The Design Frame

Once you have decided on your general approach to answering your question, you can think about the ‘scaffold’ or ‘design frame’ within which you will plan and execute your research. This chapter considers:

• some general issues in designing a piece of research;
• the main ‘frames’ to guide the way that you carry out research:
  – action research;
  – case study;
  – comparative research;
  – ethnography;
  – evaluation;
  – experiment;
  – longitudinal, cross-sectional studies and survey.

Design is about plan and structure, and as Figure 5.9 in Chapter 5 indicates, the whole programme of your research, from purposes to execution, constitutes the design. One particular part of the sequence of decisions you make will be about what I am here calling the design frame. I have called it this because it constitutes the most important element in the way that your research is structured: it’s like a chassis or superstructure that supports your research. The design frame provides the framework for your research – connecting purposes with questions with the ways in which data can be collected – though it does not prescribe how that data will be collected.

I will be concentrating on seven types of design frame here. There are others, but these seven are the most common structures used in small research projects. They are:

• action research;
• case study;
• comparative research;
ethnography;
evaluation;
experiment;
longitudinal, cross-sectional studies and survey.

It is important to reiterate that these are not designs for research in themselves. The design itself is the plan for research that you adopt from the beginning of your project. The reason I have called these design frames is that they provide the defining structure within the design. By choosing one or more of them you are consolidating decisions that you have made about the purposes of your inquiry, the kind of research that you are doing and what you want to achieve from it.

The design frames are scaffolds within which to structure what you do. But it is important also to say that these design frames are not in any way similar as structures, nor are they mutually exclusive. They can exist in combination. So, for example, action research could take shape in most of the other forms, or a case study could include a survey. This will become clearer as we look at the frames themselves.

The idea of research design per se – the idea that you plan research at the beginning and carry on according to the blueprint you have drawn until the end – is a bit of a hangover from the days when social research was expected to be as similar as possible to natural scientific research. As such, it would be mainly experimental and it would come complete with very specific instructions on procedures, methods and apparatus, with the idea that anyone could come along after you and repeat the experiment you were doing. The key idea behind all of this was replicability, or in other words the ability for someone else to repeat what you were doing. If then, after many repeats of your experiment, many others had the same finding as you, having followed the same design and procedure, the scientific community could be sure that the finding was secure.

Nowadays, social scientists working in an applied field such as education, social work or criminal justice are particularly aware of the difficulties of conforming to these kinds of expectations for our research. For a start there are now recognised to be difficulties, sometimes insuperable, of managing a social situation in the same way as a chemistry experiment, but there is also now less certainty that there is one best way of organising research. As I tried to show in Chapter 5, if we are trying to research something in the social world we don’t even necessarily even have to adopt the position of neutral, disinterested observer. By contrast, we can be open, involved interpreters of events, responding and changing as new information appears.

The latter point is particularly relevant as far as ‘research design’ is concerned, since it implies that there should be far less rigidity about such design than had hitherto been expected. This ties in with the expectation of a recursive plan (rather than a linear plan) that I talked about in Chapter 1. The design, it implies, should not be set in stone, ready to be replicated exactly by the next researcher. Given that this is
the case, some have spoken about emergent design, in other words letting the design ‘happen’ as you find out more about the situation in which you are interested. This idea of emergent design is an important one for social research in the interpretative tradition, and though the word ‘design’ is still used, it really turns the idea of ‘design’ on its head, since something that ‘emerges’ cannot be ‘designed’. We should perhaps look for a new word for the process.

However, I’m not going to do that now, since ‘design’ is the word we are stuck with and everyone knows what they mean by it – or at least they think they do. The trouble with using ‘design’ is that it implies all of the traditional features of experimental design that I have talked about (e.g. specification of sample, apparatus and so on), and these features carry with them other expectations. There are, for example, expectations about sample size (the bigger the better), reliability (you have to be sure of getting the same result if you do the same again) and validity (you have to be sure that you are finding out what you set out to find). But these are not the ground rules for interpretative research.

Yes, I said not. It is not expected that you can generalise from interpretative research: your ‘sample’ gives you insights rather than generalisations. So your ‘sample’ is small – even as small as one. Be happy with that. In fact the notion of the sample is a misnomer in interpretative research since your informant (or whatever) is not a sample from a wider population. They have integrity in their own right. It is not expected that if someone else does the study they will find something identical to what you found. Quite the contrary: someone else will almost certainly find something very different from what you found and this is to be expected. They will be interpreting with their personal history, interests, predilections and idiosyncrasies, you with yours.

So, the word ‘design’ should be interpreted with caution in social research. In certain kinds of research it will be more fixed; in others less. Expectations about it in one kind of research will not always apply in another.

Some general issues in design

As I have just tried to indicate, design is something of a misnomer and leads people to have cast-iron expectations about the structure of research, ignoring the tenets of different types of research – the presuppositions that ground it. Given these different types of research, there can be no expectation that the ground rules and methods of one kind of research will be appropriate in another. It is as well to be aware of this since a major error committed in many dissertations is the use of one set of expectations with a form of research that does not carry these expectations. It is with this warning that I discuss now some general issues in designing a research project.
Sampling

The notion of sampling really belongs in experimental research (see p. 124) and research that seeks relationships among variables. There is the assumption that you are taking your group or groups for your research from a manageable sample which is representative of a larger population. A population in experimental design means something rather more than the one we speak about in everyday parlance (e.g. ‘The population of the USA is 350 million’). It means the total number of all possible individuals relating to a particular topic which could (if we had all the money and resources we wanted) be included in a study. So, there is a population of teaching assistants, a population of school-aged children, a population of people on benefits, a prison population, and so on. Assuming that the sample is truly representative of this wider population, the findings of your well designed research can then be generalised to the population.

Of course, the sample may not be representative: there may be selection bias of one kind or another. There are many ways in which one can ensure that the sample is representative of this wider population. One is by taking a random sample. This is what it says on the tin: it is a sample that is randomly taken – in the same way that names could be picked from a hat: the process is random. However, it is not good enough for the sample just to be randomly taken. If you took as a random sample of university students the first dozen people you found in the bar on Wednesday night, your sample would be susceptible to various sources of distortion – how do you know that bar-dwellers are representative of the student population generally? It may be that those who go to the bar are less likely to go to the library, or that they tend to be representative of one ethnic or religious group more than another. That is why this kind of sample (the first dozen in the bar) is called a convenience sample, and why a convenience sample has many problems associated with it if you are expecting to generalise from it. To take a true random sample, you would have to ensure that you were drawing a large enough subset of the population so that the chance of getting a distorted picture was reduced to the minimum. You may have noticed that polls from market research organisations, when they are looking at voting intentions, take a sample of around 1,000 people, and this may (or may not, given that the voting population of the UK is over 40 million people) be a satisfactory sample.

A way of improving a sample’s representativeness is to stratify it. In taking a stratified sample you ensure that the sample you are taking reflects in important ways the characteristics of the actual population. So, suppose you were interested in the views of heating engineers, you could ensure a simple form of stratification in the sample you collected.
by matching the sample with facets of the (whole) population of heating engineers nationally. You could make sure that your sample mirrored known features of the population – say, with regard to gender, age, and highest qualification, as shown in Figure 6.1.

**Figure 6.1  Stratifying your sample**

What’s wrong with the idea of samples?

Professional researchers have extended the notion of sample to all kinds of study, not just experimental research and research concerning the collection of data from groups of individuals. So, sometimes people will speak of a ‘sample’ in relation to a case study where only one case is being studied. But this is a quite unnecessary extension of the use of the word ‘sample’ and is clearly ludicrous. For even in a colloquial definition, there
is the expectation that a sample will be in some way representative. Chambers dictionary gives as its definition of sample the following:

**sample säm’pl, n.** a specimen, a small portion to show the quality of the whole.

So, even in everyday usage, ‘sample’ carries the meaning that the sample is in some way reflective of the whole. But often when the word ‘sample’ is used in social research it carries none of this ‘reflective of the whole’ notion. So, a *snowball sample* – which involves the respondent telling the researcher who the next respondent might be, and that respondent doing the same, and so on – does not pretend to any kind of representativeness. Likewise, a *purposive sample*, which involves simply the pursuit of the kind of person in whom the researcher is interested, professes no representativeness. For this reason, these kinds of ‘samples’ are sometimes called *non-probabilistic* samples, because they do not lend themselves to the kind of design on which inferential statistics using probability estimates are used. Personally, I think it would be easier if they weren’t called ‘samples’ at all.

As I indicated earlier, this carrying of experimental meaning of ‘sample’ to other kinds of research is accompanied by dangers since it carries with it the quiet assumption that all of the other paraphernalia of experimental research goes along with it. This should not be the case.

Sadly, this mixing of tenets and ground rules has sometimes been deliberate rather than merely accidental and of course the confusion transmits itself to you as a student – and you have no idea what to make of it. It’s only someone who is as old as I am who can realise what has happened.

What has happened is that when interpretative research was trying to establish its credentials as authentic ‘social scientific’ research forty or fifty years ago, researchers went to great lengths to ape the language of traditional experimental and relational research. They did this to boost the status of interpretative research. So Glaser and Strauss (1967), for example, drew a distinction between what they called the *theoretical sample* and the *statistical sample*. By a ‘theoretical sample’ they meant the amount of sampling that would have to be done in order for the researcher to be sure that no more categories were being encountered as they went through more and more data. They say: ‘As he sees similar instances over and over again, the researcher becomes empirically confident that a category is saturated’ (Glaser and Strauss, 1967: 61). If this sounds a little vague, it’s because it is, with the word ‘empirically’ thrown in just to make things sound more scientific. (When do you think, ‘Ah! I am now empirically confident!’? When do you feel that warm, gooey glow of empirical confidence?) If you are interested, I have explored elsewhere (Thomas and James, 2006) the reasons why interpretative researchers sometimes seem to think that they...
need to parrot the language of experimentalists. My feeling is that it is not only unnecessary to do this (particularly in relation to the sample), but also that it leads to misunderstanding about the nature of interpretative research. Sadly, the feeling that one has to use the language and methods of experimentalists is a relic of the days when interpretative research was felt to be not quite good enough. Even more sadly, it leads inexperienced researchers to inappropriate research designs.

**Variables**

Things that we want to measure in the social world *vary*. (If they were always the same – didn’t vary – we wouldn’t be interested in measuring them!). So, *variables* are measurable attributes of things that change. Age changes, so it is a variable. Scores on tests vary so they too are variables. Anything that can be counted could be a variable: age, class size, time spent sitting down, hair length, reading age, level of stress, number of words written on a page, and so on. We can also treat what could be called ‘on–off’ matters as variables. Thus gender, of which there are of course only two varieties, can be a variable. The amount of a variable is called its value.

**Reliability**

Reliability refers to the extent to which a research instrument such as a test will give the same result on different occasions. Really, the idea of reliability has been imported into applied social research from psychometrics – the ‘science’ of testing people’s individual characteristics such as ability, attainment and personality. In my opinion, psychometrics is where the notion of reliability should have stayed. It is far too heavily drawn on particularly by student social researchers, who spend too much time thinking about it and writing about it in their reports. Because it is something quite concrete (complete with bullet-pointed sub-varieties) in the rather difficult area of methodology, students sometimes fall on it with eager relief: ‘Ah, here’s something I can get my teeth into!’ and spend two or three pages writing about the sub-varieties of reliability in a wholly irrelevant way.

Certainly if you are collecting data you want your measuring instruments to be consistent from one time to the next. In other words, if you are giving a test to a group of children and then give it again soon after, you would expect it to provide much the same results each time: this is called *test–retest reliability*. Or if two different people gave the same test to the same group, you would expect the test to provide very similar results each time: it should have good *inter-rater reliability*. Or you may devise some kind of measure of classroom activity, such as being ‘on task’ (see the example of Karen on p. 127). If this is the case, you will
want to know that the measure you are using is accurately assessing the feature or activity on which you are focusing.

Of course, it is as well to be aware that biases can occur in the use of any instrument – such as teachers giving a test ‘helpfully’ because they like the new teaching programme which is being assessed by the test. But if teachers were craftily or unconsciously introducing bias in this way it would not be a technical matter; it would be a matter for intelligent appraisal. We should be alert to matters such as this (e.g. giving a test ‘helpfully’) and try to minimise their effects. The compiling of types of errors and biases into technical taxonomies (observer error, observer bias, subject error, subject bias, etc.) diverts attention from the real subject of the research, and it suggests that there are technical ‘fixes’ to the problems of any kind of individual or social assessment. There aren’t. Worse, the idea that these fixes may exist may even distort the construction of instruments, so that they may be ‘reliable’ but in fact tell us nothing useful.

It is possible to use formulae to help you to establish a coefficient of agreement across times and/or observers which will give a figure representing the reliability of the instrument being used. However, in my experience, in applied social research done by students at undergraduate and postgraduate levels, these are rarely if ever used – nor do they need to be in most cases. It is far more important that you pay attention to whether the instrument is doing what you want it to do and isn’t picking up something irrelevant or something that is simply easy to measure (such as ‘out of seat’ as an indication of being ‘off task’ – but are pupils necessarily ‘off task’ when they are out of their seats?). Does it matter if, given on two occasions, the same instrument says something rather different? Surely it is almost bound to, given the nature of people (as distinct from the nature of nitrogen).

As in the case of sampling, there is a confusion that sometimes exists between different kinds of research, with an assumption that reliability should apply in all kinds. It shouldn’t. In interpretative research you are interpreting on the basis of you being you, interviewing someone else being them. Who you are – your ‘positionality’ (of which more below) – will affect this interpretation and you would not expect someone else to emerge with the same interview transcripts as you. So reliability is, in my opinion, irrelevant in interpretative research.

Validity

There are two meanings to validity in social science research, but before giving those, I should preface my comments by saying that everything that I have said about reliability applies also to validity: it is a concept imported from psychometrics and from experimental design; it is discussed far too often; it can distract you from the proper purpose of your research.
What are the two meanings? I have separated these meanings because, as with reliability, some students appear to latch on to the notion of validity and beyond this latching fail to understand the basic differentiation of types of validity (partly because in test construction several subtle differentiations are enumerated). The two types I am enumerating here are instrument-based validity and experimental validity. Unfortunately, the types have become hybridised and confused in such a way that it is now nigh-on impossible to understand what validity means. Here is my simplified version.

**Instrument-based validity**

With a measuring instrument such as a test, validity is the degree to which the instrument measures what it is supposed to be measuring. *Construct validity* is the Big Daddy of instrument-based validity – the sun around which all the other little instrument-based validities spin. Construct validity is the extent to which the results of a test (or another instrument) correlate with the theoretical construct for which it is seeking to act as an assessment. So, do the results from the test for extraversion, scores on this should largely parallel judgements made by a psychiatrist making judgements about extraversion – the latter being taken to be an unimpeachable assessment of this personality trait (and this is of course in itself questionable, but we won't dwell on this). All the other forms of validity you will find are offshoots, really, of construct validity. There are, for example, *content validity* (does it cover all of the necessary content?), *predictive validity* (does a test in intelligence, say, predict the kinds of outcomes for how intelligent people are supposed to fare?), *face validity* (does it look as if it's doing what it should be doing to the person being tested – has it, in other words, got street cred?) and *ecological validity*. They all really are facets of construct validity and there is no real need to try to disaggregate them in your mind unless you are constructing a psychometric test.

Again, the taxonomy of types here is an example of social scientists becoming more interested in the methodology than the subject, and the existence of such a complex taxonomy of types in my opinion distracts from an intelligent case-by-case appraisal of an instrument's use. For example, an IQ test could be constructed that had high marks for construct validity, but this would disguise to the lay reader the extent to which the concept of IQ is itself challenged as a notion – a construct. It's the danger of technical considerations taking precedence over others.

**Experimental validity**

There are many things that happen in psychological and social life that can mess up an experiment. The extent to which the design of an experiment attends to these and eliminates them is reflected in its ‘internal validity’. (Why ‘internal’? Don’t ask me, ask Campbell and Stanley (1963)
who came up with the term.) Chemists (lucky them) can pressurise some nitrogen in a bell jar and see how much it contracts to demonstrate Boyle's Law. They don't have to worry about the nitrogen getting older, or having been exposed to the test before, or deciding to quit the experiment for family reasons, or being a particular kind of nitrogen that doesn't like bell jars. No such luck for social researchers. If we construct an experiment we have to keep all of these things (and more) in mind. If an experiment is somehow constructed (designed) in such a way that it manages to eliminate all of these threats to the conclusions being taken seriously we can say that it has good internal validity.

One particularly important threat is that of mistaking causal direction. Is the wind caused by trees shaking their leaves? Are eggs the right size for eggcups because of good planning by hens? I have explored the latter elsewhere in relation to how research on reading difficulty is interpreted (see Thomas, 2002), and it is of course always necessary to consider direction of causation. What causes what? To you and me it's daft even to suggest that the wind is caused by mass leaf-shaking by trees, though a visitor from another planet which experienced no wind, not even a gentle breeze, might be tempted to explore the proposition. It becomes more difficult, though, when an association is found between something like skill in auditory memory and early reading success. It is easy to run away with the idea that the memory is the cause of the reading skill. In fact, though, it may be that reading ‘trains’ the auditory memory, and that better reading therefore is actually the cause of the better memory.

However, as is so often the case, the problem is not technical. The problem is usually not that no one has considered the possibility that \( y \) may be causing \( x \) rather than \( x \) causing \( y \), but rather that they do not want to consider the possibility, for any one of a variety of reasons. (The sinister implications of this are discussed very nicely by Gerald Coles (2000) in Misreading Reading: The Bad Science That Hurts Children.)

Generalisation and generalisability

In everyday life, when we make judgements about the future – predictions – these are usually on the basis of generalisation from experiences we have had in the past. Events that repeatedly occur in certain circumstances enable you to generalise – to say to yourself that these events will tend to occur in these same kinds of circumstances in the future. You notice that when there are no clouds it doesn't rain, so you make a reasonable generalisation about the low likelihood of precipitation when there are no clouds in the sky. While such generalisations serve us well most of the time, they do not match up to the expectations that science has when it seeks to offer laws and theories – which are also based on generalisation but, one might say, generalisation-plus. Scientists cannot proceed simply on the basis of common-or-garden generalisation. As the great philosopher Bertrand
Russell (1956: 91) put it, the person who says that unsupported bodies in air fall ‘has merely generalized, and is liable to be refuted by balloons, butterflies and aeroplanes’ [italics added].

What Russell is saying here is that while generalisation is important, it has to be more than a mere rule of thumb based upon everyday observation of life’s patterns. Science’s generalisations have to have a bit more oomph about them than that. Good generalisations – generalisations that provide accurate predictions – are at the cornerstone of scientific progress. It’s just that we have to find ways of making our generalisations more than the ‘mere’ generalisations Russell was talking about. And it is even more difficult to make sensible, accurate generalisations in social science than it is in physics or chemistry.

In social research, summarisability is about the extent to which research findings can be applied in settings other than the setting in which the original research took place. To make such a generalisation requires that some quite strict criteria are met about the way that the study is set up. In the main, the extent to which you can generalise hangs on the extent to which your sample is representative of the population. As I noted above in discussing sampling, the whole idea of taking a sample is that it is a sample of a wider population.

Remember that summarisability is of importance only when you want to generalise. If you are conducting a case study with only one subject, it is more or less meaningless to worry about generalisation (though people do worry about it, I can assure you). You cannot generalise from one case. The Germans have a phrase, *Einmal ist keinmal*, which means ‘What happens only once might as well not have happened at all’. I personally wouldn’t go quite that far, but I would agree with a watered-down version of this, namely that we can learn no general lessons from things that happen only once. But how many more do we have to study before we can say that we can generalise? This depends on a great many factors, not least the adequacy of your sample as a representative sample.

As with so much in social science, there are no hard and fast rules about generalisation. Many would no doubt go along with the short-story writer Damon Runyan and his comment in *A Very Honourable Guy*: ‘The race is not always to the swift, nor the battle to the strong. But that’s the way to bet’. In other words, one’s generalisations may be far from perfect as ways of judging the future, but one can, to varying extents, rely on them as rules of thumb in ordering our lives or in interpreting the findings from our research.

**Positionality**

In interpretivist research there is an assumption that knowledge is situated in relations between people. This is sometimes called *situated knowledge.* With this assumption taking the foreground, the person doing the research takes a central role in the interpretation – in the discovery of this
situated knowledge. The researcher therefore has an undeniable position and this position affects the nature of the observations and the interpretations that they make. There is an acceptance in this of the importance of the person – their likes and dislikes, their backgrounds and their pastimes, their vested interests and expectations. The researcher is an active, not passive, agent in acquiring knowledge of the processes, histories, events, language and biographies of the research context. Because of the importance of the nature of the relation between the researcher and research participants, the researcher’s biography – including class, gender, ethnicity, ideas and commitments – needs to be made explicit.

There is also the assumption that in doing research you are in some sense serving a certain set of needs, and these needs are not necessarily unequivocally good or altruistic. They will not necessarily be about the good of education (in the widest sense) or the child. They may be self-serving or benefiting a particular group. This should be made explicit.

The explicit acknowledgement of this may present some dilemmas. Let’s imagine that you are a newly qualified teacher having completed your PGCE recently and you wish to study now for a Master’s in Teaching and Learning for which you have to complete a research project. You discuss it with Zena, your Head of Year, and she suggests that you use your research project to evaluate the school’s new inclusion policy. You already know from staffroom discussion that the policy is thought to be a joke – devised to keep Ofsted and the local authority happy but serving little purpose in practice. You are therefore faced with a predicament. Given that you intend to undertake an interpretative study, how would you construct your eventual write-up, for the university will be expecting a discussion of positionality – of your interests, uncertainties, allegiances, and so on – while Zena and the school management may expect something entirely different, something that pretends a cleaner, more dispassionate and supposedly ‘objective’ assessment of what they assume to be ‘the facts’. So, for you, this idea from Zena is non-viable as a potential research project because you would have to make all of this explicit in the write-up. And the issue now becomes ‘How will I find a way to reject the Head of Year’s suggestion?’, given that the school is giving you some time in your timetable to complete your project. The problem here is about the conflict between your integrity as an interpretative researcher and the expectations and interests of the school.

In presenting interpretative research you should accept your subjectivity and not be ashamed of it or afraid of it. Given that it is central, your dissertation or thesis should be written as if you realise the underpinning principles that guide the conduct of this kind of research. An interpretative study will therefore be reported in an entirely different way from an experimental study. You should begin – right at the beginning, in the introduction – with a full discussion of positionality: of yourself, why you are interested in this topic, what your personal circumstances are, and so on. You will always write in the first person, saying, for example,
'I believe that …' rather than 'The researcher believes that …' This may seem obvious, but it is a common mistake for students conducting interpretative research – research that demands that they put themselves at the centre of the analysis – to report their work as if they have just conducted an experiment and to omit any information about themselves. If you do this kind of research, readers need to know who you are and where you stand, metaphorically as well as literally.

**Triangulation**

Triangulation is a term that has been borrowed from surveying and geometry, where it refers to the use of fixed reference points organised in triangles. By knowing an angle and the length of two sides of a triangle, the third can be accurately calculated and distances can be checked and cross-checked.

In social science the term is used simply in a metaphorical way, based on its origins in geometry and surveying. There is no intimation that triangles should be involved, or that things have to be done in threes. Rather, the term is used to indicate that viewing from several points is better than viewing from one. Given the instinctive uncertainty – the critical awareness – that should be the hallmark of the good social science researcher, another viewpoint or another analytic method may make you decide to reject an explanation that you had come up with from your first analysis of findings. Or it may encourage you to have more confidence in the explanation you proposed on the basis of your first analysis. For this reason, using several methods, or viewing things from several directions, is sometimes built into a piece of research at the beginning.

Denzin (1978) outlined several types of triangulation, including **investigator triangulation** where more than one person is involved in interpretation and analysis, **theory triangulation** where more than one theoretical framework would be involved and **methodological triangulation** where more than one method would be used to collect data. The last of these is the most common meaning, though the emphasis on method (that is to say the data-gathering method) should not be taken to imply that the use of more than one design frame is excluded in the concept of triangulation. So I would add to Denzin’s categories **design frame triangulation**. In other words, you would be triangulating if you used both a case study and a longitudinal study together in the same piece of research.

Opinions differ on the need for triangulation. Some interpretative researchers argue that a piece of interpretative research has value and completeness in itself. It doesn’t need any verification from other kinds of research. It has integrity as a singular inquiry. The argument for the integrity of the singular case, singularly done, is a
powerful one. But the argument for corroboration, for the need for alternative kinds of evidence, each corroborating the other, is to my mind even more powerful, and triangulation is really simply about corroboration. I discussed the importance of corroborative evidence on p. 16.

The design frames

Action research

Action research is research that is undertaken by practitioners (for example, teachers, social workers, nurses or doctors) for the purpose of helping to develop their practice. It is usually done at the same time as performing that practice. The central aim is change and the emphasis is on problem-solving in whatever way seems most appropriate. It is flexible in design: the assumption is built in firmly at the beginning that as the research proceeds you will go back to revisit your aims, assumptions, beliefs and practices, think about all of these critically and then revise them.

It may be done by an individual or a group, perhaps in collaboration with a consultant such as a university tutor. For this reason it is sometimes called ‘participatory’. In other words, the assumption is that it is different from the way that traditional social research has sometimes been conceived: that is to say, as research done by researchers on subjects. It is research primarily done by the ‘subject’ or practitioner, with the aid of others.

The idea of action research came from the great social psychologist Kurt Lewin. In his paper ‘Action research and minority problems’ (Lewin, 1946) Lewin was critical of social research, saying that ‘... research that produces nothing but books will not suffice’. He describes action research as ‘... research leading to social action’ using ‘a spiral of steps, each of which is composed of a circle of planning, action, and fact-finding about the result of the action’ (see Figure 6.3). The basic idea is of a continual refinement of your thinking, built upon a foundation of reflection about the problem and ways of going about solving it. The process could be shown as Figure 6.4.

So action research is a bit like a coil or spring, where you are continually moving forward up the coil by reflecting on action and changes that you have made.
Much has been written about action research over the years and there are many different ideas about the way that it should or should not be done. My own view is that there are four basic ideas at the core of action research: that it ...

1. is research done by practitioners, at their own behest – not someone else’s;
2. is primarily about developing practice and empowering practitioners;
3. involves a commitment to change and to action based on reflection;
4. involves moving forward, always building on what you are discovering, using the process of planning, reflection and re-planning (shown in Figure 6.4).

Beyond this, it can take almost any form that you wish it to take. Indeed, Jean McNiff and her colleagues (McNiff et al., 2003) suggest that action research is more of a ‘form of dialogue’ than a technique, and that it is about practitioners thinking for themselves and making their own choices, asking themselves what they should do and accepting the consequences of their actions.

For me, this sums up the spirit of action research. Beyond this, it may take shape in almost any form. For example, you may wish to do a small experiment within an action research frame. It may involve a case study. Or you may undertake action research that involves some kind of evaluation.

Example 6.1
Emily is a newly qualified teacher of history in an inner London secondary school. One of her Year 9 students, Rashid, is surly, disengaged and withdrawn, but when challenged can be aggressive and even physically violent to other students or to (Continued)
staff. At the end of last term he had refused to hand in a pen at the end of the session and when confronted pushed Emily out of the way to exit the classroom. For this he was temporarily excluded from school.

In working with her local university for a postgraduate qualification, Emily decided to do an action research project based not only on Rashid but also on challenging behaviour from other students. The action research framework involved trying to develop her practice in such a way that confrontational incidents would occur less often. The plan is shown in Table 6.1.

A number of issues arose for Emily in doing this; for example, with regard to point 2, some rethinking was necessary about the nature and causes of challenging behaviour. In point 3, some innovative ideas came from separate discussions with her head of year and her university tutor. The latter suggested trying to make a special relationship in some way with Rashid and, capitalising on a brief moment when Rashid had shown an interest in one of her lessons on evacuated children in the Second World War, Emily asked her uncle, who had been an evacuee, if he would email Rashid about his experiences – which he was glad to do, sparking up an unlikely web friendship. Still on point 3, her head of year helped Emily to devise a list of appropriate and inappropriate behaviours to take when physically challenged by a child.

Having completed this ‘circle of steps’ and reflected on the consequences in the final step, Emily was then able to move on to the next circle of steps, having seen

---

**Table 6.1** A circle of steps

<table>
<thead>
<tr>
<th>Action research steps</th>
<th>Questions/issues/actions (in brief)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Define the problem</strong></td>
<td>The incidence of confrontational behaviour from students and one student in particular.</td>
</tr>
<tr>
<td>Examine the idea or problem and gather information about it</td>
<td>Is it a problem just for Emily? How far is it a problem of her making? Is it the school’s problem or the student’s problem? Read around the area of challenging behaviour; ask colleagues what they have done in similar circumstances with Rashid and others.</td>
</tr>
<tr>
<td>Plan action</td>
<td>Define what you mean by ‘challenging behaviour’; keep a record and a diary of incidents of challenging behaviour; when does challenging behaviour occur, in what circumstances; plan ways of developing relationships with the students (challenging and unchallenging); list action to be taken if it occurs.</td>
</tr>
<tr>
<td>Take action</td>
<td>Take the action planned in Step 3.</td>
</tr>
<tr>
<td>Reflect on the consequences</td>
<td>Examine records and diary. Consider the effects of the action taken. What was effective? What wasn’t? Discuss with colleagues, advisor, consultant and/or tutor. Move to the next action cycle.</td>
</tr>
</tbody>
</table>
what worked and what didn’t work. Rashid certainly responded to the efforts Emily had made on his behalf and began to act, if not with enthusiasm, with more courtesy, respect and a new-found gentleness. She considered a number of possible actions having completed the first circle of steps: she considered, for example, how far the action she had taken with Rashid was relevant and generalisable to other students in her class. She therefore decided to run a number of small focus groups with three or four of her students in each, and various focus materials comprising newspaper and magazine articles for discussion. With these, she would try to do with the others what she had done with Rashid – find something of interest and use this to help develop a relationship with the students.

Part of her reflection was also to realise that the special relationship she was nurturing with Rashid as part of the research was extra-special, even with the focus groups she had subsequently planned. Given that this might create additional difficulties with other students in the class, she included as her move to the next circle of steps some special activities also with other children in the class – for example taking a group of the quieter and more able children to a local museum at the end of term.

Case study

A case study involves in-depth research into one case or a small set of cases. The case may be a child, a teacher, a class, a school, a social services department ... the list could go on. The aim is to gain a rich, detailed understanding of the case by examining aspects of it in detail. The data you collect can be from different facets of the question you are examining, and these data – perhaps from numbers or from interviews or informal observations – may be combined to tell your finished story.

There is no intimation in the case study that you will be generalising from this case to others. How could you? It’s one case. In other words, you are not studying this case in order to understand others. You are studying it in order to understand it in itself. For the research methodologist Martyn Hammersley (1992) this choice of one case (or a small number) is made with a trade-off in mind. You are choosing this very restricted sample in order to be able to gain greater detail, but at the expense of being able to make useful generalisations to a broader population.

How do you choose your case? Robert Stake (1995: ix) suggests the following advice:

A case study is expected to catch the complexity of a single case. A single leaf, even a single toothpick, has unique complexities – but rarely will we care enough to submit it to case study.

In other words, you don’t study a particular case just for the sake of studying it. There have to be particular circumstances surrounding the case that make it of special interest. Stake proceeds:
We study a case when it itself is of very special interest. We look for the
detail of the interaction with its contexts. Case study is the study of the
particularity and complexity of a single case, coming to understand its
activity within important circumstances.

As he goes on to suggest, the choice made of the case may be no ‘choice’
at all, because we have some kind of intrinsic interest in that case, whether
it be a child with particular difficulty or the involvement in some pro-
gramme of innovation at, say, a school or in a social services department.
This is what he calls an ‘intrinsic study’. Or it may be what he calls an
‘instrumental study’ – in other words, undertaken for the purpose of help-
ing to understand or to illustrate something in
relation to a research question.

It depends what you want to find out. The tra-
dition of case study has come to be associated
with ‘naturalistic’ inquiry, but it should not be
seen like this. Case study can include a range of
as many different methods and procedures as
necessary for understanding what is going on in
a particular situation. It’s like an umbrella, cov-
ering a whole range of inquiry-activity.

A word of warning about case study – the
social systems analyst Geoffrey Vickers in his highly influential book *The
Art of Judgement* puts it this way:

Case histories are a laborious approach to understanding. For situations
are so varied that even a large number of cases may be a misleading sam-
ple, whilst each is so complex that even a detailed description may be too
summary; and none is comprehensible outside the historical sequence in
which it grew. (Vickers, 1965: 173)

Vickers’s reservations are meant about big case studies undertaken by pro-
essional researchers and intended to inform institutional change. On the
smaller scale, undertaken in a classroom, say, and intended to inform per-
sonal understanding, the concerns about being ‘laborious’ are not evident.
However, the general points he makes are worth remembering: generali-
sation is not possible; detail and contextual understanding are essential. It
is the rich picture with ‘thick description’ (see p. 202) that we need.

**Example 6.2**
Robin is in his second year of teaching as a Year 6 teacher in a primary school. Last
year, in the run-up to the SATs, he noticed some disturbing signs of unease and
anxiety among his pupils: more absence and illness and a general sense of disquiet
that he couldn’t put his finger on. For his project as part of a university course, he
decided during the next round of SATs to undertake a case study of his class,
focusing in a detailed way on the attitudes and responses of pupils and teachers.
to these assessments. Starting with the notion of a link between these changes in
behaviour and the national testing programme, he knew that his case study
would relate the one to the other.

Since the government’s regime of testing affected him most directly in Y6 (since
this is age at which the SATs are given), Robin knew that of all the teachers in the
school it was his own views on the impact of the assessments that were most rel-
levant and important. He was particularly concerned about the emotional impact
on the children, so in the six weeks leading up to the SATs he kept a diary of his
impressions of the behaviour of the children and his own feelings over that time.
He committed to spending 15 minutes at the end of each day writing the diary,
which would produce 30 diary entries (6 weeks × 5 days) for subsequent analysis.
He also committed to focusing on five children in his class, whose work and
demeanour he would monitor closely but informally over the same period. He
would for specified periods observe the children informally. He would take notes
about the children’s work, their attendance and their punctuality and he would
make a point of talking to each of them briefly at least once a week, asking about
their work and about how they were feeling. He would then keep an audio diary
of his impressions of these encounters and the children’s comments.

Although the SATs affected his class most directly, Robin had a suspicion that
the culture of assessment made its effects felt through the age range, so he
decided to interview all of the other teachers in the school for their opinions on
the influence that SATs had made, even if only informally, on their own work and
the attitude of children to learning.

Robin’s analysis concentrated on drawing themes from the data, for example
about anxiety in the children and the formalisation of style of teaching by himself
and other teachers. He related these themes to the form in which expectations
about success in the SATs were communicated to teachers by government depart-
ments and their agencies, and the implicit or explicit messages that accompanied
these expectations.

Possible questions for Robin in the case study:

• How would he choose the five children?
• What ethical and access issues would arise?

Classic examples of case study are in Stephen Ball’s (1982) Beachside
Comprehensive and Colin Lacey’s (1970) Hightown Grammar, which each
give a detailed analysis of what goes on in one school over a period of
time, and James Patrick’s (1973) A Glasgow Gang Observed, giving a
detailed narrative of what went on in the gang from the point of view of
a young man infiltrating it (I discuss Patrick’s work more on p. 77). These
were doctoral studies and most dissertation work will not be expected to
go into any way near such detail as these authors did in their work.
However, they are enormously useful to read – in the work of both Ball
and Lacey for examples of the kind of data-gathering that can be done
and the way that case study work can be linked intelligently to national
policy, and in Patrick’s work for the style of involvement and reporting
discourse that he uses, which is as good as reading a novel.
Ethnography

If you do an ethnography you are working right in the middle lane of interpretative research (see pp. 75–8). This is as interpretative as it gets. Confusingly, this design frame may be called ‘case study method’ but as we have just seen, case study can cover a much larger range of study methods than merely this one.

The term ‘ethnography’ comes from a field of social research that emerged in the early part of the twentieth century as a branch of anthropology, the study of humankind and its cultures. Ethnography evolved as a response to what was considered to be an inappropriate way of studying other cultures or communities, which had before the 1920s been treated almost as scientific objects of study by anthropologists. Writing was often judgemental, wearing the moral spectacles of the Western viewer. And if you are interested, look at Stephen Jay Gould’s (1996) *The Mismeasure of Man* to see how some anthropological ‘scientists’ used to do their work by comparing the size and shape of the skulls of different races. Displacing this style of work, the new ethnography aimed to get to the heart of people’s understandings of life through fieldwork with people rather than supposedly objective study on them. James P. Spradley, a great exponent of ethnography in the middle of the twentieth century, put it thus:

Field work … involves the disciplined study of what the world is like to people who have learned to see, hear, speak, think, and act in ways that are different. Rather than studying people, ethnography means learning from people … Instead of collecting data about people, the ethnographer seeks to … be taught by them. Spradley (1979: 3)

Spradley draws on the example of a young American ethnographer, Elizabeth Marshall, who, interested in the culture of the Kalahari Bushmen, went to the Kalahari with her family in the 1950s. Marshall describes her meeting with a young woman:

Presently she smiled, pressed her hand to her chest, and said: ‘Tsetchwe.’ It was her name.

‘Elizabeh,’ I said, pointing to myself.

‘Nisabe,’ she answered, pronouncing after me and inclining her head gravely. She looked me over carefully without really staring, which to Bushmen is rude. Then having surely suspected that I was a woman, she put her hand on my breast gravely, and, finding that I was, she gravely touched her own breast …

‘Tsau si’ (women), she said. (Cited in Spradley, 19: 3–4)

You can glean from this brief example of Marshall’s writing that the purpose of the ethnographer was to understand from within, treating
the people with whom she is working with respect, as different but equal. It is typical of the style of reporting of ethnographers, who work through *fieldwork*, which Spradley describes as ‘... asking questions, eating strange foods, learning a new language, watching ceremonies, taking field notes, washing clothes, writing letters home, tracing out genealogies, observing play, interviewing informants, and hundreds of other things’ (ibid.: 3). (Note, though, that ‘fieldwork’ now means something much more general, i.e. any empirical work of any nature.)

He goes on to note that for Tsetchwe to turn the tables and go to a small Wisconsin town to try to understand its culture would involve her doing everything that Elizabeth had to do, which first and foremost would mean ditching a belief in *naive realism*, which is to say the belief that ‘love, rain, marriage, worship, trees, death, food, and hundreds of other things have essentially the same meaning to all human beings’ (ibid.: 4, original emphasis).

Ethnographers have to try to forget the meanings that these phenomena have for themselves in order to understand their significance in the other culture. This is hard of course and it means, usually, a long immersion in the cultures they are studying – learning new languages and learning about unfamiliar customs and practices – to try to understand how sense is made out of these and how meaning is invested in them. In order to make this sense, ethnographers have to use their own resources as people: as Bob Burgess (1982: 1) puts it, ‘The main instrument of social investigation is the researcher’. As an ethnographer, there is no attempt to deny your own personal knowledge or to put it to one side, though you might be expected to take an entirely new perspective on the scene that confronts you. In the sense that the person of the researcher is central to the process of research (and you are not in any way attempting to be ‘objective’), you must use your knowledge of people, social systems and structures and how they relate rather than rejecting this knowledge, but use it in a way that enables you to ‘see outside yourself’. You can do this by, for example, using Geertz’s thick description (discussed on p. 202). In other words, look to see what turns the twitch into a wink or the parody of a wink. Use your understanding of people to understand this situation that you are focusing on, and tell the reader how you are doing it.

Another facet of the positioning of *yourself* as an instrument of investigation is that you are a participant in the situation that you are observing and studying, which is why the term *participant observation* is often associated with this kind of research. You are not, in other words, trying to be detached or invisible, but rather engaged, fully involved, and gaining insights from this engagement. Participant observation is not limited to observation, pure and simple, though. It consists of watching, certainly, but also conducting interviews, listening to conversations, keeping a diary, taking notes and much more – anything that will help you to record and understand what is going on.
For an interesting (and short) example of the ethnographic genre, see Clifford Geertz's *Deep Play: Notes on the Balinese Cockfight*, to which an Internet link is given under Further Reading at the end of the chapter.

Ethnography is now a well accepted way of doing certain kinds of applied social research. In fact, some of the most influential research in education has been ethnographic in character. However, doing a small ethnographic project presents some major differences from the classic studies of which I have given examples. For a start, unlike the professional anthropologists cited here, you are unlikely to be entering a culture which is unfamiliar to you. Rather, you are likely to be entering a very familiar situation; in fact, the situation in which you are interested may be one with which you are intimately connected, perhaps as an employee or a regular visitor. Second, you will not have the time for the deep immersion of the professional anthropologists – but this may not be as necessary because of the familiarity you already have. This in itself has pluses and minuses: you may know the ‘stage’ and be an accepted ‘actor’ on it, but this means that you have to work extra hard to see it in a fresh light.

Though your immersion in the world of your study will be different in these important respects from the immersion of the ‘classic’ anthropology, beyond this the rationale behind ethnographic work is the same, and is about participation, engagement, thick description and understanding.

**Example 6.3** Qualifying as a teacher three years ago, Amy is now head of maths in a large inner-city comprehensive school, but she wishes eventually to move into learning support. As part of a Master’s degree in education she decides to do her dissertation research on the way that support works in the school. She wishes to understand how students see support, and what support can offer them. In particular, she asks whether there are any ways that the support department in the school can enhance its offer to its students.

She decides that her research project will be an ethnography of support: how it works and how it is viewed and understood by students and staff. She negotiates with the Head of Learning Support to use some of her regular non-contact time to work as an assistant in the learning support department. She asks to be treated as a teaching assistant for these periods, and ensures that she will be working only with classes with whom she has no dealings in her maths teaching. Her sessions of involvement with the department will include attendance at meetings and work with two classes per week as a participant observer: one science class and one languages class.

She prepares for this by planning carefully to observe each of the environments into which she goes. She decides not to take notes as she does this, as this will affect the way that the students and staff react to her during the sessions. However, she will write a diary immediately (or as soon as possible) after every session. She will also conduct unstructured interviews with both teacher members of the learning support staff as well as two of the teaching assistants associated with the department. The interviews will give free rein to these staff
to express their views about the operation of support. On top of this, she will interview four students – two from each of the classes that she is working with. Further, on her movements around school, she will keep an eye and an ear open for comments or actions that in any way give clues as to the way in which learning support is viewed in the school. Is it seen genuinely as a way of including all students, or merely as a way of dealing with students who are seen to be ‘special needs’?

An excerpt from Amy’s diary is as follows:

I make my way to Xanthe’s class again, a little late, having been talking to the head of support. I haven’t spoken in detail to Xanthe yet about the team teaching idea. Last week’s meeting with her was too tense to broach the topic. When I come in this week the situation seems rather similar. Xanthe is taking the class from the front and the children are messing around as they were last week. Xanthe holds up a small glass tank with a funnel upside down over something (a small plant, I think) and compares it with another similar set of objects. The children have to compare the two sets and say how they are different:

CHILD: There’s more water in that one.
CHILD: One’s all dirty – yeugh.

The session is to show that carbon dioxide is taken up by plants:

CHILD: That dirt’s carbon dioxide!

Explanations are again punctuated with inconsequential intrusions:

CHILD: Anyone got a rubber!

These become so intrusive that Xanthe eventually loses her temper and raises her voice. This has some effect on the students, who temporarily quieten down. She capitalises on this by saying:

TEACHER: Right we’ll have two minutes of complete silence – not a word.

The children respond to this: they are now clear about the rules and what is expected of them. Again I feel redundant – not only redundant, but worse that I am compounding the situation by my presence, and perhaps that Pauline the teaching assistant is in a similar position. She must feel some degree of inhibition about doing what comes naturally to her to quell the sorts of problems that she is experiencing. Again, there is no way that I feel I can intervene in the immediate situation without making things worse …

Amy analyses her findings using a variety of methods (some of which are described in Chapter 9). Her main aim in the analysis is to understand how learning support is viewed and understood in the school, so she is careful to try to see through the eyes of her informants, with that view moderated and influenced by her own reflections as recorded in the diary. Each element, though discrete, goes to inform each of the others and in her discussion she integrates and synthesises the analysis from each.

Possible question for Amy in undertaking the ethnography:

- How will she be able to use this diary when it makes comments that are implicitly critical of a colleague?
Evaluation

Evaluation research is probably the most common kind of research done by professional researchers in education and the social sciences, being undertaken to assess how effective a programme of activity has been. When done by professional researchers it is often carried out as a form of what is sometimes called 'contract research': a university researcher is contracted and paid to provide an independent assessment of some new initiative. A government department, for example, may pay a university research team to look at how effective the policy of employing greater numbers of teaching assistants has been.

The pattern is given by Pawson (2006) and corresponds roughly to Figure 6.5.

As you can see from Figure 6.5, evaluation research is different from action research, since there is – in the usual form of evaluation – no assumption that what is being studied feeds back in any systematic way to the activity, intervention or initiative being evaluated. (In action research, remember, there is the assumption that what is being discovered during the research actively feeds back and contributes to the development of the programme of change.) Only after the evaluation has been completed may the person contracting the evaluation decide to continue with the programme, modify it in some way or completely ‘pull the plug’ on the innovation – depending on the findings of the evaluation. Ultimately, it may be the case that enough pieces of evaluation are conducted in different places and in different circumstances and by different teams for there to emerge a body of research about the particular kind of innovation or programme in question. This body of work can then be drawn together and examined in a systematic review that will inform the development of future practice, but this is a blunter process of feedback than that used in action research.

You will appreciate from this discussion that evaluation research is often large in scale. While this is the case, it is quite possible to set up evaluation on a smaller scale for an undergraduate or postgraduate research project. Here you will be looking to assess the impact, effectiveness or other consequences of some newly introduced programme, usually within the frame of an institution’s set of policies or of some initiative that you or colleagues have personally introduced. The change is introduced and the evaluation examines the consequences.
Ideally, an evaluation will look at what happens (1) before, (2) during and (3) after an innovation is introduced, with as long a period of examination as possible in each of these three phases. In practice (and certainly this is the case for student project work), evaluation may begin only during the period of implementation, or even after this, though this is of course not ideal.

Given that an evaluation focuses on before-and-after elements there may be a temptation to think that it has to collect data relating to pre-specified variables – things that you can count – so that it is possible to tell whether there has been a measurable increase or decrease due to the intervention. Certainly this may be the case – for example, reading ages could be taken before and after introducing a new system of reading in groups, and if the evaluation took place over a long enough period it may be possible to draw some conclusions about the effectiveness of that new system. However, an evaluation may collect almost any kind of data that those who are involved in the implementation find acceptable as indices of effectiveness.

Example 6.4

Gemma is studying for a BA in Education and English and is on a placement in a primary school – a placement which her university expects to provide Gemma with information for a double-module research project. Knowing that the school is implementing a new policy for playground supervision she decides to make the implementation of the policy the subject of her research. The school is completing the policy and just preparing for implementation as Gemma is able to start her research.

After negotiating with her class teacher and the head teacher, she decides to undertake a three-part evaluation. **First**, at the beginning of term she will collect some data about playground behaviour before the implementation of the policy. **Second**, she will examine the implementation of the policy itself. **Third**, at the end of term she will collect data about playground behaviour again, subsequent to the completion of the formal process of implementation.

Several constraints from the ‘real world’ could have affected the data that Gemma could have collected here, but she capitalised on the fact that she was just in time to collect some ‘before’ data before looking at the process of implementation.

Gemma devised a checklist for observing in the playground which she used in the first and third phases (in other words, in the before and after phases). The checklist, which she would complete at set times during two weeks of observation before and after, included recording observed incidents of physical aggression from child to child, incidents of verbal abuse and other incidents of teasing or aggression. She would record the circumstances of these. At the same time she would record the circumstances surrounding these incidents and briefly interview staff on duty during each of the periods of intervention, asking also for reflections on the effects of the policy on the circumstances in question. She would emerge with data about ‘critical incidents’ backed up with (or ‘triangulated with’) informed comment from staff.

*(Continued)*
For the assessment of the implementation (i.e. the second stage) Gemma undertook interviews with staff – three teachers, three teaching assistants and three midday assistants – about how they felt about the policy, how they had been involved in devising it, how easy it had been to conduct it in practice and how optimistic they were about change. She interviewed six children – one from each year group – asking some basic questions about the policy (to find out how much they knew about it) and whether they felt it was altering playground behaviour.

Her analysis focused on bringing together the different kinds of data in each of the three phases, and then trying to understand how the implementation had happened in the context of the total evaluation. In other words, understanding how and why there was a fall in the incidence of difficult and bullying behaviour at playtimes but not at dinnertime. Her conclusion – properly tempered with caveats about the limitations of the data – was that the policy worked better when implemented by teachers and teaching assistants than it did when implemented by midday assistants, who had not been involved in the writing of the policy.

Possible questions for Gemma after the evaluation:

- Were there enough participants involved in the research to find out what was going on?
- How could play-fighting be distinguished from aggression in the observation?
- Were there enough ‘data points’ to be sure that the evaluation was reliable?
- Should the evaluation have taken place over longer periods before and after the implementation?

Experiment

As we noted in Chapter 1, one important branch of research is that which seeks to find out whether, or to what extent, one thing causes another to happen. This is in fact often the aim in social research – in other words, trying to see if \( x \) causes \( y \) – but the different design frames are differently suited to answering such questions. Experiment is taken by some to be the most reliable way of determining that such causation is occurring.

In everyday parlance, ‘to experiment’ can just mean to try something out, and ‘an experiment’ can simply mean an investigation of some kind. In science it means something more precise: it means a test done under controlled conditions to prove or falsify an idea, a conjecture. However, in social research it means something even more specific:

- it is about demonstrating cause and effect;
- it concerns the extent to which we have to control the conditions in the situation of interest to demonstrate that cause–effect relationship.

The reason it is difficult to say that \( x \) causes \( y \) in social research is that there are a very large number of factors and issues at play in any social situation. If you change one of these factors and claim that the change you have
made has been responsible for the improvement (or deterioration) in another factor you are likely to be challenged by someone saying ‘How do you know that it wasn’t a, b or c (and not x) that caused y to change?’ Using the same example I gave earlier, if a teacher introduces a new scheme of reading and she can show that the reading age of the children in her class rises following the introduction of the new scheme, how can she know that it is the reading scheme that has caused the improvement in reading? Couldn’t it perhaps be that the children would have improved anyway – even with the old reading scheme? And if the improvement looks dramatic so that the teacher says ‘This extraordinary improvement can only be put down to the new scheme’, couldn’t the improvement perhaps be attributed to the new energy that the teacher is giving to her teaching, given the stimulus to her interest provided by the new scheme?

Experiment seeks to rule out the possibility that factors such as this (and many others) could be responsible for the change. In its simplest form it seeks to isolate two variables – the one that you assume may cause change (e.g. the reading scheme), and the one that you are assuming may have change effected on it (the reading age in the example that we have been using) – from the myriad variables that might be at play in a situation. In sciences such as physics or chemistry an experimenter holds everything else – every other variable – constant in order to be able to say that any change made to the first variable has caused any measured change in the second.

But in social science we can’t be sure that x has caused y, because the thing we are interested in measuring (e.g. reading age) has an annoying habit of changing anyway, irrespective of what is being done to it. This is why, in social scientific experimental research, we have to do something such as bringing in an extra group, as alike as possible to the first group, and giving them everything that the first group had so that we can eliminate sources of variation – every source of variation, that is, except the one we are deliberately varying. So, an experiment in a social science situation usually involves two or more groups being treated in exactly the same way, save one, namely the manipulation of the first variable. Any difference that then exists between the two groups after the manipulation is made to happen by the experimenter is taken to exist because of the experimental treatment. This classic form of the experiment is usually shown as in Table 6.2.

A lot can go wrong in all of this (see the discussion of experimental validity on p. 107), not least due to the effect of the myriad confounding variables that might affect

### Table 6.2  The classic experiment

<table>
<thead>
<tr>
<th></th>
<th>Pre-test</th>
<th>Treatment</th>
<th>Post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment group</td>
<td>Take first measure</td>
<td>√ give treatment</td>
<td>Take second measure</td>
</tr>
<tr>
<td>Control group</td>
<td>Take first measure</td>
<td>× no treatment</td>
<td>Take second measure</td>
</tr>
</tbody>
</table>

A control group serves as a basis of comparison in a piece of research. The control group receives no experimental manipulation.
the result of an experiment. These confounding variables include the galaxy of minor and major things of everyday life that cannot be factored into any social situation: power failures, the colour of the paint on the walls, the morning’s argument with the spouse, and so on. The philosopher Alasdair MacIntyre (1985: 99), in his devastating critique of social science, puts it this way: ‘There is at the outset no determinate, enumerable set of factors, the totality of which comprise the situation.’ He suggests that social science will always either be ignorant of the potential effects of the indeterminate set of factors or will fail to notice the most important of them. He gives the example of the potential importance of the length of someone’s nose. He noted that an historian had suggested ...

... that the cause of the foundation of the Roman Empire was the length of Cleopatra’s nose: had her features not been perfectly proportioned, Mark Antony would not have been entranced; had he not been entranced he would not have allied himself with Egypt against Octavian; had he not made that alliance, the battle of Actium would not have been fought – and so on. (MacIntyre, 1985: 99)

But keeping these caveats about validity in mind, what are the practicalities?

The way that an experimenter states the expected outcome of manipulating the first variable is via a hypothesis. An example of a hypothesis is: ‘Raising children’s self-esteem in Year 5 will raise their attainment in English’. You can measure self-esteem with tests, and you can measure attainment in English with tests, and when you have completed your experimental study you can see whether your hypothesis is true. (You may come across the term ‘null hypothesis’, particularly in psychological studies, but I feel there is no need to complicate things with this, which seems to me to be an example of social scientists making things more complicated than necessary.)

It’s important to understand some jargon about experiment, and I’ve been trying my hardest to avoid it until now, but you will need it if you decide to do an experiment and read any further about this design frame. This is given in Table 6.3.

Figure 6.6 is a greatly simplified representation of the process of an experiment, and experimenters have to set up experiments very carefully (ensuring, for example, that the groups are as alike as possible) to maximise the probability that the conclusions they draw about x causing y are correct. There are many opportunities for going wrong in all of this – in the setting up of an experiment or in interpreting its conclusions statistically. Going wrong can mean that you may assume your hypothesis is correct when it is not, or that you assume your hypothesis is false when in fact it is correct. These errors are usually summed up in some more jargon: Type I and II errors: