matching variables to select a comparison or control group. This is different from the matching strategy used in true or randomized experiments where the experimenter matches participants and randomly assigns one member of the matched pair to either treatment or comparison group. Here the researcher tries to find participants who match the participants who are receiving a treatment. This approach is unfortunately subject to the threat to internal validity known as *regression to the mean*. While this threat is always present when matching is used without random assignment, it shows itself particularly clearly in situations where some treatment intended to assist those with difficulties or disadvantages is being assessed. Suppose that a comparison is being made between the achievements of a ‘disadvantaged’ and a ‘non-disadvantaged’ control group. The pre-treatment levels of the disadvantaged population will almost inevitably differ from those of the non-disadvantaged population, with the strong likelihood that those of the disadvantaged population tend to be lower. Hence in selecting matched pairs from the two populations, we will be pairing individuals who are pretty high in the disadvantaged group with individuals pretty low in the non-disadvantaged group. Figure 5.1 indicates what is likely to be going on.

Because pre-test scores are not 100 per cent reliable (no scores ever are), they will incorporate some random or error factors. Those scoring relatively high in their population (as in the selected disadvantaged group) will tend to have positive error factors inflating their pre-test score. Those scoring relatively low in their population (as in the selected non-disadvantaged group) will tend to have negative error factors reducing their pre-test score. On post-test, however, such random factors (simply because they are random) will be just as likely to be positive as negative, leading to the ‘regression to the mean’ phenomenon – post-test scores of originally extreme groups tend to be closer to their population means.

As can be seen from the figure, the effect of this is to produce a tendency for the disadvantaged groups to score lower *even in the absence of any treatment effects*. Depending on the relative size of this effect and any treatment effect, there will appear to be a reduced treatment effect, or zero, or even a negative one.

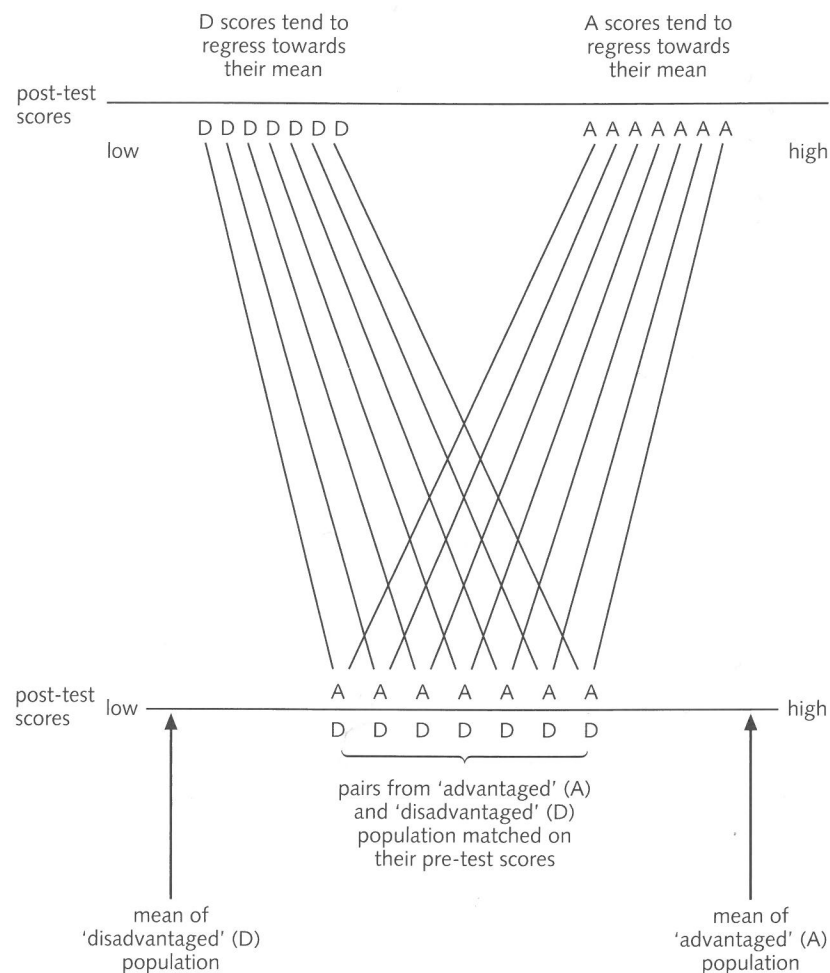


Figure 5.1: Effects of ‘regression to the mean’ when using extreme groups

Time-series designs

In the simplest form of this design, there is just one experimental group, and a series of observations or tests before and after an experimental treatment. The time-series approach is widely used in some branches of the social sciences (e.g. in economics) and has a well-developed and complex literature, particularly on the analysis of time-series data (Glass, Willson and Gottman, 2008). Textbooks covering this field suggest rules of thumb for the number of data points needed in the before and after time series, typically coming up with figures of 50 or more.

This extent of data collection is likely to be outside the scope of the small-scale study targeted in this book. However, there may well be situations where, although 50 or so observations are not feasible, it is possible to carry out several pre- and post-tests. Certainly, some advantages accrue if even one additional pre- and/or post-test (preferably both) can be added. This is essentially because one is then gathering information

about possible trends in the data, which help in countering several of the threats to the internal validity of the study. Coryn, Schröter and Hanssen (2009) provide a detailed example. With more data points, say five before and five after, the experimenter is in a much stronger position to assess the nature of the trend – does the series appear to be stationary (i.e. show no trend to increase or decrease)? Or does it appear to increase, or decrease? And is this a linear trend, or is the slope itself tending to increase? Techniques for the analysis of such short time series are available, although not universally accepted (see Chapter 16, p. 460).

As with other quasi-experimental designs, interpretation is based on a knowledge of the design itself in interaction with the particular pattern of results obtained, and contextual factors. Figure 5.2 illustrates a range of possible patterns of results. Collecting data for a time-series design can become a difficult and time-consuming task. The observations must be ones that can be made repeatedly without practical or methodological problems. Simple, non-obtrusive measures (e.g. of play in a school

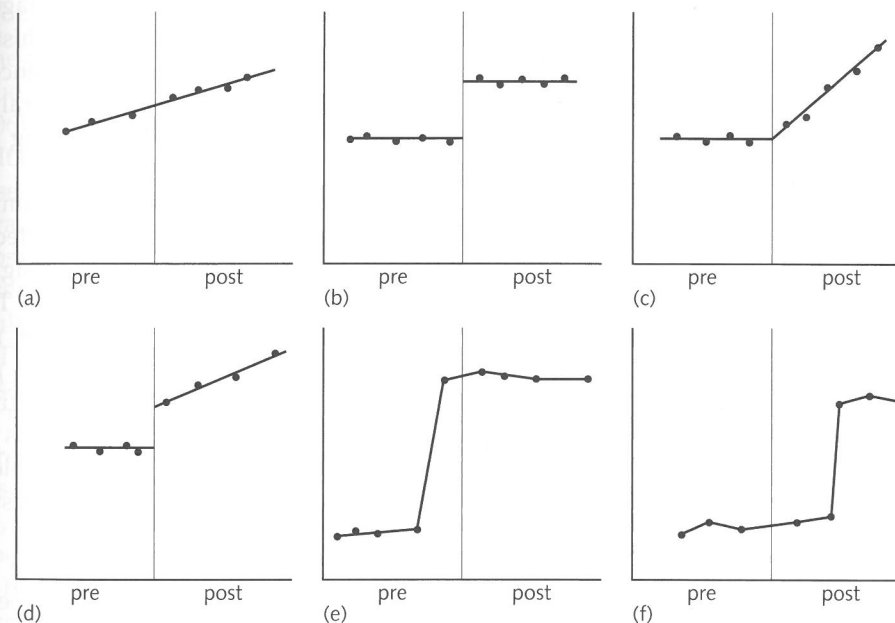


Figure 5.2: Patterns of possible results in a simple time-series experiment. (a) No effect. Note that making single pre- and post-tests (or taking pre- and post-test averages) would suggest a spurious effect. (b) Clear effect. Stable pre and post – but at a different level. Several threats to validity still possible (e.g. history – something else may be happening at the same time as the treatment). (c) Again, clear effect, but of a different kind (move from stability to steady increase). Similar threats will apply. (d) Combines effects of (b) and (c). (e) ‘Premature’ and (f) ‘delayed’ effects. Such patterns cast serious doubts on the effects being causally linked to the treatment. Explanations should be sought (e.g. may get a ‘premature’ effect of an intervention on knowledge or skill, if participants are in some way preparing themselves for the intervention).

playground) are more appropriate than, say, the repeated administration of a formal test of some kind.

If pre-existing archive material of some kind is available, then it may be feasible to set up a time-series design, even with an extended time series, at relatively low cost of time and effort for the experimenter. Increasingly, such material is gathered in conjunction with management information systems. However, it will require very careful scrutiny to establish its reliability and validity, and general usefulness, for research purposes. It will usually have been gathered for other purposes (although if, as is sometimes the case, you are in a position to influence what is gathered and how it is gathered, this can be very helpful), which may well mean that it is inaccessible to you, or is systematically biased, or is being collected according to different criteria at different times, or by different people, or is inflexible and won't allow you to answer your research questions.

While many time-series designs involve a single group before and after some treatment or intervention, more complex designs are possible. For example, a non-equivalent comparison group can be added with the same series of pre- and post-treatment tests or observations being made for both groups. The main advantage of adding the control group is its ability to test for the 'history' threat. A 'selection-history interaction' is still possible, though, that is, that one of the two groups experiences a particular set of non-treatment related events that the other does not. In general the plausibility of such a threat will depend on how closely comparable in setting and experiences the two groups are.

One way of discounting history-related threats is to use the group as its own control, and to take measures on a second dependent variable which should not be affected by the treatment. In a classic study Ross, Campbell and Glass (1970) used this design to analyse the effect of the introduction of the 'breathalyser' on traffic accidents. They argued that serious accidents should decrease following the introduction of new legislation in Britain which brought in the 'breathalyser' *during the hours that pubs were open* (the 'experimental' dependent variable), but should be less affected during commuting hours when the pubs were shut (the 'control' dependent variable) – this was before the days of extended opening hours for British pubs. They were able to corroborate this view, both by visual inspection of the time series and by statistical analysis.

Other time-series designs involving the removal of treatment, and multiple and switching replications, have been used. A lot of the interest in these designs has been in connection with so-called 'single-case' or 'single-subject' research, particularly in the behaviour modification field (e.g. Barlow, Andrasik and Hersen, 2006). Although having their genesis in a very different area of the social sciences, time-series designs show considerable similarities to single-case research designs (see below p. 118).

Regression discontinuity design

This rather fearsomely named design is conceptually straightforward. As in the true experiment, a known assignment rule is used to separate out two groups. However, whereas with the true experiment this is done on a random basis, here some other principle is used. In probably its simplest form, all those scoring below a certain value on

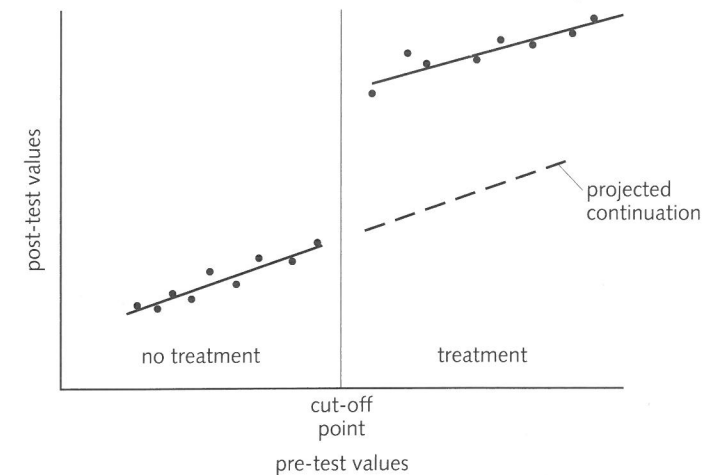


Figure 5.3: Illustrative outcome of a regression discontinuity design.

some criterion are allocated to, say, the experimental group; all those scoring above that value are allocated to the control group (or vice versa). Trochim (1984, 1990) and Lesik (2006) discuss the use of the design.

It might be, for example, that entry to some compensatory programme is restricted to those scoring below a particular cut-off point on some relevant test; or conversely that entrance scholarships are given to those scoring above some cut-off. Figure 5.3 illustrates a possible outcome for this type of design. As with other quasi-experimental designs, the pattern of outcome, design and context must be considered when seeking to interpret the results of a particular experiment.

There is a superficial similarity between the graphs obtained with this design and those for the time-series design. Note, however, that, whereas the latter shows time on the horizontal axis of the graph, the regression discontinuity design has pre-test scores along this axis. The issues are in both cases about trends in the data: are they present, do they differ and so on. 'Eyeballing' the data, i.e. visual inspection to assess this, forms a valuable part of the analysis, although most would argue that this needs to be supplemented by more formal statistical analysis. Conceptually the analyses for the two designs are equivalent, although different statistical techniques have to be used.

📖 The website gives examples of studies using different quasi-experimental designs.

Concluding thoughts on quasi-experimental designs

Quasi-experimentation is more of a style of investigation than a slavish following of predetermined designs. The designs covered above should be seen simply as suggestions for starting points. If you are not in a position to do true experiments, then with sufficient ingenuity you ought to be able to carry out a quasi-experiment to counter those threats to internal validity that are likely to be problematic.

Single-case experiments

A distinctive approach to carrying out experiments originated in the work of B. F. Skinner (e.g. Skinner, 1974). It has subsequently been developed by his followers, variously known as Skinnerians, Radical Behaviourists, or Operant Conditioners – among other labels. Sidman (1960) has produced a very clear, though partisan, account of this approach concentrating on the methodological issues and strategies involved. The work of Skinner arouses strong passions and, in consequence, his approach to experimental design tends to be either uncritically adopted or cursorily rejected.

There is much of value here for the real world investigator with a leaning to the experimental – mixed, as in Skinner's other work, with the unhelpfully polemical, the quirky and the rather silly. The approach is variously labelled, commonly as 'small-N', 'single-subject', or 'single-case' designs. This latter has the virtue, which would probably have been resisted by Skinner, of making the point that the 'case' need not necessarily be the individual person – it could be the individual school class, or the school itself, for example. It does, however, carry the possibility of confusion with 'case study', which as defined in this book is a multi-method enterprise (though this may incorporate a single-case experiment within it). Skinner's search was for a methodology which produced meaningful, reliable data at the level of the individual – and which didn't require statistical testing to decide whether or not an effect was present. His view was that 'eyeballing' the data was all that should be needed. If such visual inspection did not produce clear results then, another, better, experiment should be designed, continuing until clear, unambiguous results are obtained. However, there is the difficulty that what may be viewed as a clear result by one person may not be as clear to others. Which takes us back to using statistics, albeit ones appropriate for single-case data. Barlow *et al.* (2006) provide detailed discussions of the design and analysis (including statistical analysis) of single-case experiments. See also Chapter 16 p. 461.

Box 5.12 gives an overview of several types of single-case design.

BOX 5.12

Overview of a range of single-case designs

Note: These designs call for a series of measures on a dependent variable (DV) (or, more rarely, on two or more such variables). Typically the study is repeated with a small number of participants to establish the replicability of the findings.

1. A–B designs

Baseline phase (A) of a sequence of observations prior to intervention followed by a second phase where the intervention is introduced (B) and a further sequence of observations. Effectiveness of intervention shown by difference in observations made in B from those made in A (note similarity to interrupted time-series design).

2. A–B–A designs

As A–B but adding a third phase which reverts to pre-intervention baseline condition (A).

3. A–B–A–B designs

Addition of a second intervention phase (B) to A–B–A design. Avoids possible ethical problems of finishing with a return to baseline.

4. Multiple baseline designs

(a) *Across settings.* A DV is measured or observed in two or more situations (e.g. at home and at school). Change is made from a baseline condition (A) to the intervention (B) at different times in the different settings.

(b) *Across behaviours.* Two or more behaviours are measured or observed. Change is made from a baseline condition (A) to the intervention (B) at different times for the different behaviours.

(c) *Across participants.* Two or more participants are measured or observed. Change is made from a baseline condition (A) to the intervention (B) at different times for the different behaviours.

Additional phases can be added leading to multiple baseline versions of the A–B–A and A–B–A–B designs.

A–B design

The simplest A–B design involves two experimental conditions (the terminology is different from that used in the previous designs of experiments, but is well established). The first condition (A) is referred to as the *baseline*; the second condition (B) corresponds to the treatment. Both conditions are 'phases' which extend over time, and a sequence of tests or observations will be taken in each phase.

The investigator looks for a clear difference in the pattern of performance in the two phases – this being an actual 'look' as typically the data are 'eyeballed' by Skinnerians with their principled antipathy to statistical analysis. A distinctive feature of the Skinnerian approach is that the baseline phase is supposed to be continued until stability is reached, that is, so that there is no trend over time. In practice, this is not always achieved. The restriction to a stable baseline obviously assists in the interpretation of the data, but even so, the design is weak and subject to several validity threats (e.g. history–treatment interaction). *Because of this the design is probably best regarded as 'pre-experimental', with the same strictures on its use as with the other pre-experimental designs considered in the preceding section on quasi-experiments (p. 109).* The design can be strengthened in ways analogous to those employed in quasi-experimental design – effectively extending the series of phases either over time, or cross-sectionally over different baselines.

It is also a pragmatic point as to whether the necessary baseline stability can be achieved, although Skinnerians would consider it an essential feature of experimental control that conditions be found where there is stability. As with lengthy time-series

designs, this approach presupposes an observation or dependent variable where it is feasible to have an extended series of measures. Skinnerians would insist on the dependent variable being *rate of response* but this appears to be more of a historical quirk than an essential design feature.

A–B–A design

This improves upon the previous design by adding a *reversal* phase – the second A phase. The central notion is that the investigator removes the treatment (B) and looks for a return to baseline performance. Given a stable pre-treatment baseline, a clear and consistent shift from this during the second phase, and a return to a stable baseline on the reversal, the investigator is on reasonably strong ground for inferring a causal relationship.

The problems occur when this does not happen, particularly when there is not a return to baseline in the final phase. This leaves the experimenter seeking explanations for the changes that occurred during the second phase other than that it was caused by the treatment (B); or evaluating other possible explanations for the failure to return, such as carry-over effects from the treatment. The design is also open to ethical objections, particularly when used in an applied setting. Is it justifiable deliberately to remove a treatment when it appears to be effective? This is not too much of an issue when the goal is to establish or demonstrate some phenomenon, but when the intention is to help someone, many practitioners would have reservations about the design.

A–B–A–B design

A simple, though not entirely adequate, answer to the ethical problems raised by the preceding design is to add a further treatment phase. In this way the person undergoing the study ends up with the – presumed beneficial – treatment. All additions to the sequence of baseline and treatment phases, with regular and consistent changes observed to be associated with the phases, add to one's confidence about the causal relationship between treatment and outcome. There is no reason in principle why this AB alternation should not continue as ABABAB – or longer. However, it does involve extra time and effort which could probably be better spent in other directions. The design does also still call for the treatment to be withdrawn during the sequence, and there are alternative designs which avoid this.

Multiple baseline designs

The approach in this design involves the application of the treatment at different points in time to different baseline conditions. If there is a corresponding change in the condition to which the treatment is applied, and *no change in the other conditions* at that time, then there is a strong case that the change is causally related to the treatment.

Three versions of the design are commonly employed: multiple baselines *across settings*, *across behaviours* and *across participants*.

- In the across settings design, a particular dependent variable (behaviour) of a participant is monitored in a range of different settings or situations and the treatment is introduced at a different time in each of the settings.
- In the across behaviours design, data is collected on several dependent variables (behaviours) for a particular participant and the treatment is applied at different times to each of the behaviours.
- In the across participants design, data is collected on a particular baseline condition for several participants and the treatment is applied at different times to the different participants.

The general approach is illustrated in Figure 5.4.

Other designs have been used and are briefly explained here.

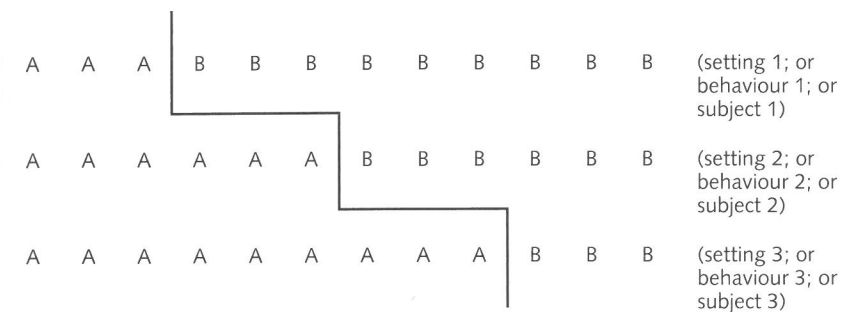


Figure 5.4: The multiple baseline design.

Changing criterion designs

A criterion for performance is specified as a part of the intervention. That criterion is changed over time in a pre-specified manner, usually progressively in a particular direction. The effect of the intervention is demonstrated if the behaviour changes to match the changes in criterion.

This design is attractive in interventions where the intention is to achieve a progressive reduction of some problem behaviour or progressive increase in some desired behaviour. It is probably most useful for situations involving complex behaviours or where the intention is to try to achieve some major shift from what the person involved is currently doing. Certainly the notion of 'successive approximations' which is built in to this design sits very naturally with the 'shaping of behaviour' approach central to Skinnerian practice and applicable by others.

Multiple treatment designs

These involve the implementation of two or more treatments designed to affect a single behaviour. So, rather than a treatment being compared with its absence (which is effectively what baseline comparisons seek to achieve), there are at least two separate

treatments whose effects are compared. In its simplest form, this would be ABC, i.e. a baseline condition (A), followed in sequence by two treatment conditions (B and C). This could be extended in several ways: ABCA or ABACA or ABACABCA, etc. The latter gives some kind of assessment of 'multiple treatment interference' – the extent to which there are sequence effects, where being exposed to one condition has a subsequent influence on the apparent effect of a subsequent condition.

There are several more complex variants. In one (known as a *multiple schedule design*), each treatment or intervention is associated in a consistent way, probably for a substantial number of times, with a particular 'stimulus' (e.g. a particular person, setting or time) so that it can be established whether or not the stimulus has consistent control over performance. An alternative (known variously as a *simultaneous treatment* or *alternating treatment design*) is for each of the settings to be balanced across stimulus conditions (persons, settings or times) so that each of the settings has been associated equivalently with each of the stimuli. This then permits one to disentangle the effects of the settings from 'stimulus' effects. Barlow *et al.* (2006) give details.

These designs have several advantages. As the main concern is for differential effects on the two conditions, the establishment of a stable baseline becomes less crucial. In the two latter variants, there is no need for treatment to be withdrawn or removed, and the relative effects of the different conditions can be determined without the need for lengthy successive involvement with different phases. It is possible to generate what might be called 'combined designs' by putting together features from the individual designs considered above. For example an ABAB design could be combined with a multiple baseline approach.

The general approach taken to design in single-case experimentation bears some similarities to that taken in quasi-experimentation. There is concern for specific threats to validity which might make the study difficult or impossible to interpret, particularly in relation to causality. These are taken into account in developing the design. However, there is often substantially greater flexibility and willingness to modify the design than is found in other types of experimentation. It is common to review and possibly alter the design in the light of the pattern of data which is emerging. Decisions such as when to move from one phase to another are made in this way. The attraction of a combined design is that additional or changed design features (which can counter particular threats to validity) can be introduced reactively to resolve specific ambiguities.

This approach is foreign to the traditional canons of fixed design research, where a design is very carefully pre-planned and then rigidly adhered to. Interestingly, it has similarities to the flexible designs discussed in the next chapter.

There are also similarities between single-case experimentation and the way in which experiments are carried out in some branches of the natural sciences. Statistics play little part; the trick is to set up the situation, through control of extraneous variables that would cloud the issue (Ernest Rutherford, the eminent physicist, is widely quoted as saying 'If your experiment needs statistics, you ought to have done a better experiment'). By doing this, the cause-effect relationship shines out for all to see, in realist terms ensuring the context is such that the mechanism(s) operate. It may be necessary to 'fine-tune' your study so that unforeseen eventualities can be accounted for. With sufficient experimental skill and understanding you should be able to find out something of importance about the specific focus of your study (whether this happens to be a person

or a radioactive isotope). You will need to test it out on a few individuals just to assure yourself of typicality – which was not only Skinner's approach but also that of the supreme experimentalist Ivan Pavlov.

 *The website gives examples of a range of studies using single case designs.*

Passive experimentation

Several types of study which appear to be very similar to experiments do not have the *active manipulation of the situation by the experimenter* taken here to be a hallmark of the experimental approach. They are sometimes called passive experiments but are discussed here under the heading of non-experimental fixed designs.

Non-experimental fixed designs

If, following your reading of the previous chapter, it appears possible that a non-experimental fixed design may be appropriate for your project and its research questions, then perusal of this section should help in choosing a specific non-experimental design. However, before confirming that choice, it will be necessary to read the chapters in Part III of this book to help select appropriate methods of collecting data (the survey is a very common type of non-experimental fixed design and hence Chapter 10 is particularly relevant), and Chapter 18 to establish how you will analyse the data after it has been collected.

This style of fixed design research differs from the experimental one in that the phenomena studied are not deliberately manipulated or changed by the researcher. Hence it is suitable in situations where the aspects of interest are not amenable to such changes for whatever reason. These include variables or characteristics which:

- are not modifiable by the researcher (e.g. personal characteristics such as gender, age, ethnicity);
- should not be modified for ethical reasons (e.g. tobacco smoking, alcohol consumption); or
- it is not feasible to modify (e.g. placement in a school or classroom).

As these constraints apply to a high proportion of the variables of likely interest in applied social research, these designs are of considerable importance to anyone intent on carrying out a fixed design study.

Dealing with things as they are, rather than as modified by the experimenter, has the advantage of not disturbing whatever it is that we are interested in. Byrne (2002) stresses that 'It is important to realize that surveys are not experiments. They are fundamentally different and in my view much superior' (p. 62) – note, however, that he presents a radical critique of the traditional variable-oriented approach as presented here.

Non-experimental fixed designs are commonly used for descriptive purposes, and, because of their fixed, pre-specified nature are not well adapted to exploratory work. They can be used when the interest is in explaining or understanding a phenomenon. The designs have been rarely used by researchers with emancipatory concerns but there is no reason why, say, a descriptive study could not be useful for anti-discriminatory or other emancipatory purposes.

Working within the realist framework they can help establish cause in the sense of providing supportive evidence for the operation of mechanisms and for teasing out the particular situations and groups of people where enabling or disabling mechanisms have come into play. Within the realist approach, the specification of which mechanisms have operated in a *post-hoc* (i.e. after the study has taken place) manner is viewed as entirely legitimate. While prediction of the exact pattern of results may not be feasible because of the open nature of the systems which social research studies, this does not preclude an explanation of the particular pattern obtained.

Measuring relationships

Non-experimental fixed designs are commonly used to measure the relationship between two or more variables. Do pupils from different ethnic backgrounds achieve differently in schools? Are there gender differences? What is the relationship between school characteristics and student achievement? They are sometimes referred to as *correlational studies*. However, this tends to suggest that a particular statistical technique (the correlation coefficient) is to be used whereas there is a range of possibilities for analysis which is discussed in Chapter 16.

The traditional approach is to start with a conceptual framework or other approach to theory, in some provisional form at least. In realist terms, this means having a pretty clear idea (e.g. from a previous study or from other sources) of likely mechanisms and the contexts in which they will operate. This theory is then used to identify the variables and possible relationships to be studied. Research questions are formulated prior to data collection. Similarly, decisions about the methods of data collection and analysis, and the sampling strategy determining who will be asked are all finalized before data collection proper starts. And it is expected that, when these decisions have been made, they are kept to throughout the study.

When all measures are taken at the same time (or, in practice, over a relatively short period of time), it is commonly referred to as a *cross-sectional study*, and is widely used in fixed design social research. It is often employed in conjunction with the *survey method* of data collection discussed in detail in Chapter 10. The pattern of relationships between variables may be of interest in its own right, or there may be a concern for establishing causal links. In interpreting the results of these studies, statistical and logical analyses effectively take the place of the features of experimental design which facilitate the interpretation of true experiments.

The variables to be included in the study are those needed to provide answers to your research questions. These questions will, as ever, be governed by the purposes of your study and by the theory or conceptual structure you have developed. Whereas the tradition in experimental research is to label these as independent variables (those which

the experimenter manipulates) and dependent variables (those where we look for change), they are usually referred to here as *explanatory variables* and *outcome variables*, respectively.

It is possible to include more explanatory variables in this design than is feasible in experimental designs. However, this should not be taken as an excuse for a 'fishing trip' – just throwing in variables in the hope that something will turn up. To reiterate the principle – the variables are included because of their relevance to your research questions.

The choice of participants to make up the group is important. Again, your research questions effectively determine this. An issue is the homogeneity of the group. For example, while you may not be interested in this study in gender issues in themselves, you may consider it important to have both males and females in the group. It may be, however, that males and females are affected by different variables and hence there will be increased variability on the outcome variable. A solution here is to analyse the two genders separately; i.e. to perform a subgroup analysis.

When decisions have been made about the composition of the group, data can be collected. Typically quantitative, or quantifiable, data are collected often using a questionnaire, or some type of test or scale for both explanatory and outcome variables (see Chapters 10 and 12, respectively).

Analysis and interpretation

A variety of data analysis techniques can be used, including (but by no means limited to) correlational analysis. Chapter 16, pp. 462–3, gives examples. The techniques can be thought of as providing statistical control for factors controlled for in true experiments by employing a control group and randomization of allocation of participants to experimental and control groups. So, for example, a survey on attitudes to nuclear power might show gender differences. Frankfort-Nachmias and Nachmias (1996, p. 126) discuss an example from Solomon, Tomaskovic-Devey and Risman (1989) where 59 per cent of men and 29 per cent of women support nuclear power. This analysis of differences in percentages (produced by 'cross-tabulation'; see Chapter 16, p. 418) provides a kind of statistical equivalent to an experimental design which is of course not feasible due to the impossibility of randomly assigning individuals to be males or females!

Establishing the statistical significance of the relationship between gender and attitude does not enable us to conclude that these variables are causally related. Nor does it, in itself, help in understanding what lies behind this relationship. In realist terms, we need to come up with plausible mechanisms and seek evidence for their existence. As Frankfort-Nachmias and Nachmias suggest, it may be that 'the women in the study may have been less knowledgeable about technological matters and therefore more reluctant to support nuclear power. Or perhaps women's greater concern with safety would lead them to oppose nuclear power more than men' (p. 127).

Further subdivision of the group allows one to control statistically for the effects of such variables providing the relevant information (e.g. degree of knowledge in the above example) has been collected. Various statistical techniques such as path analysis can be used to analyse the data in more detail (Chapter 16, p. 443). However, the interpretation,

particularly in relation to causation, remains a challenging task and is an amalgam of theoretical, logical and statistical considerations.

Making group comparisons

An alternative approach to seeking relationships in a single group is to have a second group and make comparisons between the two groups. The groups may be naturally occurring ones already in existence or may be created especially for the study. Group selection raises the same issues about threats to internal validity as in experimental (particularly quasi-experimental) design. The threat of *differential selection* will arise if the groups differ in some ways other than those indicated by the *explanatory variable(s)*. Assuming random allocation to the two groups is not feasible, some other approach must be taken to guard against this threat. Possibilities include:

- *matching* on variables likely to be relevant (see p. 105);
- *using a statistical method of control* for existing differences (see Chapter 16; p. 463);
- *using direct control* e.g. by only selecting participants from a particular ethnic or socioeconomic background; and/or
- *analysing subgroups*.

Measures will typically be made on other background or control variables (e.g. ethnicity) which may be of interest in their own right or as helping to understand what lies behind any differences found. They could then form the basis for subsequent subgroup analyses.

While the relational and group comparison approaches may seem very different, the difference is largely in the way the study is conceptualized. The example of gender differences in attitude to nuclear power was discussed above as a single group containing both males and females. The focus was on relations between gender and attitude. It could be viewed as based on separate groups of males and females. The focus is then on differences between the two groups.

Gender is a dichotomous variable (i.e. it only has two possible values). Other explanatory variables can take on a wide range of values. When they are involved, instead of talking about separate comparison groups differing on the variable of interest, there is a wide range of differences on that variable. Separate and distinct comparison groups (two or more) can be thought of as a special case of that general situation. This approach effectively brings together both relational and comparative studies within the same framework. It also paves the way to the use of statistical techniques such as analysis of variance (and particularly in this context analysis of covariance which permits the separation out of the effects of control variables) and multiple linear regression.

Category membership

The identification of membership of particular categories or groups in non-experimental fixed design research can be difficult and complex. This may appear straightforward for

a variable such as age. However, there are situations where even age can be problematic, say, in a study of the drinking habits of young adults or in cultures where date of birth may not be recorded or remembered. Gender, as a social construct, can raise category difficulties in particular cases. An area such as ethnicity bristles with complexities particularly in multi-cultural societies (see Stanfield, 1993).

In flexible design research where participant numbers are almost always small, it is usually feasible to achieve the depth needed to deal with the complexities but fixed design research, typically with much larger numbers, necessarily has to simplify. Notwithstanding the difficulties, it would be unfortunate if this put important topics out of bounds. As Mertens (2005) comments, 'Discontinuing such research based on the rationale that our understanding of race, gender, and disability is limited needs to be weighed against the benefit associated with revealing inequities in resources and outcomes in education, psychology, and the broader society' (p. 150).

Classification of non-experimental fixed designs

This is a murky area where a plethora of terms have been used including comparison and causal-comparative designs, correlational designs, natural experiments, ex post facto, available data designs, descriptive designs and surveys. Johnson (2001) has suggested that many of the designs used can be classified in terms of the purpose, or research objective, of the study as one aspect, and the time dimension as a second one.

Purpose (research objective)

The main non-experimental fixed designs are *descriptive, predictive* or *explanatory*:

- Descriptive designs are primarily concerned with describing something; with documenting its characteristics.
- Predictive designs are primarily concerned with predicting or forecasting some event or phenomenon in the future.
- Explanatory designs are primarily concerned with developing or testing a theory about something; to identify the causal factors or mechanisms producing change.

'Primarily' flags that a project may have more than one objective. Primarily descriptive and predictive designs often attempt some explanation of what is happening.

Time dimension

Three types are commonly found, *cross-sectional, longitudinal* and *retrospective* designs:

- In cross-sectional designs, the data are collected at a single point in time (practical considerations may extend this to a relatively brief period rather than a single point).
- In longitudinal designs, the data are collected at more than one point in time or brief period. Many such designs involve an extended series of data collections (as in experimental time-series designs, discussed earlier in the chapter, p. 114) but

some just involve two data collections. Subtypes include *trend designs* (independent samples of participants at the different data collections) and *panel designs* (same participants).

- In retrospective designs, the researcher collects data at a point in time about the situation at some earlier point in time as well as the current situation (e.g. by asking questions about earlier behaviour). This is effectively a special case of a cross-sectional design which attempts to simulate a longitudinal design and obtain data relating to more than one point in time.

Longitudinal panel designs can avoid many of the difficulties in interpretation and threats to internal validity of much non-experimental fixed design research discussed above (see also Ruspini, 2002 and Menard, 2007). However, they are difficult and complex to run and typically call for considerable resources. Some of the problems include sample attrition (when participants are lost to the study), the need to devise measures which can be used repeatedly and for special methods of data analysis (see Menard, 2007).

✦ The website gives examples of studies using cross-sectional, longitudinal and retrospective designs.

Sample size in fixed designs

One of the most common questions asked by a novice researcher is 'What size of sample do I need?' The answer is not straightforward as it depends on many factors. In some real world research, the question is answered for you by the situation. You may be working in an organization where the obvious thing is to survey everybody or your resources may be so stretched that this sets the limits on the number of participants you can deal with. In such circumstances, it is particularly important to have thought through how your data are to be analysed before proceeding. There are minimum numbers for statistical tests and procedures below which they should not be used. Hence, if you plan to use a particular test or procedure then this sets minimum numbers for your design.

Following Borg and Gall (1989), Mertens (2005, p. 325) suggests 'rule-of-thumb' figures of about 15 observations per group for experimental, quasi-experimental and non-experimental designs involving group comparisons and about 30 observations for non-experimental designs involving relations in a single group. In survey research, which typically seeks to incorporate more variables than experimental and other non-experimental designs, they recommend somewhat larger numbers (see Chapter 10, p. 271). Cohen (1992) provides a similar set of rules for sample sizes for a range of commonly used statistical tests.

When the main interest, as in many surveys, is to generalize the findings to the population from which the sample is drawn, then issues such as the *homogeneity* of the population are important. If pilot work establishes considerable *heterogeneity*, this then indicates the need for a larger sample. Similarly, the more accurate you want the estimates from your study to be, the larger a sample is needed. There are statistical

techniques for determining the relationship between sampling error and sample size. Formulae have been developed to assist in the choice of an efficient sample size when it is important to limit estimation errors to a particular level. Henry (1990, Chapter 7) and Czaja and Blair (2005, pp. 142–8) give introductions. Lipsey (1990) provides a more detailed account and a useful general treatment of power analysis which covers the factors that affect the sensitivity of a design in detecting relationships (see Chapter 16, p. 448). Clark-Carter (1997) provides power tables to help in choosing sample sizes. They should be treated with care, as you need to be clear about the assumptions on which they are based. This is a matter on which it is advisable to seek assistance from a statistician if it is of importance in your study.

Further reading

✦ The website gives annotated references to further reading for Chapter 5.

CHAPTER 6

Flexible designs

This chapter:

- reiterates the rationale for referring to 'flexible' rather than the more usual 'qualitative' research design;
- covers some general features of flexible design research and the researcher qualities it calls for;
- concentrates on three traditions of flexible design research: case studies, ethnographic studies and grounded theory studies;
- briefly reviews a range of other possible approaches including the narrative, phenomenological, hermeneutic and feminist research traditions;
- discusses how to determine sample size in flexible designs; and
- concludes by considering the place of reliability and validity in this type of research, and the ways in which researcher bias and threats to validity can be dealt with.

Introduction

It is now considered respectable and acceptable in virtually all areas of social research (including applied fields such as education, health, social work, and business and management) to use designs based largely or exclusively on methods generating qualitative data. There are still some isolated outposts, particularly in fields abutting medicine and in areas of experimental psychology, where such designs are considered illegitimate or inferior to traditional quantitative designs. Other areas, such as educational research in the UK, appeared for some time to be in danger of espousing a new

orthodoxy where anything other than qualitative research was viewed as deviant, but recent governmental support for quantitative evidence-based approaches in both the US and UK has had its effects.

The position taken in this text is that for some studies and for certain types of research question, research largely or exclusively based on qualitative data is indicated. For others, research largely or exclusively based on quantitative data is needed. The qualitative/quantitative ways of labeling research *designs* are so well established that it risks miscommunication not to use them. However, as already pointed out in Chapter 2, their use is not entirely logical. In principle (and not uncommonly in practice) so-called qualitative designs can incorporate quantitative methods of data collection. All of these approaches show substantial flexibility in their research design, typically anticipating that the design will emerge and develop during data collection. As discussed in the previous chapter, so-called quantitative approaches call for a tight pre-specification of the design prior to data collection. Hence my preference for referring to them as *flexible* and *fixed designs*, respectively.

It is worth stressing that whereas fixed design research is, typically, an 'off-the-shelf' process where the task is primarily one of choice from a range of well-defined alternative designs, flexible designs are much more 'do-it-yourself'. Although there are research traditions in flexible design research, covered in the chapter, your task is much more that of constructing a one-off design likely to help answer your research questions.

The following two chapters also cover designs likely to include some flexible elements. The multi-strategy designs in Chapter 7 are ways in which substantial fixed and flexible design aspects can be incorporated in the same study. The 'designs for particular purposes' in Chapter 8 run the whole gamut from strictly fixed experimental designs common in some types of evaluation research to the highly flexible designs typical of action research.

General features of flexible designs

We first provide a general specification for the design of a flexible study, followed by accounts of three influential design traditions within flexible design research which are commonly used for real world studies: *case studies*, *ethnographic studies* and *grounded theory studies*.

The chapter also gives information about a range of other traditions within qualitative research which may be worth considering for some real world projects. Box 6.1 gives a flavour of the kind of characteristics to be found in a flexible design where serious attention has been given to the general norms and canons of this style of research. This should be seen in the context of the overall design framework developed in Chapter 4 (Figure 4.1; p. 71). In other words, thought and attention will have to be given to the *purpose(s)* of your research, to its *conceptual structure*, to the *research questions* to which you seek answers, to the *methods* of data collection and the *sampling strategy* which will be needed to get these answers.

BOX 6.1

Characteristics of a 'good' flexible design

1. Typically multiple qualitative data collection techniques (possibly some quantitative data also – if a substantial amount of quantitative data collection this becomes a multi-strategy design, see Chapter 7) are used. Data are adequately summarized (e.g. in tabular form). Detail is given about how data are collected.
2. The study is framed within the assumptions and characteristics of the flexible design approach to research. This includes fundamental characteristics such as an evolving design, the presentation of multiple realities, the researcher as an instrument of data collection, and a focus on participants' views.
3. The study is informed by an understanding of existing traditions of research, i.e. the researcher identifies, studies and employs one or more traditions of enquiry.
4. This tradition need not be 'pure', and procedures from several can be brought together. The beginning researcher is recommended to stay within one tradition initially, becoming comfortable with it, learning it and keeping a study concise and straightforward. Later, especially in long and complex studies, features from several traditions may be useful.
5. The project starts with a single idea or problem that the researcher seeks to understand, not a causal relationship of variables or a comparison of groups (for which a fixed design might be indicated). Relationships might evolve or comparisons might be made, but these emerge later in the study.
6. The study shows a rigorous approach to data collection, data analysis and report writing. The researcher has the responsibility of verifying the accuracy of the account given.
7. Data are analysed using multiple levels of abstraction. Often, writers present their studies in stages (e.g. multiple themes that can be combined into larger themes or perspectives) or layer their analyses from the particular to the general.
8. The writing is clear, engaging and helps the reader to experience 'being there'. The story and findings become believable and realistic, accurately reflecting the complexities of real life.

(after Creswell, 1998, pp. 20–2)

You don't try to get all of this cut-and-dried before starting data collection. The purpose or purposes of the study are likely to be pretty clear from the outset. However, at this stage, you may not have much of an idea about what theoretical framework is going to be most helpful. Indeed, one version of the grounded theory tradition discussed below argues that you should seek to enter the field without theoretical preconceptions (something which many would regard as impossible to achieve). It is highly likely that your research questions will be initially underdeveloped and tentative. You

obviously need to make some early decisions about methods of data collection, because if you don't, you never get started. However, in these designs you don't have to foreclose on options about methods. Ideas for changing your approach may arise from your involvement and early data collection. Or, as you change or clarify the research questions, different means of data collection may be called for. Similarly, your sampling of who, where and what does not have to be decided in advance. Again, you need to start somewhere but the sampling strategy can and should evolve with other aspects of the design.

Realism and flexible design

Within the realist framework, it is held that theory, rather than data or the methods used to produce that data, is central to explaining reality. This is fully consonant with the view developed in this text that it is the research questions which drive the design of a study, whether it be flexible or fixed. And that these questions have to be linked to theory, whether pre-existing, which is tested by the research, or generated by the process of the research itself.

Hence a realist view has no problems with flexible design research or with the use of qualitative data. As pointed out by Anastas (1999):

Flexible or qualitative methods have traditionally included the researcher and the relationship with the researched within the boundary of what is examined. Because all any study can do is to approximate knowledge of phenomena as they exist in the real world (fallibilism) the process of study itself must be studied as well. Because all methods of study can produce only approximations of reality and incomplete understanding of the phenomena of interest as they exist in the real world, the findings of flexible method research can be seen as no more or less legitimate than those of any other type of study (p. 56).

Researcher qualities needed for flexible design research

Doing flexible design research calls for flexible researchers. This approach to research makes great demands on the researcher *while carrying out the study*. It is commonly said that it involves the 'researcher-as-instrument', i.e. that rather than being able to rely on specialist tools and instruments, you to a large extent have to do it all yourself. Certainly the quality of a flexible design study depends to a great extent on the quality of the investigator. It is not a 'soft' option in the sense that anyone can do it without preparation, knowledge of procedures or analytical skills. It *is* soft, however, in the sense that there are few 'hard and fast' routinized procedures, where all you have to do is to follow the formulae. This makes life harder rather than easier – though also more interesting.

Ideally this kind of research calls for well-trained and experienced investigators but other aspects are also important. Personal qualities such as having an *open and*

enquiring mind, being a 'good listener', general sensitivity and responsiveness to contradictory evidence are needed. These are commonly regarded as skills central to the professional working with people in whatever capacity. Relevant professional experience of this kind is also likely to provide you with a firm grasp of the issues being studied in a particular study.

The professional or practitioner working with people as their job has much to contribute both *as* an investigator and *to* an investigator. As an investigator, probably carrying out 'insider' research, she will need a firm grasp of the material in this book and experience (this can lead to 'Catch-22' problems – how do you get the experience without carrying out a study, and vice versa). Working in collaboration with someone who has the methodological skills and the experience is obviously one way forward. Box 6.2 tries to provide an indication of the skills needed to be an effective flexible design researcher.

BOX 6.2

General skills needed by flexible design investigators

1. *Question asking.* Need for an 'enquiring mind'. Your task includes enquiring why events appear to have happened or to be happening. This is something you ask yourself as well as others and is mentally and emotionally exhausting.
2. *Good listening.* Used in a general sense to include all observation and sensing, not simply via the ears. Also 'listening' to what documents say. *Good* means taking in a lot of new information without bias; noting the exact words said; capturing mood and affective components; appreciating context. You need an open mind and a good memory. (Taping helps but is not a panacea.)
3. *Adaptiveness and flexibility.* These studies rarely end up exactly as planned. You have to be willing to change procedures or plans if the unanticipated occurs. The full implications of any changes have to be taken on board, e.g. you may need to change the design. There is a need to balance *adaptiveness* and *rigour*.
4. *Grasp of the issues.* The investigator needs to *interpret* information during the study, not simply record it. Without a firm grasp of the issues (theoretical, policy etc.), you may miss clues, not see contradictions, overlook the requirement for further evidence, etc.
5. *Lack of bias.* The preceding skills are negated if they are simply used to substantiate a preconceived position. Investigators should be open to contrary findings. During data collection, preliminary findings should be submitted to critical colleagues who are asked to offer alternative explanations and suggestions for data collection. See the discussions on researcher bias at various points in the chapter.

BOX 6.3

Overview of three approaches to flexible design research

1. *Case study.* A well-established research strategy where the focus is on a case (which is interpreted very widely to include the study of an individual person, a group, a setting, an organization etc.) in its own right, and taking its context into account. Typically involves multiple methods of data collection. Can include quantitative data, though qualitative data are almost invariably collected.
2. *Ethnographic studies.* Another well-established strategy where the focus is on the description and interpretation of the culture and social structure of a social group. Typically involves participant observation over an extended period of time, but other methods (including those generating quantitative data) can also be used.
3. *Grounded theory studies.* A more recently developed strategy where the main concern is to develop a theory of the particular social situation forming the basis of the study. The theory is 'grounded' in the sense of being derived from the study itself. Popular in research on many applied settings, particularly health-related ones. Interviews are commonly used but other methods (including those generating quantitative data) are not excluded.

Research traditions in qualitative research

Box 6.3 provides an overview of the main approaches to flexible design research featured in this chapter.

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. Gerring (2006) gives references to case studies in areas as disparate as anthropology, archaeology, business studies, education, international relations, marketing, medicine, organizational behaviour, politics, psychology, public administration, public health, social work and sociology. The varying strategies developed for dealing with cases in different disciplines have useful lessons, suggesting solutions to problems with case study methodology, including the thorny one of generalizing from the individual case. There is some danger in using a well-worn term like case study. 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. Following the lead set by Robert Yin (2009), who has done much to resuscitate case study as a serious option when doing social research:

Case study is a strategy for doing research which involves an empirical investigation of a particular contemporary phenomenon within its real life context using multiple sources of evidence.

The important points are that it is:

- a *strategy*, i.e. a stance or approach, rather than a method, such as observation or interview;
- concerned with *research*, taken in a broad sense and including, for example, evaluation research;
- *empirical* in the sense of relying on the collection of evidence about what is going on;
- about the *particular*; a study of that specific case (the issues of what kind of generalization is possible from the case, and of how this might be done, are important);
- focused on a *phenomenon in context*, typically in situations where the boundary between the phenomenon and its context is not clear; and
- using *multiple methods* of evidence or data collection.

The central defining characteristic is concentration on a particular case (or small number of cases) studied in its own right. However, the importance of its context or setting is also worth highlighting. Miles and Huberman (1984, p. 27) suggest that in some circumstances the term 'site' might be preferable 'because it reminds us that a "case" always occurs in a specified social and physical *setting*: we cannot study individual cases devoid of their context in a way that a quantitative researcher often does'.

While some commentators see case studies as being essentially qualitative (e.g. Stake, 1995, 2005), it is now widely accepted (e.g. Gerring, 2006; Yin, 2009) that they can make use of both quantitative and qualitative data collection methods. However, it is relatively rare to see case studies where any quantitative component has anything other than a minor role (hence they are viewed here as flexible, rather than multi-strategy, designs).

Taking case study seriously

Valsiner (1986) claimed that 'the study of individual cases has always been the major (albeit often unrecognized) strategy in the advancement of knowledge about human beings' (p. 11). In similar vein, Bromley (1986) maintained that 'the individual case study

or situation analysis is the bed-rock of scientific investigation' (p. ix). But he also notes, in an unattributed quotation, the common view that 'science is not concerned with the individual case' (p. xi). These widely divergent claims betray a deep-rooted uncertainty about the place and value of studying cases.

Case study was until recently commonly considered in methodology texts as a kind of 'soft option', possibly admissible as an exploratory precursor to some more 'hard-nosed' experiment or survey or as a complement to such approaches but of dubious value by itself. Campbell and Stanley (1963) presented an extreme version of this view:

Such studies often involve tedious collection of specific detail, careful observation, testing and the like, and in such instances involve the error of misplaced precision. How much more valuable the study would be if the one set of observations were reduced by half and the saved effort directed to the study in equal detail of an appropriate comparison instance. It seems well-nigh unethical at present to allow, as theses or dissertations in education, case studies of this nature (p. 177).

However, Campbell subsequently recanted and in later publications (e.g. Cook and Campbell, 1979) viewed case study as a fully legitimate alternative to experimentation in appropriate circumstances. They make the point that 'case study as normally practiced should not be demeaned by identification with the one-group post-test-only design' (p. 96). Their central point is that *case study is not a flawed experimental design; it is a fundamentally different research strategy with its own designs*.

It is useful to separate out criticisms of the practice of particular case studies from what some have seen as inescapable deficiencies of the strategy itself. As Bromley (1986) points out, 'case studies are sometimes carried out in a sloppy, perfunctory, and incompetent manner and sometimes even in a corrupt, dishonest way' (p. xiii). Even with good faith and intentions, biased and selective accounts are undoubtedly possible. Similar criticisms could be made about any research strategy, of course. The issue is whether or not it is possible to devise appropriate checks to demonstrate the trustworthiness of the findings (see the discussion below, p. 154).

Can case study be scientific?

As discussed in Chapter 2, the positivist 'standard view' of science (which found case study problematic) has been comprehensively demolished, although its ghostly presence lingers on in the views and practices of many quantitatively inclined social researchers. Case study does not appear to present any special difficulties for the realist view of science, developed and defended in that chapter (see Box 2.7, p. 31). The study of the particular, which is central to case study, is not excluded in principle; it is the aims and intentions of the study, and the specific methods used, that have to concern us. Carr and Kemmis (1986) reach very similar conclusions: 'What distinguishes scientific knowledge is not so much its logical status, as the fact that it is the outcome of a process of enquiry which is governed by critical norms and standards of rationality' (p. 121).

Designing case studies

The 'case' can be virtually anything. The individual person as the case is probably what springs first to mind. A simple, single case study would just focus on that person, perhaps in a clinical or medical context where the use of the term case is routine. More complex, multiple case studies might involve several such individual cases. Case studies are not necessarily studies of individuals, though. They can be done on a group, on an institution, on a neighbourhood, on an innovation, on a decision, on a service, on a programme and on many other things. (There may be difficulties in defining and delimiting exactly what one means by the 'case' when the focus moves from the individual person.) Case studies are then very various. Box 6.4 gives some indication of different types and of the range of purposes they fulfil. Yin (2003, 2004) gives details of a wide range of different case studies.

BOX 6.4

Some types of case study

1. *Individual case study*. Detailed account of one person. Tends to focus on antecedents, contextual factors, perceptions and attitudes preceding a known outcome (e.g. drug user; immigrant). Used to explore possible causes, determinants, factors, processes, experiences etc., contributing to the outcome.
2. *Set of individual case studies*. As above, but a small number of individuals with some features in common are studied.
3. *Community studies*. Studies of one or more local communities. Describes and analyses the pattern of, and relations between, main aspects of community life (politics, work, leisure, family life etc.). Commonly descriptive, but may explore specific issues or be used in theory testing.
4. *Social group studies*. Studies of both small direct contact groups (e.g. families) and larger, more diffuse ones (e.g. occupational group). Describes and analyses relationships and activities.
5. *Studies of organizations and institutions*. Studies of firms, workplaces, schools, trades unions etc. Many possible foci, e.g. best practice, policy implementation and evaluation, industrial relations, management and organizational issues, organizational cultures, processes of change and adaptation etc.
6. *Studies of events, roles and relationships*. Focus on a specific event (overlaps with (3) and (4)). Very varied, includes studies of police-citizen encounters, doctor-patient interactions, specific crimes or 'incidents' (e.g. disasters), studies of role conflicts, stereotypes, adaptations.
7. *Cross-national comparative studies*. Used for research on local and national governments and the policy process.

(after Hakim, 2000, pp. 63–72)

Whatever kind of case study is involved (and the list in Box 6.4 only scratches the surface), there is always the need, as in any kind of research, to follow a framework for research design such as that given in Chapter 4 (Figure 4.1, p. 71). The degree of flexibility of design will vary from one study to another. If, for example, the main purpose is exploratory, trying to get some feeling as to what is going on in a novel situation where there is little to guide what one should be looking for, then your initial approach will be highly flexible. If, however, the purpose is confirmatory, where previous work has suggested an explanation of some phenomenon, then there is a place for some degree of pre-structure. There is an obvious trade-off between *looseness* and *selectivity*. The looser the original design, the less selective you can afford to be in data selection. Anything might be important. On the other hand, the danger is that if you *start* with a relatively tight conceptual framework or theoretical views, this may blind you to important features of the case or cause you to misinterpret evidence. There is no obvious way out of this dilemma. Practicalities may dictate some pre-structuring, for example, if the project is on a very tight time-scale, as in much small-scale contract research.

Holistic case studies

Yin (2009) differentiates between two versions of the single case study on the basis of the level of the unit of analysis. A study where the concern remains at a single, global level is referred to as *holistic*. This would typically (though not necessarily) be how a case study of an individual would be viewed but would also apply to, say, the study of an institution which remained at the level of the whole rather than seeking to look at and analyse the different functioning of separate sub-units within the institution.

Holistic case studies are appropriate in several situations. The *critical case* is a clear, though unfortunately rare, example. This occurs when your theoretical understanding is such that there is a clear, unambiguous and non-trivial set of circumstances where predicted outcomes will be found. Finding a case which fits, and demonstrating what has been predicted, can give a powerful boost to knowledge and understanding. This is the way in which experiment is used classically – the 'crucial experiment'. It is interesting to note that some of the most illustrious of this genre, for example, the verification of Einstein's theory of relativity by measuring the 'bending' of light from a distant star at a rare eclipse, are effectively case studies (being the study of a particular instance in its context) rather than experiments (in that no experimental manipulation of variables is possible).

The *extreme case* also provides a rationale for a simple, holistic case study. A former colleague of mine now features as a case in an orthopaedic textbook because of the virtually complete recovery he made from horrific arm and leg injuries in a cycle crash, after skilled surgical and physiotherapy support, together with his own determination, confounded initial gloomy predictions. More generally, the extreme and the unique can provide a valuable 'test bed' for which this type of case study is appropriate. Extremes include the 'if it can work here it will work anywhere' scenario, to the 'super-realization' where, say, a new approach is tried under ideal circumstances, perhaps to obtain understanding of how it works before its wider implementation.

Multiple case studies

In many studies it is appropriate to study more than a single case. A very common misconception is that this is for the purpose of gathering a 'sample' of cases so that generalization to some population might be made. Yin makes the useful analogy that carrying out multiple case studies is more like doing multiple experiments. These may be attempts at replication of an initial experiment or they may build upon the first experiment, perhaps carrying the investigation into an area suggested by the first study or they may seek to complement the first study by focusing on an area not originally covered. This activity, whether for multiple case studies or for multiple experiments (or for multiple surveys for that matter; or for multiple studies involving other research design strategies), is not concerned with statistical generalization but with what is sometimes referred to as *analytic* or *theoretical generalization*. The first case study will provide evidence which supports a theoretical view about what is going on; perhaps in terms of mechanisms and the contexts in which they operate. This theory, and its possible support or disconfirmation, guides the choice of subsequent cases in a multiple case study. Findings, patterns of data etc. from these case studies which provide this kind of support, particularly if they simultaneously provide evidence which does not fit in with alternative theories, are the basis for generalization.

Put simply, cases are selected where *either* the theory would suggest that the same result is obtained *or* that predictably different results will be obtained. Given, say, three of each which fall out in the predicted manner, this provides pretty compelling evidence for the theory. This is an oversimplification because case studies and their outcomes are likely to be multi-faceted and difficult to capture adequately within a simple theory. Support for the theory may be qualified or partial in any particular case, leading to revision and further development of the theory, and then probably the need for further case studies.

Preparing a case study plan

An important feature of case study is that *if more than one investigator is involved, they typically take on essentially similar roles*. The tasks cannot be reduced to rigid formulae with division of function as in survey research. All the investigators need an intelligent appreciation of what they are doing, and why. Hence it is highly desirable that all are involved in the first stages of conceptualization and definition of the research questions. Similarly, they should all be involved in the development of the case study plan.

The plan contains details of the data collection procedures to be used and the general rules to be followed. Where there is a single investigator, the main purpose of the plan is to enhance the validity of the study but it also acts as an aide-mémoire to the investigator. When a team is involved, it also serves to increase reliability in the sense of assisting all investigators to follow the same set of procedures and rules. Box 6.5 gives suggestions for the organization of the plan.

BOX 6.5

The case study plan

It is highly desirable that an explicit plan is prepared and agreed by those involved *in the full knowledge and expectation that aspects of this may change as the work continues*.

The following sections may be helpful:

1. *Overview*. Covers the background information about the project; the context and perspective, and why it is taking place; the issues being investigated and relevant readings about the issues.
2. *Procedures*. Covers the major tasks in collecting data, including:
 - (a) access arrangements;
 - (b) resources available; and
 - (c) schedule of the data collection activities and specification of the periods of time involved.
3. *Questions*. The set of research questions with accompanying list of probable sources of evidence.
4. *Reporting*. Covers the following:
 - (a) outline of the case study report(s);*
 - (b) treatment of the full 'data base' (i.e. totality of the documentary evidence obtained); and
 - (c) audience(s).*

Note: The plan should communicate to a general intelligent reader what is proposed. It forms part of establishing the validity of the study.

*There may be several audiences, for which different reports (in style and length) are needed. Consideration of reports and audiences at this stage, and during the study, helps to guide the study. See Chapter 18.

Pilot studies


A pilot study is a small-scale version of the real thing; a try-out of what you propose so that its feasibility can be checked. There are aspects of case study research which can make piloting both more difficult to set up and, fortunately, less crucially important. It may be that there is only one case to be considered or that there are particular features of the case selected (e.g. geographical or temporal accessibility, or your own knowledge of the case), such that there is no sensible equivalent which could act as the pilot.

In circumstances like these, the flexibility of case study gives you at least some opportunity to, as it were, 'learn on the job'. Or it may be that the initial formulation leans more toward the 'exploratory' pole of case study design and later stages with the benefit of experience can have a more 'explanatory' or 'confirmatory' focus.

Yin distinguishes between 'pilot tests' and 'pre-tests'. He views the former as helping investigators to refine their data collection plans with respect to both the content of the data and the procedures to be followed. For him, the pilot is, as it were, a laboratory for the investigators, allowing them to observe different phenomena from many different angles or to try different approaches on a trial basis. I prefer to regard these as case studies in their own right with an essentially exploratory function where some of the research questions are methodological. What he calls the 'pre-test' is a formal 'dress rehearsal' in which the intended data collection plan is used as faithfully as possible and is perhaps closer to the usual meaning of a pilot study.

Every research project is a kind of case study

In one sense, all projects are case studies. They take place at particular times in particular places with particular people. Stressing this signals that the design flexibility inherent in the case study is there in all studies until we, as it were, design it out. If we have the potential for random sampling and tight control, then the delights of the experiment beckon. But even then augmentation with additional observation or unstructured interaction with participants might be very illuminating.

 The website provides references to case studies of different types as used in a range of disciplines.

Many flexible design studies, even though not explicitly labelled as such, can be usefully viewed as case studies. They take place in a specific setting, or small range of settings, context is viewed as important, and there is commonly an interest in the setting in its own right. While they may not be multi-method as originally designed, the use of more than one method of data collection when feasible can improve many flexible design studies. Hence, even if you decide to follow one of the other flexible design traditions, you will find consideration of the above sections on case study of value.

Ethnographic studies

An ethnography provides a description and interpretation of the culture and social structure of a social group. It has its roots in anthropology, involving an immersion in the particular culture of the society being studied so that life in that community could be described in detail. The ethnographer's task was to become an accepted member of the group including participating in its cultural life and practices. Anthropologists initially focused on exotic cultures such as Trobriand Islanders in New Guinea. Sociologists, initially at Chicago University, adapted the approach to look at groups and communities in modern urban society (Bogdan and Biklen, 2007), and it is currently widely used in social research (Atkinson *et al.*, 2007; O'Reilly, 2009).

A central feature of this tradition is that people are studied for a long time period in their own natural environment. Critics of the approach are concerned about researchers

getting over-involved with the people being studied, perhaps disturbing and changing the natural setting, and hence compromising the quality of the research. However, the argument is that 'in order to truly grasp the lived experience of people from their point of view, one *has* to enter into relationships with them, and hence disturb the natural setting. There is no point in trying to control what is an unavoidable consequence of becoming involved in people's lives in this way' (Davidson and Layder, 1994, p. 165; emphasis in original). Hence it becomes necessary to try to assess the effects of one's presence.

The main purpose and central virtue of this approach is often considered to be its production of descriptive data free from imposed external concepts and ideas. Its goal is to produce 'thick description' (Geertz, 1973) which allows others to understand the culture from inside in the terms that the participants themselves used to describe what is going on. There is clear value in doing this for and about cultures where little is known or where there have been misleading presumptions or prejudices about the culture of a group.

Some ethnographers display a general distrust of theorizing. However, there seems to be no reason why an ethnographical approach cannot be linked to the development of theory (Hammersley, 1985). Working within the ethnographic tradition is not an easy option for the beginner, for the reasons given in Box 6.6.

Using an ethnographic approach

Using an ethnographic approach is very much a question of general style rather than of following specific prescriptions about procedure. In process terms, it involves getting out into 'the field' and staying there. Classically this meant staying there for a long

BOX 6.6

Difficulties in doing an ethnographic study

1. To 'do an ethnography' calls for a detailed description, analysis and interpretation of the culture-sharing group. This requires an understanding of the specialist concepts used when talking about socio-cultural systems.
2. For traditional ethnographies the time taken to collect data is very extensive, often extending over years. Some current approaches (sometimes referred to as 'mini-ethnographies') seek to cut this down drastically, but this creates a tension with the requirement to develop an intimate understanding of the group.
3. Ethnographies have typically been written in a narrative, literary style which may be unfamiliar to those with a social science background (conversely this can be an advantage to those with an arts or humanities background). This may also be a disadvantage when reporting to some real world audiences.
4. Researchers have been known to 'go native', resulting in their either discontinuing the study or moving from the role of researcher to that of advocate.

BOX 6.7

Features of the ethnographic approach

1. The shared cultural meanings of the behaviour, actions, events and contexts of a group of people are central to understanding the group. Your task is to uncover those meanings.
2. To do this requires you to gain an insider's perspective.
3. Hence you need both to observe and study the group in its natural setting, and to take part in what goes on there.
4. While participant observation in the field is usually considered essential, no additional method of data collection is ruled out in principle.
5. The central focus of your study and detailed research questions will emerge and evolve as you continue your involvement. A prior theoretical orientation and initial research questions or hypotheses are not ruled out, but you should be prepared for these to change.
6. Data collection is likely to be prolonged over time and to have a series of phases. It is common to focus on behaviours, events etc. which occur frequently so that you have the opportunity to develop understanding of their significance.

period, of the order of two or more years. This is highly unrealistic for virtually all real world studies and hence this section focuses on the use of ethnographic techniques rather than on how to carry out a full-scale ethnography. Box 6.7 lists features of the ethnographic approach.

Participant observation is very closely associated with the process of an ethnographic study. Chapter 13 considers the different types of role the observer might take. Whatever degree of participation you adopt (ranging from that of full group member to one where you are involved solely as a researcher) observation is difficult, demanding and time consuming. Box 6.8 helps you to assess whether or not it is for you. If it isn't, you should probably rule out an ethnographic study. However, while it is undoubtedly true that virtually all ethnographic studies do use this type of observation, they can be eclectic in methods terms, making use of whatever technique appears to be feasible. The feature which is crucial is that the researcher is fully *immersed* in the day-to-day lives of the people being studied.

The focus of an ethnographic study is a group who share a culture. Your task is to learn about that culture, effectively to understand their world as they do. Initially such studies were carried out by cultural anthropologists who studied societies and cultures very different from their own. Even when some familiar group or sub-group within one's own society is the focus, the ethnographic approach asks the researcher to treat it as 'anthropologically strange'. This is a very valuable exercise, particularly for those carrying out 'insider research'. It provides a means of bringing out into the open presuppositions about what you are seeing.

BOX 6.8

Using participant observation

Commit yourself to doing this only if the following fits you pretty closely:

1. You see interactions, actions and behaviours and the way people interpret these, act on them etc. as central.
2. You believe that knowledge of the social world can be best gained by observing 'real life' settings.
3. You consider that generating data on social interaction in specific contexts, as it occurs, is superior to retrospective accounts or their ability to verbalize and reconstruct a version of what happened.
4. You view social explanations as best constructed through depth, complexity and roundedness in data.
5. You are happy with an active, reflexive and flexible research role.
6. You feel it is more ethical to enter into and become involved in the social world of those you research, rather than 'standing outside'.
7. You can't see any alternative way of collecting the data you require to answer your research questions.

(adapted and abridged from Mason, 1996, pp. 84–102)

Ethnography and realism

Classically, ethnography was seen as a means of getting close to the reality of social phenomena in a way which is not feasible with the experimental and survey strategies. The Chicago sociologist Herbert Blumer talked about using ethnography to 'lift the veils' and to 'dig deeper', illustrating his realist assumptions (Hammersley, 1989).

However, there is a tension within the ethnographic research community on this issue. 'Central to the way in which ethnographers think about human social action is the idea that people *construct* the social world, both through their interpretations of it and through the actions based upon those interpretations' (Hammersley, 1992, p. 44, emphasis in original). Hammersley goes on to argue, persuasively, that this constructivist approach can be compatible with realism. This calls for an abandonment of the 'naive' realism characteristic of early ethnography where it was assumed that the phenomena studied were independent of the researcher who could make direct contact with them and provide knowledge of unquestionable validity. He argues in favour of 'subtle' realism as a viable alternative to the relativist constructionist approach.

The key elements of subtle realism, elaborated in Hammersley (1992, pp. 50–4) are:

- defining knowledge as beliefs about whose validity we are reasonably confident (accepting that we can never be absolutely certain about the validity of any claim to knowledge);

- acknowledging that there are phenomena independent of our claims about them, which those claims may represent more or less accurately; and
- an overall research aim of representing reality while acknowledging that such a representation will always be from a particular perspective which makes some features of the phenomenon relevant and others irrelevant (hence there can be multiple valid and non-contradictory representations).

This represents a reprise, using rather different terminology, of some of the issues discussed in Chapter 2 where the case was made for the adoption of a realist approach.

Designing an ethnographic study

The framework for research design given in Chapter 4 (Figure 4.1, p. 71) is applicable for an ethnographic study. As with other types of study, you need to give serious consideration to the purposes of your work and to establishing some (probably very tentative) theoretical or conceptual framework. This gives you an initial take on possible research questions, which in themselves assist in the selection of data collection methods and sampling – in the sense of who you observe, where, when, etc. While you can assume that participant observation of some kind will be involved, it may be that you have research questions which call for an additional approach or that additional methods will give valuable scope for triangulation.

An ethnographic approach is particularly indicated when you are seeking insight into an area or field which is new or different (and, paradoxically, in an area with which you are very familiar). It can help gain valuable insights which can then guide later research using other approaches. Remember that your initial research questions (and the other aspects of the framework you started out with) are highly likely to change and develop as you get involved.

There is no one specific design for an ethnographic study. Depth rather than breadth of coverage is the norm, with a relatively small number of cases being studied. Description and interpretation is likely to be stressed (Atkinson *et al.*, 2007; O'Reilly, 2009).

✦ *References to studies from different areas and disciplines using an ethnographic approach are provided on the website.*

While ethnography is a distinctive approach, it can be linked with either the case study or grounded theory approaches. A case study can be approached ethnographically or an ethnographic study can be approached by means of grounded theory.

Grounded theory studies

A grounded theory study seeks to generate a theory which relates to the particular situation forming the focus of the study. This theory is 'grounded' in data obtained

during the study, particularly in the actions, interactions and processes of the people involved. It is closely associated with two American sociologists, Barney Glaser and Anselm Strauss. Their highly influential text introducing this approach (Glaser and Strauss, 1967) has been followed by several more accessible introductions including Glaser and Strauss (1999) and Corbin and Strauss (2008).

Their approach was in reaction to the sociological stance prevalent in the 1960s which insisted that studies should have a firm *a priori* theoretical orientation. It has proved particularly attractive in novel and applied fields where pre-existing theories are often hard to come by. The notion that it is feasible to discover concepts and come up with hypotheses from the field, which can then be used to generate theory, appeals to many.

Grounded theory is both a strategy for doing research and a particular style of analysing the data arising from that research. Each of these aspects has a particular set of procedures and techniques. It is not a theory in itself, except perhaps in the sense of claiming that the preferred approach to theory development is via the data you collect. While grounded theory is often presented as appropriate for studies which are exclusively qualitative, there is no reason why some quantitative data collection should not be included. Indeed, the first studies reported in Glaser and Strauss (1967) made extensive use of quantitative data. In later years, differences have built up between the two collaborators to the extent that Glaser (1992) takes vigorous exception to the direction in which Strauss (and other colleagues) has taken grounded theory. Rennie (1998), in developing a rationale for grounded theory which reconciles realism and relativism, argues that Strauss and Corbin's approach effectively reverts back to the hypothetico-deductivism of traditional experimentalism, and that Glaser's procedures are more consistent with the objectives of the method.

Box 6.9 indicates some attractive features of grounded theory research; Box 6.10 some problems in carrying it out.

BOX 6.9

Attractive features of grounded theory research

1. Provides explicit procedures for generating theory in research.
2. Presents a strategy for doing research which, while flexible, is systematic and coordinated.
3. Provides explicit procedures for the analysis of qualitative data.
4. Particularly useful in applied areas of research, and novel ones, where the theoretical approach to be selected is not clear or is non-existent.
5. A wide range of exemplars of its use in many applied and professional settings is now available.

BOX 6.10

Problems in using grounded theory

1. It is not possible to start a research study without some pre-existing theoretical ideas and assumptions (as assumed in some versions of grounded theory research).
2. There are tensions between the evolving and inductive style of a flexible study and the systematic approach of grounded theory.
3. It may be difficult in practice to decide when categories are 'saturated' or when the theory is sufficiently developed.
4. Grounded theory has particular types of prescribed categories as components of the theory which may not appear appropriate for a particular study.

Carrying out a grounded theory study

A grounded theory study involves going out into 'the field' and collecting data. No particular type of 'field' is called for. Such studies have been carried out in a very wide range of settings. Glaser and Strauss initially worked in organizational contexts: interest in their studies of dying in hospitals (Glaser and Strauss, 1965, 1968) providing the stimulus for their first methodology text (Glaser and Strauss, 1967).

Interviews are the most common data collection method. However, other methods such as observation (participant or otherwise) and the analysis of documents can and have been used. Similarly, although grounded theory is typically portrayed as a qualitative approach to research, there is no reason in principle why some form of quantitative data collection cannot be used.

Procedurally, the researcher is expected to make several visits to the field to collect data. The data are then analysed between visits. Visits continue until the *categories* found through analysis are 'saturated'. Or, in other words, you keep on gathering information until you reach diminishing returns and you are not adding to what you already have. (A category is a unit of information made up of events, happenings and instances.)

This movement back and forth: first to the field to gather information, then back to base to analyse the data; then back to the field to gather more information, then back home to analyse the data etc. is similar to the 'dialogic' process central to the hermeneutic tradition (see below, p. 151). It is very different from a traditional linear one-way model of research where you first gather all your data, then get down to the analysis. It is close to the common-sense approach which one might use when trying to understand something which is complex and puzzling.

Sampling in grounded theory studies is *purposive* (see Chapter 10, p. 275). We do not seek a representative sample for its own sake; there is certainly no notion of random sampling from a known population to achieve statistical generalizability. Sampling of people to interview or events to observe is so that additional information can be obtained to help in generating conceptual categories. Within grounded theory, this type of

purposive sampling is referred to as *theoretical sampling*. That is, the persons interviewed, or otherwise studied, are chosen to help the researcher formulate theory.

The repeated comparison of information from data collection and emerging theory is sometimes referred to as the *constant comparative* method of data analysis. Its most standardized form is given in Strauss and Corbin (1998) and Corbin and Strauss (2008) and is summarized in Box 6.11 (note, however, that Glaser, 1992, dissents quite violently from some of their prescriptions).

Further details of the analysis process are given in Chapter 17. A summary is presented here because of the intimate interrelationship between design and analysis in a grounded theory study. It may also help you to appreciate that a grounded theory

BOX 6.11

Data analysis in grounded theory studies

The analysis involves three sets of coding:

1. *Open coding*. The researcher forms initial *categories* of information about the phenomenon being studied from the initial data gathered.

Within each category, you look for several subcategories (referred to as *properties*) and then for data to dimensionalize (i.e. to show the dimensions on which properties vary and to seek the extreme possibilities on these continua).
2. *Axial coding*. This involves assembling the data in new ways after open coding. A *coding paradigm* (otherwise known as a *logic diagram*) is then developed which:
 - identifies a *central phenomenon* (i.e. a central category about the phenomenon);
 - explores *causal conditions* (i.e. categories of conditions that influence the phenomenon);
 - specifies *strategies* (i.e. the actions or interactions that result from the central phenomenon);
 - identifies the *context* and *intervening conditions* (i.e. the conditions that influence the strategies); and
 - delineates the *consequences* (i.e., the outcomes of the strategies) for this phenomenon.
3. *Selective coding*. Involves the integration of the categories in the axial coding model. In this phase, conditional *propositions* (or hypotheses) are typically presented.

The result of this process of data collection and analysis is a *substantive-level theory* relevant to a specific problem, issue or group.

Note: The three types of coding are not necessarily sequential; they are likely to overlap. While the terms 'axial coding' and 'selective coding' are well established they are somewhat confusing as 'coding' usually refers to applying categories to data.

study is by no means an easy option and should not be undertaken lightly. It is, of course, possible to design a study which incorporates some aspects of grounded theory while ignoring others. For example, you may feel that the approach to coding is too prescriptive or restrictive. However, as with other research traditions, by working within the tradition you buy shelter and support from criticism – providing that the ways of the tribe are followed faithfully. And, less cynically, the fact that a group of researchers and methodologists have worked away at the approach over a number of years makes it likely that solutions to problems and difficulties have been found.

Realism and grounded theory

Disagreements within the grounded theory family, mentioned above, are mirrored in the stance taken about realism. Bryant and Charmaz (2007) claim that ‘The key weaknesses of Glaser and Strauss’s statement of the GTM [grounded theory method] resided in the positivist, objectivist direction they gave grounded theory’ and, in particular, that, ‘In seeking to provide a firm and valid basis for qualitative research their early position can be interpreted as justification for a naïve, realist form of positivism’ (p. 33). This reading is not unreasonable, but it was appreciated at an early stage that such positivist shackles were not an intrinsic feature of the approach.

Certainly, grounded theory as practised, has developed in more flexible ways, to the extent that, ‘The postmodernist may see this style as objectivist, realist and scientific; the positivist may see it as disconcertingly literary’ (Charmaz, 2003, p. 280). While this comment refers to reports of grounded theory studies, it reflects a more general perception that there are both objectivist and constructivist aspects to the ways of grounded theorists.

Of the two founding fathers, the position taken by Strauss (e.g. Corbin and Strauss, 2008) is the more constructivist, although considered by Anells (2006) to still reflect some aspects of positivist thought and language. Rennie and Fergus (2006) view the grounded theory method as an ‘accommodation’ of realism and relativism. For example, they consider that users of grounded theory:

are given the impression that social phenomena are external to the researcher and awaiting discovery, while being told that these phenomena are to be formulated creatively. They are encouraged to believe that with the correct procedures they will be able to access social phenomena grounded in reality, while being advised that the returns from the grounding will vary depending on the interests of the particular analyst (p. 484).

It is clear that there is no basic incompatibility between taking a realist view and using grounded theory. Grounded theory offers guidelines for building conceptual frameworks specifying the relationships among categories. If the guidelines are used as flexible tools rather than rigid rules, grounded theory gives researchers a broad method with distinct procedures that work in practice (Hallberg, 2006). As such it is suitable for pragmatic researchers of different methodological persuasions (or none).

◆ A range of references to studies from different areas and disciplines using a grounded theory approach is provided on the website.

Other traditions

This section lists several other possible approaches which have been used. The main principle for their inclusion has been that they may be useful for answering particular kinds of research question.

- *Narrative research.* Based on ‘stories.’ Can refer to an entire life story, long sections of talk, leading to extended accounts of lives, or even an answer to a single question.
- *Biographical and life history research.* A particular kind of case study where the ‘case’ studied is an individual person and the intention is to tell the story of a person’s life.
- *Phenomenological research.* Focuses on the need to understand how humans view themselves and the world around them. The researcher is considered inseparable from assumptions and preconceptions about the phenomenon of study. Instead of bracketing and setting aside such biases, an attempt is made to explain them and to integrate them into the research findings. The research methodology informed by what is often called interpretive phenomenology seeks to reveal and convey deep insight and understanding of the concealed meanings of everyday life experiences.
- *Hermeneutics.* Originally concerned with the translation and interpretation of sacred texts such as the Bible. It continues to provide a useful method for the analysis of texts and other documents. It has also been extended to include seeking understanding of any human action, with an emphasis on the importance of language in achieving that understanding, and on the context in which it occurs. The focus is on how the understanding is achieved rather than what is understood.

◆ The website gives additional information about these approaches.

Feminist perspectives and flexible designs

Some researchers following a feminist perspective reject quantitative methods and designs, and both positivism and post-positivism, as ‘representations of patriarchal thinking that result in a separation between the scientist and the persons under study (Fine, 1992)’ (Mertens, 2005, p. 232). For them flexible, qualitative designs are the only option. There is certainly an emphasis in these design traditions on non-exploitative research seeking an empathetic understanding between researcher and participants which chimes closely with feminist views. However, as Davidson and Layder (1994, p. 217) point out, feminist methodologists are not the only nor the first to advocate such approaches. Reinharz (1992) also cautions against the assumption that research using qualitative methods is inherently feminist. She also describes the wide variety of viewpoints that feminists hold on the appropriateness of different methods and methodologies.

Hence the position taken here, following Davidson and Layder (1994), is that while the critique presented by feminist methodologists of traditional social science research has yielded important insights, and helped to strengthen the case for qualitative research, the claim for a distinctive feminist research methodology has not been substantiated.

Sampling in flexible designs

Determining an appropriate sample in fixed designs, as discussed in the previous chapter, is relatively straightforward. There is almost always a concern that the sample is representative of a known population, so that statistical generalization is possible. Given this, what we find out about the sample can be regarded as telling us something about that population, probabilistically at least. The size of the sample is largely determined by the requirements of the statistical analysis we intend to carry out. Such analyses typically assume representative sampling (although, as discussed in Chapter 16, p. 445, all too often the samples have not been selected in a way which ensures this).

The most important thing to appreciate about sampling in flexible designs is that one is playing a very different kind of game. The data are almost always non-numerical and hence conventional statistical analysis is not feasible. Aspects of the qualitative data may be amenable to conversion into a numerical form, or there may be some data collected directly as numbers, enabling summary or descriptive statistics to be calculated, but the sample sizes are likely to be below those needed for statistical testing. More fundamentally, the nature of the sampling makes such testing inappropriate. It is typically *purposive* or *theoretical*, rather than seeking to be *representative* of a known population (see Chapter 10, p. 275). In these circumstances statistical generalization is not possible.

Nevertheless, although there are situations where researchers using flexible designs simply want to say something sensible about the specific circumstances of their research, many wish to make generalizations of some sort. This may be in the form of a theoretical conceptualization of what they have found (the central aim of a grounded theory study). Or, in realist terms, that they have evidence for mechanisms operating in certain contexts. Or, in very general terms, that the findings from the study somehow 'speak' to what might be happening in other settings or cases.

Generalizability of findings from flexible design research

Resist the temptation to smuggle in the concepts and approach of statistical generalization. It won't work.

Small (2009) emphasizes the basic incompatibility of statistical generalization and flexible design research. Box 6.12 summarizes one of the scenarios that he presents to illustrate his case (see also a second hypothetical scenario based on an ethnographic study of an 'average' neighbourhood).

BOX 6.12

Incompatibility of statistical generalization and flexible design research

Scenario: Study of attitudes towards immigration of working-class African Americans

Approach: Lengthy open-ended interviews with 35 respondents.

Question: How to ensure that the findings are generalizable?

Plan: Find city with large working-class African American population.

Random selection of 100 people from telephone directory. 40 target population agree to an interview; 35 follow through (highly optimistic figures). Conduct high-quality two-hour interviews.

From the perspective of statistical generalizability:

- *Problems of inbuilt and unaccounted for bias.* The 35 are those polite enough to talk, friendly enough to keep appointment based on cold call and extroverted enough to share feelings with interviewer. We know nothing about the 65 who weren't interviewed or about working-class blacks in other cities.
- *Sample too small.* Not large enough to make confident predictions about complex relationships in the population.
 - Alternatives
 - *Use a survey instead.* This would need to severely restrict the number of questions. With one simple yes/no question, to be confident statistically about the 1000 working-class blacks in one city you need an approximate sample of 300.
 - *Use a different sampling technique.* Snowball sampling where interviewees recommend other interviewees will increase the number of respondents and possibly their openness. But this would be seen as increasing bias and reducing representativeness from the perspective of statistical generalization.

(abridged from Small, 2009, pp. 11–15)

An alternative, favoured by Small, is to view the scenario of Box 6.12 as akin to multiple case studies rather than a small sample study. As discussed earlier in the chapter (p. 140), multiple case studies are effectively replications where each successive case adds to the understanding of questions at issue:

The first unit or case yields a set of findings and a set of questions that inform the next case. If the study is conducted properly, the very last case examined will provide very little new or surprising information. The objective is saturation. An important component of case study design is that each subsequent case attempts to replicate the prior ones. Through 'literal replication' a similar case is found to determine whether the same mechanisms are in play; through 'theoretical replication' a case different according to the theory is found to determine whether the expected difference is found (p. 25).

Note, however, that while the scenario has a weak sample on a sampling logic basis, it is also weak on a case logic basis. It is possible that one might achieve saturation after the 35 interviews – or even with a much smaller number. Following the case logic, one starts with a single interview, not knowing how many are going to be sufficient. The choice of a second interviewee will be based on issues arising from the first, perhaps using snowball sampling or some other means to identify a person from a particular background, age or gender. The process is then repeated several times until saturation. Viewed in this light the ‘How big a sample’ question disappears. There are obvious variants of this process where, perhaps, a small set of interviews are set up and decisions made about what lines to follow up after they are completed.

Such a process would be viewed as horrendously biased if statistical generalization is the aim. Many of the problems of bias seen by critics of qualitative research designs and procedures arise from this inappropriate aim. The techniques of verification and saturation inherent in well-conducted qualitative research (Morse, 2006) provide safeguards. They include:

- *Constantly evaluating the quality of data.* In a qualitative study, the quality of data is paramount. Investigators must attend to sampling adequacy (enough data), and sampling appropriateness (by interviewing ‘good informants’ who have experienced the phenomenon and who know the necessary information). If the proposed methods of data collection are not working and resulting in useful data, the investigator must change strategies – perhaps looking to a new study site, or types of participants, or even exploring whether the question itself is appropriate.

Sampling for scope and variation. This is necessary to ensure that comprehensive data are obtained.

Investigator sensitivity. Includes researcher reflexivity and techniques that enhance interpretation, including comparison with the literature, reflection on and comparison with known concepts, and the saturation of negative data.

- *Recognizing the progressive nature of enquiry.* Qualitative enquiry is a puzzle-solving activity. Ideas are initiated from one example, one instance or one participant, and the investigator is learning (gaining ideas) as the study progresses. Data from some participants may contribute more to the theoretical development of the study than others, some exemplars are better examples than others, and so on. This selection process is also reflected in the presentation of results, with the researcher’s commentary describing the range and richness of the categories or themes, and the best (clearest case) exemplars used.

Establishing trustworthiness in flexible design research

The trustworthiness or otherwise of findings from flexible research is the subject of much debate. Fixed design researchers criticize the absence of their ‘standard’ means of assuring

reliability and validity, such as checking inter-observer agreement, the use of quantitative measurement, explicit controls for threats to validity, the testing of formal hypotheses and direct replication. Thus, for example, while the essential test of validity of a finding in the natural sciences is that it has been directly replicated by an independent investigator, this approach is not feasible when a flexible design is used (and is also highly questionable in real world fixed design research involving people). One problem is that identical circumstances cannot be re-created for the attempt to replicate. As Bloor (1997) puts it, ‘Social life contains elements which are generalizable across settings (thus providing for the possibility of the social sciences) and other elements that are particular to given settings (thus forever limiting the predictive power of the social sciences)’ (p. 37).

Some researchers using flexible designs deny the relevance of canons of scientific enquiry (e.g. Guba and Lincoln, 1989). Others go further and reject the notion of any evaluative criteria such as reliability and validity (Wolcott, 1994). Taking an extreme relativist stance, it is maintained that using such criteria privileges some approaches inappropriately.

Altheide and Johnson (1994) argue that fields in the humanities such as history and literature employ evaluative criteria such as elegance, coherence and consistency which provide more appropriate standards for qualitative studies. While they may appear imprecise to traditional positivistically inclined researchers it is worth noting that even such a vague notion of elegance is used as a central criterion for the choice of one explanation over a rival in fields such as theoretical physics, the very heartland of natural science. More generally, accepting that social science researchers of whatever persuasion can benefit from some understanding of methodology in the humanities need not be at variance with the aspiration of remaining within the scientific fold put forward in Chapter 1. The problems of the relativist position outlined in Chapter 2 suggest a need for evaluative criteria in flexible designs. However, given the inappropriateness of the methods and techniques used in fixed design research, it is clear that different procedures for ensuring trustworthiness are called for (Kirk and Miller, 1986).

The terms reliability and validity are avoided by many proponents of flexible design. Lincoln and Guba (1985, pp. 294–301) for example prefer the terms *credibility*, *transferability*, *dependability* and *confirmability*. 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 and Brinkmann, 2009, p. 168). 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 2.

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 appropriate to the conditions and circumstances of flexible design research.

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 aspects 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 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 (1996) 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.

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, simply because 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 11, p. 281.

Interpretation

The main threat to providing a valid interpretation is through 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 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.

Bias and rigour

Issues of bias and rigour 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' central to many styles of qualitative research emphasizes the potential for bias. Padgett (1998, Chapter 8) presents a range of commonly used strategies to deal with these threats, which are discussed below.

Prolonged involvement

Involvement over a period of years was a defining characteristic of ethnography in its traditional anthropological version. Most current studies following the ethnographic approach have much more condensed fieldwork, but a period of weeks, or even months is still usual, a much more prolonged period than is typical in fixed methods research. This relatively prolonged involvement is also typical of other styles of flexible methods research and helps to reduce both reactivity and respondent bias. Researchers who spend a long time in the setting tend to become accepted and any initial 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 involving the use of multiple sources to enhance the rigour of the research. Denzin (1988b) 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 between the different sources. Thus, interviews and documents may be contradictory or 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, for example that findings collected by different methods differ to a degree which makes their direct comparison problematic. Such problems are a particular issue in multi-strategy (mixed methods) designs (Chapter 7, p. 161).

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 literally 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

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 search for instances which will disconfirm your theory. This may be in data you already have or through collection of additional data. This is sometimes referred to as 'playing the devil's advocate' and you have a responsibility to do this thoroughly and honestly. Don't be too concerned that this procedure will lead to you 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 Part 1, p. 1), and details of your data analysis.

Maxwell (1996; pp. 92–6), and Miles and Huberman (1984; 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. And 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, and 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 12. 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. At a technical level, the general non-standardization of many methods of generating qualitative data precludes formal reliability testing. Nevertheless, there are common pitfalls to all types of data collection and transcription including equipment failure, environmental distractions and interruptions, and transcription errors. Easton, McComish and Greenberg (2000) suggest strategies to minimize the risk from these problems.

In a more general sense, however, researchers using flexible designs do need to seriously concern themselves 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) 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 school, 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), sometimes referred to as *analytic* or *theoretical generalization*. For example, in realist terms, the study may provide convincing evidence for a set of mechanisms and the contexts in which they operate generalizable from, say, the particular intensive care unit studied to many other such units.

Further reading

◆ The website gives annotated references to further reading for Chapter 6.

CHAPTER 7

Multi-strategy (mixed method) designs

This chapter:

- explains what multi-strategy designs are;
- rejects the incompatibility thesis which claims that this kind of research is not possible;
- discusses the mixed methods movement;
- presents a typology of multi-strategy designs;
- stresses the centrality of research questions to multi-strategy design;
- emphasizes its compatibility with both pragmatic and realist stances;
- suggests ways of dealing with discrepancies between the findings of quantitative and qualitative elements;
- provides examples of multi-strategy research; and
- concludes by warning that this increasingly advocated approach is no easy option.

Introduction

This chapter focuses on designs where there is a substantial element of qualitative data collection as well as a substantial element of quantitative data collection in the same research project. The term *mixed methods* is commonly used for these designs (sometimes multiple methods – but see below). I prefer to stress the fact that they involve not only combining methods in some way but also use more than one research strategy – and so prefer to refer to them as *multi-strategy designs* (a terminology also favoured by

Alan Bryman who has published extensively in this area, e.g. Bryman, 2004). However, in the interests of communication I have included both terms in the chapter title. Using both fixed and flexible design strategies in the same research project raises a number of issues, some theoretical, some very practical.

Using two or more methods of collecting qualitative data in a project is commonly done and non-controversial. Case studies have followed this approach for many years, typically combining two or more methods of collecting qualitative data. Multiple quantitative data collection methods are also common (e.g. where data from a structured observation schedule are linked to a questionnaire survey). See Chapter 14, p. 385, where I refer to these as *multiple methods* studies, for further discussion of the issues involved.

The last 20 years have seen a considerably increased interest in multi-strategy (mixed method) designs. Apart from a large handbook on the topic (Tashakkori and Teddlie, 2003), specialist texts including Creswell and Plano Clark (2007), Greene (2007) and Teddlie and Tashakkori (2009) concentrate on how to carry out this type of research. The *Journal of Mixed Methods Research* and the *International Journal of Multiple Research Approaches* both started in 2007 and are sources of published articles from a wide range of different fields. An annual UK conference on the topic has been held since 2005. Some advocates of multi-strategy (mixed method) designs are evangelical in their zeal. It is seen by them as an idea whose time has come, a 'third way' to do research, arising phoenix-like from the smoking ashes of the quantitative–qualitative wars.

The quantitative–qualitative incompatibility thesis

The 'incompatibility thesis' is that multiple strategy research is not possible because qualitative and quantitative research are associated with two distinct paradigms that are incompatible with each other. Sale, Lohfeld and Brazil (2002) assert that 'Because the two paradigms do not study the same phenomena, quantitative and qualitative methods cannot be combined' (p. 43). Guba (1987) puts it more colourfully, 'The one [paradigm] precludes the other just as surely as belief in a round world precludes belief in a flat one' (p. 31).

Howe (1988) provides a comprehensive and convincing rebuttal of this thesis. He supports the view that, far from being incompatible, combining quantitative and qualitative methods is a good thing, and that 'there are important senses in which quantitative and qualitative methods are inseparable' (p. 10). A principle in the incompatibilist argument is that abstract paradigms should determine research methods in a simple one-way fashion. This principle was queried in Chapter 2 (p. 41) where an alternative, pragmatic, view was put forward. This is that there is a more complex two-way relationship between research methods and paradigms, where paradigms are evaluated in terms of how well they square with the demands of research practice. Crudely, if as is increasingly the case, research practitioners are successfully carrying out multi-strategy research, then the incompatibility thesis is refuted.

This is not to deny that there are major differences, particularly in research design and analysis, when dealing with quantitative and qualitative methods. The two preceding chapters discussed in detail the two research traditions of fixed and flexible design

research which appear poles apart. However, as Howe (1988) points out, it is possible to overemphasize these differences, and fail to realize that there are many similarities.

The chief differences between quantitative and qualitative designs and analysis can be accounted for in terms of the questions of interest and their place within a complex web of background knowledge. Because quantitative research circumscribes the variables of interest, measures them in prescribed ways, and specifies the relationships among them that are to be investigated, quantitative data analysis has a mechanistic, non-judgemental component in the form of statistical inference. But, as Huberman (1987) notes, this component is small in the overall execution of a given research project, and it is far too easy to overestimate the degree to which quantitative studies, by virtue of employing precise measurement and statistics, are eminently 'objective' and 'scientific'. One gets to the point of employing statistical tests only by first making numerous judgements about what counts as a valid measure of the variables of interest, what variables threaten to confound comparisons, and what statistical tests are appropriate. Accordingly, the results of a given statistical analysis are only as credible as their background assumptions and arguments, and these are not amenable to mechanistic demonstration (p. 12).

In other words, fixed designs call for a complex web of qualitative judgements. Campbell (1978) goes so far as to argue that all research has a qualitative grounding. A further illustration of the blurring of the line between quantitative and qualitative research is a study by Gueulette, Newgent and Newman (1999), cited by Onwuegbuzie and Leech (2005), which analysed over 300 randomly selected studies labelled by their authors as representing qualitative research and found that over 40 per cent of the articles actually involved the blending of qualitative and quantitative methodologies.

It could still be argued that, despite the appearance that current research practice gives, quantitative and qualitative methods are in some sense incompatible. This view had greater force when methodologically minded researchers thought that they had to make a choice between a positivist and an interpretivist (or constructionist) approach. Positivist and interpretivist approaches are undoubtedly incompatible. Hence, if the positivist paradigm underpins quantitative methods, and an interpretivist paradigm underpins qualitative methods then, despite appearances, the two are incompatible. However, as discussed in Chapter 2, positivism has long ceased to be a viable option (though the message has still not got through to some researchers) and post-positivist approaches, including the more sophisticated variants of realism, as well as pragmatism, allow one to move beyond making the forced choice on which the incompatibility thesis relied. Howe (1988) concludes trenchantly:

Questions about methodology remain, but they ought not be framed in way that installs abstract epistemology as a tyrant or that presupposes the moribund positivist–interpretivist split. The fact that quantitative and qualitative methods indeed might be historical outgrowths of incompatible positivist and interpretivist epistemologies no more commits present-day researchers to endorsing one or the other of these epistemologies than the fact that astronomy is an outgrowth of astrology commits present-day astronomers to squaring their predictions with their horoscopes (p. 15).

The mixed methods movement

The development of a somewhat evangelical movement where so-called mixed methods research evolved into a new research paradigm is commonly cited as an aftermath of the quantitative–qualitative ‘paradigm wars’. An early period in which the positivist quantitative paradigm was dominant between the 1950s and mid-1970s was followed by one in which the qualitative interpretivist/constructivist research paradigm became established as a viable alternative in the mid-1970s to the 1990s. Mixed methods, as a research paradigm, is seen as emerging from the 1990s onwards, establishing itself alongside the previous paradigms so that ‘we currently are in a three methodological or research paradigm world, with quantitative, qualitative, and mixed methods research all thriving and coexisting’ (Johnson, Onwuegbuzie and Turner, 2007). The distinctive nature of the mixed methods approach and the core ideas and practices on which the paradigm stands have been spelled out by Creswell (2003) and Tashakkori and Teddlie (2003) among others. A contrast is made between a mixed methods approach and research paradigms advocating the use of either quantitative or qualitative methodologies. Its defining characteristics are typically cited as:

- quantitative and qualitative methods within the same research project;
- a research design that clearly specifies the sequencing and priority that is given to the quantitative and qualitative elements of data collection and analysis;
- an explicit account of the manner in which the quantitative and qualitative aspects of the research relate to each other; and
- pragmatism as the philosophical underpinning for the research (Denscombe, 2008).

However, not all mixed methods studies fit within this definition (see, for example, the very varied set of studies in Weisner, 2005).

Many researchers within the mixed methods movement do explicitly espouse a pragmatic approach (e.g. Morgan, 2007; Onwuegbuzie and Leech, 2005), but this is a controversial area – see below, p. 170.

While referring to approaches of this kind as ‘mixed methods’ research is well established I will, as discussed above, talk about multi-strategy designs as a better descriptor. Feel free to substitute ‘mixed methods’ if you prefer to vote with the current majority.

Types of multi-strategy designs

Research strategies and methods can be combined in a variety of ways. Box 7.1 presents a simple typology, based upon the order or sequence of the design elements and the priority that they are given. The two ‘transformative’ designs are typically used when there is a dedication to social change of some kind, reflecting an emancipatory or

BOX 7.1

A typology of multi-strategy designs focusing on the sequencing and status of data collection methods

1. *Sequential explanatory design*. Characterized by the collection and analysis of quantitative data followed by the collection and analysis of qualitative data. Priority is typically given to the quantitative data and the two methods are integrated during the interpretation phase of the study. The qualitative data function to help explain and interpret the findings of a primarily quantitative study.
2. *Sequential exploratory design*. Characterized by an initial phase of qualitative data collection and analysis followed by a phase of quantitative data collection and analysis. Priority is given to the qualitative aspect of the study. The findings are integrated during the interpretation phase. The primary focus of this design is to explore a phenomenon.
3. *Sequential transformative design*. One method precedes the other with either the qualitative or the quantitative method first. Priority may be given to either method. The results are integrated during interpretation. This design is guided primarily by a theoretical perspective (e.g. by the conceptual framework adopted).
4. *Concurrent triangulation design*. Qualitative and quantitative methods are used separately, independently and concurrently. Results are compared to assess their convergence.
5. *Concurrent nested design*. Involves the embedding or nesting of a secondary method within a study with one main or primary method. The primary method can be either quantitative or qualitative.
6. *Concurrent transformative design*. Guided primarily by the researcher’s use of a specific theoretical perspective, as in the sequential transformative design above.

(based on Creswell, 2003, pp. 213–19)

empowerment purpose. A very similar typology, with slightly different labels, has been put forward by Leech and Onwuegbuzie (2009).

Maxwell and Loomis (2003) accept the value of such typologies but stress their limitations. They consider that they do not capture the actual diversity of the designs researchers have used, and that typically they do little to clarify the actual functioning and interrelationship of the qualitative and quantitative parts of the design. Rather than viewing a research design as choosing from a set of possible arrangements or sequences they propose an interactive model with five components – purposes or goals, conceptual framework, research questions, methods, and validity (Maxwell, 2005, p. 5). Their model is very similar to that developed in Chapter 4 (p. 71), but with the ‘interactive’

relationship between the components stressed by linking them using two-way arrows. It follows the mantra repeated throughout this book; that the research questions are at the heart of the design. Maxwell and Loomis advocate an integration of the typological and their interactive design approach. Typologies help in deciding on the *type of study*; in making broad decisions about how to proceed; in the sequencing and ordering of different approaches; and their relative dominance. The design model is a tool for designing and analysing an *actual study* (see below, p. 168).

Benefits of multi-strategy designs

Many benefits have been claimed for combining quantitative and qualitative data collection methods in a project. Box 7.2 lists some of them. Several of these benefits can accrue in multiple-method projects where the methods used are all quantitative, or all qualitative (see Chapter 14, p. 385). However, it will be clear that there is a greater variety of potential benefits when approaches associated with the two different paradigms of quantitative and qualitative research are brought together.

Complexities of multi-strategy designs

Bryman (2004) summarizes the results of interviewing a sample of 'mixed method' researchers. They cited a range of concerns, including:

- *Skills and training.* Skills and training is seen as a problem area. The skills and inclinations of many researchers are either quantitative or qualitative and they feel uncomfortable with the other tradition. Most commonly, this takes the form of qualitative researchers expressing unease about involvement in the more advanced forms of quantitative data analysis.
- *Timing issues.* Quantitative and qualitative components sometimes have different time implications. Most frequently, this takes the form of quantitative research being completed more quickly than the qualitative component.
- *Limits of multi-strategy research.* Multi-strategy research is not obviously beneficial when the rationale for combining quantitative and qualitative research is not made explicit. In such cases, it is difficult to judge what has been gained by employing both approaches. In some studies, qualitative data are used only or mainly to illustrate quantitative findings. In such cases, the qualitative findings are largely ornamental and do not add a great deal to the study.
- *Lack of integration of findings.* Responses indicated that only a small proportion of studies fully integrate the quantitative and qualitative components when the research is written up.

Mason (2006), while acknowledging that they can have several benefits, is concerned that multi-strategy designs can produce disjointed and unfocused research, and can severely test the capabilities of researchers. Researchers 'need to have a clear sense of the logic and purpose of their approach and of what they are trying to achieve,

BOX 7.2

Potential benefits of multi-strategy designs

1. *Triangulation.* Corroboration between quantitative and qualitative data enhances the validity of findings.
2. *Completeness.* Combining research approaches produces a more complete and comprehensive picture of the topic of the research.
3. *Offsetting weaknesses and providing stronger inferences.* Using these designs can help to neutralize the limitations of each approach while building on their strengths, leading to stronger inferences.
4. *Answering different research questions.* Multi-strategy designs can address a wider range of research questions than is feasible with single method fixed or flexible designs.
5. *Ability to deal with complex phenomena and situations.* A combination of research approaches is particularly valuable in real world settings because of the complex nature of the phenomena and the range of perspectives that are required to understand them.
6. *Explaining findings.* One research approach can be used to explain the data generated from a study using a different approach (e.g. findings from a quantitative survey can be followed up and explained by conducting interviews with a sample of those surveyed to gain an understanding of the findings obtained). This can be particularly useful when unanticipated or unusual findings emerge.
7. *Illustration of data.* Qualitative data can illustrate quantitative findings and help paint a better picture of the phenomenon under investigation. Bryman (2006a) refers to this as putting 'meat on the bones' of dry quantitative data.
8. *Refining research questions (hypothesis development and testing).* A qualitative phase of a study may be undertaken to refine research questions, or develop hypotheses to be tested in a follow-up quantitative phase.
9. *Instrument development and testing.* A qualitative phase of a study may generate items for inclusion in an instrument (e.g. questionnaire, test or scale, or structured observation schedule) to be used in a quantitative phase of a study.
10. *Attracting funding for a project.* Agencies funding research projects are showing increased interest in interdisciplinary research involving collaboration between disciplines traditionally using different approaches (e.g. in health-related areas where collaboration on projects between nursing, medical and other professionals is increasingly promoted and encouraged).

(based, in part, on Bryman, 2006a)

because this ultimately must underpin their practical strategy not only for choosing and deploying a particular mix of methods, but crucially also for linking their data analytically' (p. 3). However, she admits that sometimes mixing methods and data can become possible more by accident than design, especially where existing data sets

become available unexpectedly or serendipitously, or where access is available to a potential data source.

Mason stresses that, in the real world, practical, political and resource issues will establish certain constraints and contexts for those wishing to carry out multi-strategy research projects. These include:

- power, status and inequalities within and between teams, and for individual researchers, and between disciplines and fields of interest;
- constraints and opportunities of research funding; responsibilities to and expectations of funders and other stakeholders;
- access to and ownership of data;
- opportunities for collaboration, for sole working, for authorship;
- spread of skills and competencies;
- time, resources and capacity to learn new skills; and
- possibilities for strategic planning of outputs, e.g. for different purposes and audiences (pp. 11–12).

She concludes that 'it is just as important to recognize how these factors play out in one's own real life research, as it is to be clear about a desired strategy for mixing methods – since these are inextricably related and mixed methods research practice will involve dealing with both in tandem'.

Designing and carrying out multi-strategy research

The basic approach to research design discussed in Chapter 4 still applies. There is the additional task of clarifying and making explicit your rationale for, and the purpose of, using this type of design as discussed above. Essentially, why are you mixing quantitative and qualitative methods?

Hence, in the design framework of Figure 4.1 (p. 71) we have the elements of:

- purpose(s);
- conceptual framework;
- research questions;
- methods; and
- sampling procedures.

In multi-strategy research, consideration of purpose(s) has to be extended from the general issues covered in Chapter 4 to 'Why a multi-strategy design?' In similar vein, the other elements (particularly the conceptual framework – see Greene, Caracelli and Graham, 1989) have to take note of the particular issues raised by using both quantitative and qualitative methods of data collection. We are effectively marrying fixed and flexible design elements in the same overall project. Issues already discussed earlier in this chapter include:

- Implications for the overall design of likely difference in time scales for the qualitative and quantitative elements. Is the completion of one phase essential before a consequent phase can be started?
- As a sole researcher, do you have the necessary skill set to carry out both qualitative and quantitative elements? And to analyse and interpret both data sets? Can you get help where and when needed?
- In a research team, is there agreement about who does what? Are all team members in agreement about the design approach, etc.?

Research questions rule

The centrality of research questions for the research process has been the mantra of this text. This view enjoys considerable support in the research community active in multi-strategy social research. It is, in part, accounted for by the pragmatist stance taken by many which regards the research question(s) as the driver for carrying out research. However, there is not a necessary linkage between the two and researchers can, and do, approach multi-strategy research from other philosophical or theoretical standpoints – sometimes from none which is discernible.

Chapters 5 and 6 discussed research questions in the context of fixed design and flexible design strategies respectively. As multi-strategy designs necessarily include both fixed and flexible design elements or phases, they therefore call for the inclusion of a research question, or questions, covering both aspects. Keeping the focus on it being a single research project is helped by having a single main research question which to be answered properly needs both quantitative and qualitative data collection. Separate sub-questions, focusing on the different elements individually, can then be developed.

Onwuegbuzie and Leech (2006) present a detailed discussion of the development of research questions in this field and provide several examples of main research questions focusing on the concurrent/sequential distinction.

Examples of research questions for which a sequential design is appropriate:

1. *What is the difference in perceived barriers to reading empirical research articles between graduate students with low levels of reading comprehension and those with high levels of reading comprehension?* Here, the quantitative research component would generate levels of reading comprehension and the qualitative research element would generate the perceived barriers to reading empirical research articles. The overall research design is sequential because the quantitative phase of the study would inform the qualitative phase. The researcher would administer a test of reading comprehension, rank these comprehension scores, and then select students who attained scores that were in the top third and bottom third, say, of the score distribution. These students could then be interviewed and asked about their perceptions of barriers that prevent them from reading empirical research articles.
2. *What is the difference in perceived atmosphere of classroom between male and female graduate students enrolled in a statistics course?* To address this question, you would use qualitative techniques (e.g., interviews, focus groups, observations) to examine the

experiences of students enrolled in a statistics course. On finding that the negative experiences of some of the study participants are extreme, relative to other members of the class, you might decide to compare statistically scores on the final statistics examination between these two sets of students. The overall research design would be sequential. A qualitative phase might involve a case study research design. The quantitative research phase would call for a descriptive, correlational, or causal-comparative research design.

3. *What are the characteristics of participants who do not fit the theory emerging from an initial phase of the design?* (a generic question type widely applicable when grounded theory is used). Qualitative techniques (e.g. interviews, focus groups, observations) could be used to collect and analyse qualitative data using a grounded theory approach until theoretical saturation is reached. Cases which do not fit the emergent theory could be identified. Such negative and other non-negative cases could be compared with respect to one or more sets of existing quantitative scores. Alternatively, new quantitative data could be collected and the two groups compared with regard to the new data. The overall research design would be sequential with the qualitative phase represented by a case study or grounded theory design, and the quantitative research phase by a correlational or causal-comparative design (examples based on Onwuegbuzie and Leech, 2006).

Examples of research questions for which a concurrent design is appropriate:

1. *What is the relationship between graduate students' levels of reading comprehension and their perceptions of barriers that prevent them from reading empirical research articles?* To answer this question information about both the levels of reading comprehension and the perceived barriers to reading empirical research articles must be obtained. Levels of reading comprehension would be gleaned from the quantitative component of the study, perceived barriers to reading empirical research articles from the qualitative part. The overall research design would be concurrent because the quantitative phase of the study did not inform or drive the qualitative phase or vice versa.
2. *What are the implications of the 'No Child Left Behind' Act on parents?* (a generic question type widely applicable to other legislation). A research question such as this could lead to a descriptive research design for the quantitative component of the study, (possibly with a variety of different data sets) and, say, a case study design for the qualitative element. Alternatively, the overall design could be thought of as a case study incorporating both quantitative and qualitative data collection. If both elements are essentially exploratory, a concurrent design will minimize the overall duration of the study (examples based on Onwuegbuzie and Leech, 2006).

Pragmatism, realism or 'anything goes'?

It has already been pointed out that many of the researchers who advocate 'mixed methods' designs couple this with an explicit endorsement of the virtue of a pragmatic stance. Interviews with them by Bryman (2006b, p. 124) found 'a tendency to stress the

compatibility between quantitative and qualitative research and a pragmatic viewpoint which prioritizes using any approach that allows research questions to be answered regardless of its supposed philosophical presuppositions'. Bryman (2006b) provides support, viewing a pragmatic approach as providing a way of redirecting attention to methodological rather than metaphysical concerns.

Pragmatism can be seen as providing a licence to carry out multi-strategy research, safe in the knowledge that a body of leading researchers in the field have followed this path. For Onwuegbuzie and Leech (2005) what they term 'pragmatic researchers' are simply those who learn to utilize and to appreciate both quantitative and qualitative research. From this they consider that several advantages flow, including:

- researchers can be flexible in their investigative techniques;
- a wide range of research questions can be addressed;
- they are more likely to promote collaboration among researchers (including those of different philosophical orientations);
- they are more likely to view research as a 'holistic endeavour'; and
- as they have a positive attitude to both qualitative and quantitative approaches, they are likely to favour using qualitative techniques to inform the quantitative aspect of a study and vice versa (p. 383).

This could be seen as a pretty minimal theoretical underpinning to multi-strategy research, verging on an 'anything goes' philosophy where, by the fact of carrying out this kind of project, you qualify as a pragmatist. There is a danger of being open to the criticism of carrying out incoherent projects lacking a rationale and of dubious validity. The situation can be rescued by taking seriously the design task discussed in the previous section. Given clarity of the purposes of the study, a thought-through conceptual structure, and in particular a feasible research question or questions, as well as attention to the other aspects of the design framework, a convincing methodological rationale can be established.

It is, of course, possible to take on board the philosophical tenets of pragmatism, as discussed in Chapter 2, p. 27. Scott and Briggs (2009) develop a sophisticated 'pragmatist argument for mixed methodology' in the field of medical informatics, basing the argument on this eminently real world field's confluence of pragmatist clinical practice, empirical social science and information technology.

Notwithstanding the dominant pragmatic tendency in much multi-strategy research, other theoretical rationales have been put forward. While there are several possibilities, some are effectively ruled out if one takes the 'third way' argument seriously. Viewing multi-strategy research designs which are neither exclusively quantitative, nor exclusively qualitative, but a genuine attempt to develop a hybrid third way, restricts the choice. Post-positivists find much going under the banner of qualitative research deeply uncongenial. Interpretivists and constructionists are probably even less sympathetic to the ways in which essentially quantitative strategies such as experiments and surveys are conducted. Among the attractions of realist approaches is their capacity to embrace both quantitative and qualitative ways of carrying out social research, seized on by Lipscomb (2008) and by McEvoy and Richards (2006), who have argued for (critical) realism as a natural partner for multi-strategy research.

The view taken in this text is that realism (including critical realism) has much to offer the real world researcher. Adoption of realist terminology and associated concepts (e.g. and in particular, generative mechanisms) encourages a productive way of thinking about many of the issues which arise in designing a study, and interpreting and understanding its findings. Multi-strategy research, rather than introducing new and specific realist concerns, provides a context where they appear particularly apposite. Maxwell and Mittapalli (2010, pp. 160–2), in a chapter advocating the use of ‘realism as a stance for mixed methods research’ review several examples of the explicit uses of realism in this field. They include realist approaches to evaluation research, a field where quantitative and qualitative approaches are often combined (Henry, Julnes and Mark, 1998; Pawson and Tilley, 1997) – see Chapter 8, p. 178; and several studies adopting a critical realist perspective (e.g. Clark, MacIntyre and Cruickshank, 2007; Lipscomb, 2008; Mingers, 2006; Olsen, 2004).

Dealing with discrepancies in findings

Findings from the qualitative and quantitative elements or phases of a project may, or may not, corroborate each other. If they do, fine. You have greater confidence in the findings and their validity. If they don’t, all is not lost but you do have to do further work to try to establish the reason(s) for the discrepancy. Greene (2007) in her discussion of dealing with divergent findings emphasizes their value for deepening understanding of the phenomena studied (see, especially, pp. 79–82).

Moffatt *et al.* (2006) discuss different ways of dealing with apparent discrepancies between qualitative and quantitative research data in a study evaluating whether welfare rights advice has an impact on health and social outcomes. These include:

- *Treating the methods as fundamentally different.* A process of simultaneous qualitative and quantitative data set interrogation enables a deeper level of analysis and interpretation than would be possible with one or other alone and demonstrates how multi-strategy research produces more than the sum of its parts. It is not wholly surprising that methods come up with divergent findings if they ask different, but related questions, and are based on fundamentally different theoretical paradigms. Combining the two methods for cross-validation (triangulation) purposes is only a viable option if both methods are examining the same research problem. Moffatt *et al.* approached the divergent findings as indicative of different aspects of the phenomena in question and searched for reasons which might explain these inconsistencies. They treated the data sets as complementary as each approach reflected a different view on how social reality ought to be studied.
- *Exploring the methodological rigour of each component.* It is standard practice at the data analysis and interpretation phases of any study to scrutinize methodological rigour. In this case, they had another data set to use as a yardstick for comparison and it

became clear that interrogation of each data set was informed to some extent by the findings of the other. Possible reasons why there might be problems with each data set were investigated individually but they found themselves continually referring to the results of the other study as a benchmark for comparison. With regard to the quantitative study, the sample size had insufficient power to detect small differences in the key outcome measures. Other factors provided some explanation for the lack of a measurable effect between intervention and control group and between those who did and did not receive additional financial resources. The number of participants in the qualitative study who received additional financial resources as a result of this intervention was small but they argue that the fieldwork, analysis and interpretation were sufficient to claim that the findings were therefore an accurate reflection of what was being studied. However, there still remained the possibility that a reason for the discrepant findings was due to differences between the various sub-samples.

- *Exploring data set comparability.* They compared the qualitative and quantitative samples on a number of social and economic factors. There were negligible differences in test scores between the groups at baseline, which led them to discount the possibility that the samples were markedly different on these outcome measures.
- *Collecting additional data and making further comparisons.* Quantitative and qualitative follow-up data verified the initial findings of each study.
- *Exploring whether the intervention under study worked as expected.* The qualitative study revealed that many participants had received welfare benefits via other services prior to this study, revealing the lack of a ‘clean slate’ with regard to the receipt of benefits, which was not anticipated.
- *Exploring whether the outcomes of the quantitative and qualitative components match.* The qualitative study revealed a number of dimensions not measured by the quantitative study, such as, ‘maintaining independence’ which included affording paid help, increasing and improving access to facilities and managing better within the home. Secondly, some of the measures used with the intention of capturing dimensions of mental health did not adequately encapsulate participants’ accounts of feeling ‘less stressed’ and ‘less depressed’ by financial worries. The data demonstrated the difficulties of trying to capture complex phenomena quantitatively. They also demonstrated the value of having alternative data forms on which to draw whether complementary (where they differ but together generate insights) or contradictory (where the findings conflict). The complementary and contradictory findings of the two data sets proved useful in making recommendations for the design of a definitive study.

The strategies adopted in this study have general relevance to the further exploration of discordant results. They highlight the dangers of relying on the findings from any study which used a single method of data collection (including relying on mono-method RCTs when seeking to evaluate complex interventions with a social component).

📖 *The website gives further examples of dealing with divergent findings in multi-strategy research.*

Examples of multi-strategy research

The Moffat *et al.* (2006) paper discussed in the previous section, although primarily methodological in focus, also provides a good example of a multi-strategy design.

✧ *The website gives further examples of multi-strategy research.*

Concluding comments

To carry out a multi-strategy research project, you are likely to have to call on material and suggestions from many of the chapters of this book. All of the other chapters in Parts I and II will probably have some relevant aspects, then the chapters in Part III covering the specific methods you select and all of the chapters in the remaining parts. In this sense, it is similar to evaluation research (it is, of course, perfectly feasible for an evaluation to use a multi-strategy design).

You also need to cover the material in this chapter, giving particular attention to design aspects. So, a multi-strategy design is not to be selected lightly, particularly by a lone and/or new researcher. Not only do you need to have the requisite skills to use both qualitative and quantitative data collection techniques successfully but you also need the time to actually carry out at least two very different types of data collection – and to analyse and interpret the resulting data. Obviously experience and the existence of a team of researchers reduce many of these concerns. A lack of integration of findings from qualitative and quantitative analyses in much research has already been referred to (p. 166) and is addressed at the end of Chapter 17, p. 492.

The mixed methods advocates make a strong case for this type of research design, and it appears likely to be of increasing importance in the next few years. However, a poorly designed and/or executed multi-strategy design is worse than a competent mono-method study.

Further reading

✧ *The website gives an annotated list of further reading.*

CHAPTER 8

Designs for particular purposes: evaluation, action and change

This chapter:

- stresses the ubiquity and importance of evaluation;
- discusses different forms of evaluation research;
- covers the planning and carrying out of evaluations;
- emphasizes the political dimension of evaluations;
- introduces needs assessment and cost-benefit analysis;
- explains the distinctive features of action research and other participatory approaches;
- considers the place of research in producing social change; and
- discusses some of the problems associated with doing this.

Introduction

Much real world research is concerned with evaluating something. Real world researchers also often have an ‘action’ agenda. Their hope and intention is that the research and its findings will be used in some way to make a difference to the lives and situations of those involved in the study and/or others. This takes us into the somewhat specialist fields of *evaluation research* and *action research*.

Researchers tend to bemoan the lack of influence that research has on practice. Some reasons for this ineffectiveness, and what might be done about it, are discussed later in the chapter.