STEPHEN GORARD

Research Design

Creating Robust Approaches for the Social Sciences

SAGE

Los Angeles | London | New Delhi
Singapore | Washington DC

Process evaluation – usually in-depth field work conducted as part of a trial to check how well the intervention is implemented (fidelity to treatment) and how participants react to it.

Propensity score – used for matching control cases in a natural experiment, based on identifying characteristics most likely to be linked to participation in the intervention arm.

Randomised controlled trial or RCT – evaluation of an intervention which is manipulated so that at least one randomly allocated sub-group receives the treatment and at least one does not.

Regression discontinuity – method for generating an unbiased comparator group in a natural experiment by looking at cases either side of an important threshold.

Researcher effect – unconscious impact the researcher can have on the nature of the findings, including demand characteristics in interviews, and differential awareness of some findings.

Sample – cases involved in a study, deliberately selected to represent a wider population.

Synthesis – process of summarising existing evidence on a topic, including systematic review, secondary analysis, and meta-analysis.

Warrant – the logical argument leading from research findings to the conclusions drawn.

1

Introduction – 'Design' as Distinct from Methods

ONE

What is research design?

SUMMARY

- Attention to research design at an early stage is necessary for rigorous social research.
- Many areas of social science do not pay enough attention to design.
- Many existing resources for new researchers over-emphasise research methods at the expense of design.
- The elements of a research design include the cases studied, their allocation to sub-groups, the timing and sequence of data collection, and any interventions.
- These elements can be portrayed in a simple design notation to represent new or existing designs.
- An important part of research design involves thinking beforehand about the kinds of conclusions that you might want to draw.

1.1 Introduction to design

This is a book about research design for social scientists. It argues that research design has been largely ignored in the development of new researchers, at the expense of a focus on methods of data collection and analysis. Perhaps this is because so many people generally care so little about their social science. To understand what I mean by this, consider areas of innovation where research design is strong. These might include the development of transport such as cars or elevators, of consumables such as medicines or packaged foods, and of gadgets from toasters to internet-capable televisions. In all of these areas, and many more, all of the products are tested before use. In many countries it is illegal to market such a product without rigorous testing. Even if it were not illegal, a strong pressure to test all products would come from the consumer. People want their aircraft to fly rather than crash, and their medicines to work rather than to poison them inadvertently. So, the research to test these things must be designed with a robust approach. Of course medicines and aircraft still fail, despite testing. This regrettable fact is not an argument against testing; it is an argument for more and better designed testing.

People should also care about the quality of studies in economics – witness the worldwide economic downturn in 2007/08 that was almost entirely un-predicted by the thousands of professional economic researchers in each country. The public should care about the billions of public money spent on school 'improvement' schemes that have no discernible impacts on the desired outcomes. Similar concerns should arise in research relevant to housing, crime, social services, business leadership, politics, international development, well-being, social inequalities, marketing, and a host of other fields. Perhaps people do not care as much as they might because, even where research in social science has serious public implications, the 'result' could be a long way off in the future, or hard to discern in the present. People rarely fall ill or die as a direct result of poor social science research. Now, this should not mean that they do not want improvements in public services like housing, education, or criminal justice. But perhaps their concern is less immediate than the fear that a badly designed plane might crash, because the consequences of poor design in social science could be less visually dramatic.

Two other reasons may be that social science research is often ignored by its potential users such as politicians, and practitioners in the public services, and that its research findings are often of very poor quality anyway. None of these reasons is an excuse, but in combination they might form an explanation for how and why social science research gets away with ignoring research design. What this book does is to imagine that more people genuinely care about the quality of social science research, in the way that they care about the effectiveness and safety of aircraft and medicines. The book imagines that when a child is taken into care, or a government changes the sentencing guidelines for criminal courts, then the public would demand that these decisions are made using the best possible evidence.

Design is not chiefly about techniques or procedures. It is more about care and attention to detail, motivated by a passion for the safety of our research-based conclusions. At its simplest, research design is about convincing a wider audience of sceptical people that the conclusions of the research underlying important decisions are as safe as possible. This is perhaps the major difference between the objects of design in medicine and engineering, where things can be seen to work or fail quickly, and in most social sciences, where we can only seek to be convincing. If something works, that is convincing in itself, but where we do not know whether something works, we can at least demand to be convinced that it *ought* to work. We should want to be convinced that it is worth risking the happiness of a family by removing a child from its parents, risking public safety by releasing prisoners early, or spending public money on almost any intervention. Such decisions might be correct, or they might be a wasted opportunity or worse. It is

the task of social scientists to help make such decisions as foolproof as possible. At present, despite a small amount of excellent work in every field, this is just not happening sufficiently.

New researchers largely complete their development lacking any understanding of research designs, and this is reflected in the inadequate work of many areas of public policy research. There are many examples of public policy interventions, some covered in this book, that have been well-intended and rolled out into practice on the basis that they seem plausible and unlikely to do any harm. Yet when they have been rigorously evaluated, they have been found to be ineffective or even harmful. This means that ineffective and even harmful initiatives can divert scarce resources away from effective ones – a particular problem in the current economic downturn, when decisions are being made to abandon programmes on a whim rather than in terms of genuine cost-effectiveness. So, policy-makers and public auditors are increasingly calling for good research evidence on the development of cost-effective and efficient policy and practice solutions, establishing causal-type relationships between innovative changes and their desired effects. This is a key ethical issue for publicly-funded research.

In an attempt to improve the situation, this book is for a range of audiences. These suggested readers include newer researchers in those areas of social science where design is already important – including health promotion studies, for example. For them, the purpose of the book is to provide a relatively gentle introduction that can lead to more advanced templates for rigorous research design. The book is also for newer researchers in areas where research design is present only in a limited fashion. It should encourage them to go beyond focusing almost exclusively on longitudinal designs in sociology, or merely laboratory experiments in psychology. For them, the purpose of the book is to set the common design(s) within their disciplines into a wider context, and to suggest that a mature social science requires a greater variety of designs. Perhaps, most urgently, this book is for newer researchers in those many areas of social science where design is almost completely absent, where methods resources do not even address design, or it is confused with instrument design, *post hoc* statistical procedures, or bizarre issues like 'paradigm wars' (Gorard 2004a). This is probably the situation in most fields, including economics – the supposed 'queen' of the social sciences.

This is most definitely a book for readers who do not know what research design is, did not take a course on it as a doctoral researcher, who would otherwise feel content to continue with their existing approach to generating evidence for public consumption, and whose mentors, supervisors and colleagues feel the same. As this book argues, such complacency is unethical and unwarranted. In the example areas listed so far there are key issues of safety, efficiency and equality. People have lost their jobs as a result of an economic downturn caused partly by untested financial products, for example. The public should care about such things, but the researchers who work in such areas often claim to care about them even more. If they do care, they will want to ensure that they design their research to be as rigorous as possible. Ignoring design is one way of saying openly to the world – 'I don't care about the quality of my research, the wasted opportunities it represents, the waste of peoples' time participating in or reading it, or the dangers to the very people that the research is meant to help'.

1.2 Design and methods

An important point for readers to understand is that research design is not about methods of data collection and analysis. What all rigorous research designs, and variants of them, have in common is that they do not specify the kind of data to be used or collected. No kinds of data, and no particular philosophical predicates, are entailed by common existing design structures such as longitudinal, case study, randomised controlled trial or action research. A good intervention study, for example, could and should use a variety of data collection techniques to understand whether something works, how to improve it, or why it does not work. Case studies involve immersion in one real-life scenario, collecting data of any kind ranging from existing records to *ad hoc* observations. The infamous 'Q'-words of qualitative and quantitative, and mixed methods approaches are therefore not kinds of research design; nor do they entail or privilege a particular design. Of course, all stages in research can be said to involve elements of 'design'. The design of instruments for data collection is one example. But research design, as usually defined in social science research, and as discussed throughout this book, is a prior stage to each of these. Thinking about methods before design is similar to an architect ordering building materials or setting deadlines for construction before settling on what the building will be (de Vaus 2001).

This point is quite commonly confused in the literature, where randomised controlled trial designs are seen as tied to 'quantitative' methods of data collection and analysis (Ceglowski et al. 2011), or it is assumed that a life-course research design must be 'qualitative' (Fehring and Bessant 2009). This point is also confused in some research methods resources, even those purportedly about design, including Creswell and Plano Clark (2007) who are really writing about methods issues not about research design. These writers and many like them contribute to the widespread misunderstanding of design issues. Do not be misled. Otherwise, judgement about what should be a design issue, such as how well the research will cater for rival explanations of the evidence, will be confused with judgement about the perceived merits of a method, such as whether to use a survey or interviews.

A study that followed infants from birth to adolescence, weighing them on 1 January every year, would be longitudinal in design. A study that followed infants from birth to adolescence, interviewing their parents about their happiness every year, would also be longitudinal. A study that did both of these would still be longitudinal, even though some commentators would distractingly and pointlessly categorise the first study as 'quantitative', the second as 'qualitative', and the third as 'mixed methods'. In each example, the design – 'longitudinal', or collecting data from the same cases repeatedly over a period of time – is the same. This illustrates that the design of a study does not entail a specific form of data to be collected, nor does it entail any specific method of analysis; nor does any method require a specific research design.

Almost all existing research resources for newer researchers concern methods of data collection and analysis, and almost all of the rest concern red herrings about paradigms, or treating serious subjects like epistemology as though they were fashion items to be tried on and rejected on a whim. This is true even of many texts that claim to be about research design. This book is very different. Methods of investigation and the philosophy of social science are important, and aspects of both appear throughout the book. But they are not its starting point or its focus.

1.3 The elements of design

The elements of design covered in this book include the cases (participants) involved, the ways in which cases can be allocated to sub-groups, the time sequence of data collection episodes, and any manipulated interventions. These elements are the same, except perhaps for some terminology, as those presented by de Vaus (2001) and Shadish et al. (2002). The book presents these elements of design using a shorthand notation, as a convenient way of expressing more complex designs, and the differences between them. The notation is very simple, and all designs will also always be fully described and illustrated with examples where they first appear in a chapter. Do not be alarmed. What follows here is a brief introduction to the notation.

In a design, the cases are the participants in a study or the objects of a study. The letters R, C, M and N are used to denote groups of cases, allocated to their groups randomly (R), by using a cut-off point (C), through matching (M) or none of these (N). The letter O is used to represent an episode of data collection, which could be observation, measurements, conversations, text or indeed any form of data. If it is necessary to distinguish two or more different types of data collection, a sub-script will be added to the standard notation O. Thus, O_1 and O_2 might represent two different kinds of data taken from the same cases (such as a standard test and an interview). This vagueness about what the methods of data collection are is deliberate (see above). The letter X is used to represent an intervention or change of some sort that might influence the cases to which it is applied. Again, if it is necessary to distinguish two or more different types of intervention, a sub-script will be added to the standard notation X. Thus, X_1 and X_2 might represent two different kinds of treatments given to the same cases. I also use a square bracket, as in [X], to denote an intervention that occurs naturally rather than created by the research. Time is represented by a flow of events from left to right, and different groups of cases are denoted by different lines on the page. A simple example could be:

(→Time→)			
N	X	O	(Group 1)
N		O	(Group 2)

This shows a study of an unknown number of cases, sub-divided into two groups (two lines on the page), and divided naturally or non-randomly (N for each). The first group of cases is given a treatment or intervention (X) and the second group is not (blank). Both groups then have the same unspecified data collected from them (each have an O without any sub-scripts). The diagram shows that the data collection (O) occurs after the intervention (X), and the intervention occurs after the allocation to groups (N) because of their order in the line representing time from left to right. There are varieties of design notations, and more complex issues involved, but this shorthand will do for the present. It will enable me to present the designs in this book as an easy picture, once you get the hang of the notation, and should allow you to make notes on any research you are reading and to communicate designs to colleagues.

1.4 The structure of this book

Research design in the social sciences is a way of organising a research project or programme from its inception in order to maximise the likelihood of generating evidence that provides a convincing answer to the research questions for a given level of resource. Chapter Two presents a simplified cycle for a field or programme of research and how this relates to the elements of design. The next section of the book looks at the rationale for research designs. It provides grounds for deciding on which design is most appropriate for a given study. Chapter Three looks at research questions, how we might generate them, and best express them in order to achieve useful and meaningful answers. Chapter Four introduces the idea of a warrant for research claims, as the part of an argument that could convince a sceptical person to believe the answers to the research questions. Chapter Five is all about the nature of causal claims, which have a special place in explanatory social science research.

The third section of the book concerns the various elements of a design. There are many elements to consider in a research design, but they commonly include the selection of cases of interest to be used in the research (Chapter Six), the appropriate allocation of cases to sub-groups and their subsequent comparison (Chapter Seven), what happens over time (Chapter Eight), and any intervention to be evaluated (Chapter Nine). A specific design or project may have only some of these elements, but some well-known designs involve all of them. These elements of a research design can be combined and varied in many ways, so that each new project might devise a completely new kind of design. On the other hand, there is a variety of standard designs that it is worth being familiar with, both to assist when reading the research of others, and to give some idea of the range available for your own research.

The next section moves to slightly more advanced issues relating to design. A range of further and currently less common research designs is presented in Chapter Ten. Chapter Eleven discusses traditional and generic threats to the validity of research conclusions, and introduces some important new ones. The key issue of how to differentiate between patterns or simply 'noise' in the data is addressed in Chapter Twelve. Chapter Thirteen looks at the ethics of research design, and conflicts of interest in the conduct of research.

Finally, Chapter Fourteen sums up the argument that a robust approach to social science research design is necessary, and offers a few guidelines for choosing a design and developing a grant application, using the principles and ideas in this book.

Each chapter also ends with three 'exercises' that readers might like to consider while reading. These exercises will tend to get more complex and involve greater judgement as the book progresses. They are followed by my notes and suggestions for discussion, which are an important part of the argument and narrative of the book. They often introduce material in a different way, or even suggest ideas not covered elsewhere, and so should be treated as an integral part of the text. Each chapter ends with a suggestion for further reading on the same topic.

Initial exercises on research design

1 Using the simple design notation described in this chapter, a piece of research might be presented as:

N	O	X	O	(Group 1)
N			O	(Group 2)

Assume that this design represents an evaluation of a new training course for social workers. The intervention (X) is the training course. In one region, a group of volunteers (Group 1) take the training course, and their remaining colleagues (Group 2) do not. The volunteer group are initially given a test of the skills that the training is intended to improve. After training, the volunteers and their colleagues (Group 2) are given the same test.

a How many groups of cases are there in this research design?
b How have the cases been allocated to groups – and have they been allocated by chance or not?

c How can we tell from the design notation that Group 1 was given a test before the training?

d If the volunteers score better on the test after the training than they did before, suggest a few reasons why this is not necessarily evidence that the training is the cause of the improvement.

e If the volunteers also score better on the final test than their colleagues, suggest a few further reasons why this is not necessarily evidence that the training is the cause of the difference between Groups 1 and 2.

f Finally, if the colleagues do as well as the volunteers in the skills test, suggest at least one reason why this is not necessarily evidence that the training is ineffective.

2 Imagine designing a new piece of research that tries to follow a group of all men leaving a specific prison after their custodial sentence, in a specific month. The researcher will interview each person once as soon as possible after they leave prison, and then monitor them a year later to see if they have a job, have re-offended, and so on. What would the simplest version of this research look like in design notation?

3 Select a journal article reporting new research in your own area of interest. Try to present the design of this research using the simple notation introduced in this chapter.

Notes on initial exercises

1 a There are two groups in this design because there are two lines of notation, with each line representing the research process as experienced by one group.

b The cases have not been allocated to groups by chance. We know this because each line begins with N, denoting a non-random division between them. In the example, the first group consists of volunteers, and the second of everyone else in the study.

c We know from this design that Group 1 is given the skills test before the training, because the first episode of data collection (O) appears to the left of the treatment (X). Time is assumed to flow from left to right.

d The volunteers might score better on the test after the training than they did before simply by chance, especially if the difference is small. They might also do better through practice, because they have already taken the same test before the training. Or something else might have happened between the two tests, such as formative experiences at work. Any of these explanations and a host of others could show that the difference between the before-and-after tests is not related to the training. This is why it is important to have a comparable group that are also tested but do not receive the training.

e The volunteers might score better on the test than their colleagues, by chance or as a result of practice, or due to some other experience that is unique to the volunteers. But the simplest explanation could be that the two groups are clearly not comparable. By volunteering, the group that receives the training has perhaps shown itself to be more enthusiastic, better motivated and keener to improve their skills than the other group. They might therefore have performed better in the test than their colleagues even without the training. We do not know from this design.

f It is possible that the training is effective, even if the colleagues do as well as the volunteers in the skills test. Again, there are many possible reasons for this, including chance, or that the effect of the training is too small to be detected, or even that the volunteers were those who felt most in need of training, having a lower level of skill initially. Another quite common problem is termed 'contamination'. Where volunteers who receive the training and other colleagues work together in the same offices or departments, the colleagues may learn about the training second-hand via inadvertent cascading in conversation. The volunteers might show their friends materials from the training course to help them as well. So, the training could be effective but the results not show up as a difference in the scores because, in reality, the training has affected both groups. These ideas begin to give some idea of the complexities of design and the difficulties of designing a study whose results will convince a sceptical audience.

2 The simplest notation that matches the design for the prison leaver study could be:

N O_1 O_2

There is only one group (one line). The group is a natural cluster in one prison (not random), and there are two different episodes of data collection (not a repetition of the same data collection, as in question 1). There is no intervention – the prisoners have already left prison at the outset.

3 A surprising number of articles report research without specifying a design. In many cases this is because so few of the elements of a design are included in the study that it is not worth discussing. For example, a simple survey of business leaders might have this design:

N O

It does not matter how complex the subsequent analysis is, nor how sophisticated the questionnaire is. There is only one designed group. This is so even if the analysis later divides the cases temporarily into sub-groups like male/female or by the size of their businesses. As there is only one group, there is no pre-specified method of allocating cases to groups. There is no intervention and no time sequence. It is a snapshot study. The same design notation would be used if individual interviews replaced the survey, because the design is independent of the precise methods of data collection. And the same notation would be used if a series of focus groups replaced individual interviews. The 'groups' in a research design, represented by different lines in the notation, are those for whom the research process is different. If research involves six focus groups all doing the same thing, there is only one 'group' for design purposes. The same applies to 100 interviewees, or 1,000 survey respondents.

Suggested reading

Chapters One and Two in de Vaus (2001) *Research Design in Social Research*. London: SAGE.

TWO

Introducing designs in the cycle of research

SUMMARY

- Different types of research have different purposes, from synthesising what we already know about a topic, through developing a new theory, to monitoring how well a new approach works in practice.
- Different research designs will be particularly suitable for these different types of research. No one design is always applicable or best suited for an entire programme of study.
- Designs can be classified in a number of ways, such as whether they involve a pre-specified comparison or not, whether there is a sequence of data collection episodes, and whether they involve a deliberate intervention or not.
- Research findings are intrinsically more convincing when the research design uses the 'better' or positive half of such classifications. For example, a comparative claim is intrinsically more convincing with a pre-specified set of comparator sub-groups.
- Such consideration reveals that, all other things being equal, something like a case study will always tend to be the least convincing design, while something like a randomised controlled trial will tend to be the most convincing.

2.1 Introduction

This chapter starts by looking at the concept of an over-arching research cycle, and suggests that different designs might be more or less appropriate at different phases in that cycle. Each design covered in this book, and many others, has a place in social science. This section tries to show how they fit together in a programme working towards the solution of a social science puzzle.

2.2 Research cycle

Whatever area of social science you work in, it is likely that the field as a whole will look something like Figure 2.1. This is a simplified description of a full cycle for a research programme. It is based on a number of sources, including the genesis of a design study (Gorard with Taylor 2004), the UK Medical Research Council model for undertaking complex medical interventions (MRC 2000) and one OECD conception of what useful policy research looks like (Cook and Gorard 2007). The cycle is more properly a spiral which has no clear beginning or end, in which phases overlap, can take place simultaneously, and could iterate almost endlessly.

The various phases illustrated should be recognisable to anyone working in areas of applied social science, like public policy. I discuss each of these steps in slightly more detail below, and then in succeeding chapters. Any one study or researcher might contribute to only some of these steps, but the field as a whole ought to progress in something like this fashion. Unfortunately, some fields seem

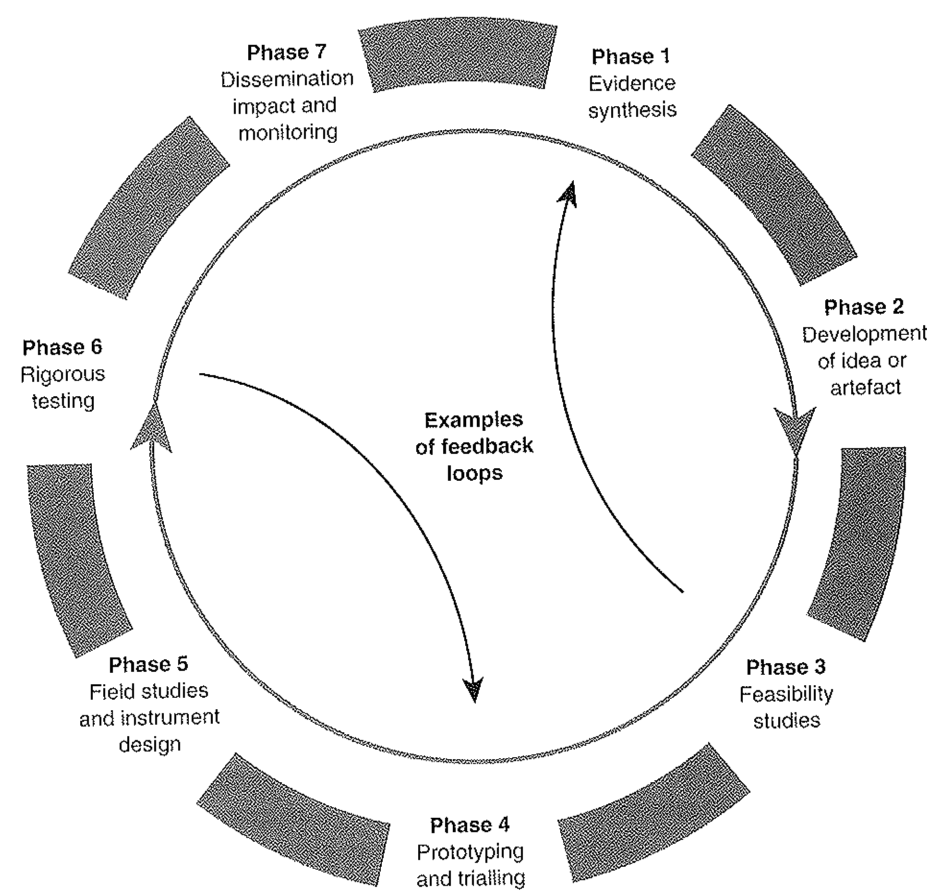

Figure 2.1 An outline of the full cycle of social science research and development

to have got stuck in a limited range of phases (Gorard et al. 2011). Of course not all ideas come to fruition. There should be questions that are answered by a synthesis alone. Then some ideas should halt at a pilot stage where they show no promise. But equally, this should not always happen. Some ideas ought to be developed all of the way to become tested 'products' engineered into use for other researchers, policy-makers or practitioners. These products could be theories, practical protocols, policy interventions or genuine artefacts like software or training manuals. The key point for this book is that different designs will be more appropriate at different phases in the cycle. So a healthy field, as a whole, will use a wide range of research designs – which means of course that everyone needs to know something about all designs, even if only to be able to conduct an appropriately critical literature review in their field.

Evidence synthesis

The cycle might start with an evidence synthesis, which should use existing datasets and previously published literature in an unbiased way to produce a summary of what is already known in relation to the research question(s). The cycle might also end there, if existing evidence answers the question satisfactorily. In practice it tends not to, partly because the external funding structure for research discourages such honesty (Chapter Thirteen), and partly because researchers are generally so poor at conducting a synthesis of existing evidence. The characteristics of a good synthesis are discussed in Chapter Three. Once the existing evidence synthesis is complete, it should be clear what remains unanswered in the area of interest, and this can lead to a definition of the problem to be solved, the research purpose, and research questions.

Development and preparation

A high proportion of existing published work in social sciences seems to be in Phases 2 to 3 of the cycle. This is what may be envisaged as the research development phase. However, very little of this work currently starts with a serious attempt at research synthesis, offering instead only partial literature reviews and confusing conceptual frameworks. Even less of it moves on from this development work towards the preparation of a definitive large-scale study, or to creating something useful from the knowledge gained. Some of the reasons why this may be so were touched on in Chapter One.

Proto-typing and trialling of any new idea is vital, if resource is not to be wasted on a large-scale definitive study that has no chance of success. Trying ideas out at this development stage will tend to be done via small scale work, to minimise the risk and cost in case the idea does not work. With minimal risk and cost, several alternative ideas can be tried out in parallel. Feasibility studies can be as cheap as thought experiments (Chapter Nine), as simple as case studies, or they can be complex designs for multi-method data collection. Their purpose in the cycle is to assess the likelihood of success of the idea, and so to assist with the decision whether to proceed further in the cycle, or not. The decision can also be influenced by the potential gain from the idea, the resources required, and ethical issues.

Evaluation

Each prior phase might lead to a realisation that little more can be learnt and that the study is over, or that the programme needs radical revision and iteration to an earlier phase, or progression to a subsequent phase. However, the overall cycle can be envisaged as tending towards an artefact or 'product' of some kind. This product might be a theory (if the desired outcome is simply knowledge), a proposed improvement for public policy, or a tool/resource for a practitioner. In order for any of these outcomes to be promoted and disseminated in an ethical manner they must have been tested. A theory, by definition, will generate testable propositions. A proposed public policy intervention can be tested realistically and then monitored *in situ* for the predicted benefits, and for any unwanted and undesirable side effects. It is no good knowing that an intervention works if we do not also know that it is unpopular and likely to be ignored or subverted in practice. Similarly, it would be a waste of resource, and therefore unethical, simply to discover that an intervention did not work in Phase 6 and so return to a new programme of study in Phase 1. We would want to know why it did not work, or perhaps how to improve it, and whether it was effective for some regular pattern of cases but not for others. So in Phase 6, like Phase 1, the researcher or team who genuinely wants to find something out will naturally use a range of methods and approaches including measurement, narrative and observation. Methods really are independent of design.

2.3 A simple design typology

The basic elements of research design introduced in Chapter One can be combined to form a wide variety of study designs. Most of these combinations would not have a well-known name but are as valid as any other combination nevertheless. The value of a design can only be judged in relation to the research questions it is intended to answer (Chapter Three). This in turn depends upon where the research programme is currently focused in terms of Figure 2.1. Some combinations of design elements have well-known names like 'case study' or 'longitudinal'.

One way of classifying such standard designs is in terms of whether they are active or passive. An active design would include a controlled intervention, introduced as part of the study. Examples of active designs that you may already have heard of include randomised controlled trials (RCTs), and laboratory experiments. Quasi-experiments, action research, interrupted time series, regression discontinuity, and design studies may also involve a specific intervention. All of these are intrinsically more convincing in testing a causal claim than completely passive designs (Chapter Nine). A passive design may consider changes over time but these changes do not occur as part of the research itself, because there is no specific or controlled intervention. Examples include standard cohort research, other longitudinal designs, case studies, and comparative or cross-sectional approaches.

Another way of classifying designs could be in terms of whether they involve a pre-specified comparator group, or not (Chapter Seven). Obviously, comparative studies are intrinsically more convincing in testing a comparative claim than non-comparative ones. Examples include RCTs, natural experiments, and comparative research. Non-comparative designs include standard longitudinal approaches and case studies.

A third way of classifying designs could be in terms of whether they involve repeated measures or some other planned elapse of time between the start and a final measure, or not. Such longitudinal designs are intrinsically more convincing in demonstrating a before-and-after claim than cross-sectional ones (Chapter Eight). Examples of designs with an automatic longitudinal element include standard cohort research, other longitudinal designs, RCTs, and natural experiments. Those usually without a longitudinal element include case studies, and comparative and cross-sectional approaches.

It is interesting that only RCTs and other experiments appear in the 'better' or positive half of each of these classifications. They are better for causal, comparative *and* time-dependent claims. In addition, RCTs have the advantage over quasi-experiments of having cases allocated to comparator groups at random. As shown in the rest of the book this means that given the right conditions and questions, an RCT or equivalent laboratory experiment is the best and most convincing design to use. Much of the rest of research, as in the full cycle, can be envisaged as working towards such a trial.

It is also notable that case studies are always in the worse half of each of these classifications. In themselves, they have no comparator, no intervention and no longitudinal element. As a design they are simply an episode of (possibly extended) data collection:

N   O

This notation represents observations (O) or data collected at one point in time with no intervention, no pre-defined comparator groups, and so no rule about how cases were allocated to comparator groups. With so few of the elements of research design present here, there is little that such research can do beyond exploratory initial descriptive preparation for subsequent studies. Such work might be useful in generating ideas and possible explanations for a causal model perhaps. However, the authors of such work rarely seem to generate such ideas. Even where the authors of such work describe it as exploratory, they do not then progress with it to a later phase of the cycle. Instead they seem merely to move to another exploratory study (see Gorard et al. 2011 for examples).

A case study, in isolation, will never be the preferred design for any study that aims to be convincing or definitive. Case studies can be valuable, especially towards the start of the research cycle, largely because they are simple and quick to set up. I realise that some commentators would say that case studies are to be preferred because they allow a researcher to study a case in-depth. But you must recall the point from Chapter One. Designs are independent of the methods of data collection. It is as feasible that a case study was an examination of the financial accounts of one company as that a longitudinal study involved an in-depth observation of someone's adjustment to a new job over the first six months. No design has a monopoly on depth or breadth of data.

The need for warranted conclusions requires the researcher to identify the kind of claims to be made – such as descriptive, associative, or causal – and then ensure that the most appropriate possible design is used. A comparative claim *must* have an explicit and suitable comparator, although it is truly shocking how often this is not found. The warranting principle in Chapter Four is based on this consideration – if the claim to be drawn from the evidence is not actually true then how else could the evidence be explained? The research-based claim should be the simplest explanation for the available evidence. What the research design should do is eliminate (or at least test or allow for) the greatest possible number of alternative explanations before the final claim to knowledge is made. In this way, the design eases the analysis process, and provides part of the warrant for the research claim. Design makes research better.

There are a host of already known designs other than those mentioned so far, and presumably many more waiting to be combined from the elements of research. Which design is used for any study should depend largely on the kind of claims and conclusions to be drawn. And these in turn depend on the research questions to be answered, which are the subject of the next chapter.

Exercises on designs

1 Two doctoral researchers at a social science conference describe their research projects. The first researcher is a social historian, looking at the impact on the diet of agricultural labourers of the enclosure of common land in England. The researcher examines records and other sources relating to meals for the labourers in one area, from around 25 years before the recorded start of land enclosure, and then from around 50 years later. The second is an education researcher who has persuaded their own institution to schedule their adult education classes more flexibly, in order to encourage a wider range of participants. This researcher has conducted a survey of people taking adult education classes in the year before the change took place, and in the year after. The survey asked people about their occupational and educational backgrounds.

a What is the simplest design notation for each study?
b How would the design look different if the second researcher had used interviews with the adult learners, rather than a survey?
c Both researchers want to argue that there was a change from the first to the second episode of data collection, and that this change was due to enclosure in the first study and the rescheduling of classes in the second. What is the biggest problem for their argument?
d How would you advise them to re-design each study to try and overcome this problem?

2 A third researcher at the same conference has used quite a complex research design, involving multiple groups and episodes of data collection. A member of the audience asks what the design is called, and then mocks the presenter for not knowing. Is it reasonable for a researcher to describe a design but not know what it is called?

3 Why might an intervention that was found to be reasonably effective in Phase 6 of a research cycle turn out to be much less effective when rolled out into widespread practice? Come up with a range of suggestions, perhaps through discussion in a group.

Notes on exercises on designs

1 a The simplest design notation for the first study could be:

N O

N [X] O

Different lines are used here because the cases are different before and after the intervention of interest. Of course, the first researcher did not really intervene to create the mediaeval land enclosure system in England (how could they?). This is why I put the X in square brackets. But this limitation is the same for all history, retrospective accounts, archaeology, palaeontology and so on, and so I think it is reasonable to represent the design like this. The first researcher is looking at evidence from one time period and then looking at evidence from a later period, with an important intervening change. In some respects it does not matter whether this change has already happened or not, or quite how long ago it happened.

The simplest design notation for the second researcher could also be:

N O

N [X] O

They have gathered evidence from one cohort of adult learners, waited until their institution altered something important, and then gathered evidence from a later cohort. In essence, the design is very similar to the first despite the surface differences in the topics.

b If the second researcher used interviews rather than a survey, this would make no difference to the design. Data collection is only about what goes on within each O episode, and this is traditionally unspecified at the level of research design.

c Both researchers face the same problem. They cannot tell whether any difference between the two time periods is due to the intervening specified change; whether it would have happened anyway; is due to differences between the people; or whether it is just natural variation caused by incompleteness of records or the vagaries of data collection. Their argument would be unconvincing, as it stands.

d Probably the simplest thing both researchers could do is to add a relevant comparison group to their design, to combine with the before-and-after element. The second researcher could look at the before-and-after participation rates in another institution that did not reschedule classes, or they could have rescheduled only some of their classes in the first year to see what happened. Therefore, the design could be:

```
N    O
N          [X]    O
N    O
N                 O
```

This design is discussed further as a difference-in-difference, in Chapter Ten. It is harder for the first researcher, but they could pick any of a variety of slightly weaker comparisons. For example, they could have looked at diets 50 years before the start of the study, or 50 years after it, and so tried to judge if the change specifically during the onset of land enclosure was remarkable in any way. Therefore, the design could be:

```
N    O
N         O
N              [X]    O
```

As you see, passive designs for causal questions can get complicated very quickly. This kind of design is discussed further, as an interrupted time series, in Chapter Ten.

2 It might be a little embarrassing for the third researcher if they were using what was clearly a simple and well-known design like a cohort study but had not learnt the name. Names are a useful shorthand as long as both the researcher and the audience mean the same thing by it. Not knowing the name would not affect the validity of the research of course. However, the implication here is that the design is not well-known. Perhaps it is some combination of two or more off-the-shelf designs. In this case it does not matter at all if it has no name. Even if the researcher gave it a name this would be useless for the presentation because it is then almost certain that the audience would not know what the name meant. Given that the design will have to be explained anyway the name might just be confusing. In general, knowing the names of designs is over-rated. Where the design is simple, such as a case study, there is often considerable confusion and disagreement between experts about what it means. It is therefore ambiguous and needs spelling out. Where the design is unfamiliar and complicated then the name will not lead to recognition and could be ambiguous. So, it will still need spelling out. Design notation, or similar, is more likely to clarify the design, and is safer than just using a name.

3 There are many reasons why an intervention that was found to be effective in Phase 6 of a research cycle might turn out to be much less effective when rolled out into widespread practice. The results of the evaluation might have been faked or inadvertently misunderstood by the researchers involved (Chapter Four), perhaps because of a conflict of interest (Chapter Thirteen). Did you consider that? The intervention might have been better resourced as an experiment than when it was rolled out. This often happens. The evaluation might have involved only volunteers, whereas the rollout might be imposed on all, leading to sullen participation. The context for the evaluation might have been more suitable for success than the greater variety of settings encountered on rollout. The general training to use the intervention might be worse than it was during the experimental phase. And so on.

Suggested reading

Gorard, S. (2010) 'Research design, as independent of methods'. In C. Teddlie and A. Tashakkori (eds) *Handbook of Mixed Methods*. Los Angeles: SAGE.