CHAPTER GUIDE

This chapter is concerned with the characteristics of quantitative research, an approach that has been the dominant strategy for conducting social research. Its influence has waned slightly since the mid-1970s, when qualitative research became increasingly influential. However, it continues to exert a powerful influence in many quarters. The emphasis in this chapter is very much on what quantitative research typically entails; though at a later point in the chapter, the ways in which there are frequently departures from this ideal type are outlined. This chapter explores:

- the main steps of quantitative research that are presented as a linear succession of stages;
- the importance of concepts in quantitative research and the ways in which measures may be devised for concepts; this discussion includes a discussion of the important idea of an indicator, which is devised as a way of measuring a concept for which there is no direct measure;
- the procedures for checking the reliability and validity of the measurement process;
- the main preoccupations of quantitative research, which are described in terms of four features: measurement; causality; generalization; and replication;
- some criticisms that are frequently levelled at quantitative research.

Introduction

In Chapter 1, quantitative research was outlined as a distinctive research strategy. In very broad terms, it was described as entailing the collection of numerical data and as exhibiting a view of the relationship between theory and research as deductive, a predilection for a natural science approach (and of positivism in particular), and as having an objective conception of social reality. A number of other features of quantitative research were outlined, but in this chapter, we will be examining the strategy in much more detail.

It should be abundantly clear by now that the description of the research strategy as 'quantitative research' should not be taken to mean that quantification of aspects of social life is all that distinguishes it from a qualitative research strategy. The very fact that it has a distinctive epistemological and ontological position suggests that there is a good deal more to it than the mere presence of numbers. In this chapter, the main steps in quantitative research will be outlined. We will also examine some of the principal preoccupations of the strategy and how certain issues of concern among practitioners are addressed, like the concerns about measurement validity.

The main steps in quantitative research

Figure 3.1 outlines the main steps in quantitative research. This is very much an ideal-type account of the process; it is probably never or rarely found in this pure form, but it represents a useful starting point for getting to grips with the main ingredients of the approach and the links between them. Research is rarely as linear and as straightforward as the figure implies, but its aim is to do no more than capture the main steps and to provide a rough indication of their interconnections.

Some of the chief steps have been covered in the first two chapters. The fact that we start off with theory signifies that a broadly deductive approach to the relationship between theory and research is taken. It is common for outlines of the main steps of quantitative research to suggest that a hypothesis is deduced from the theory and is tested. This notion has been incorporated into Figure 3.1. However, a great deal of quantitative research does not entail the specification of a hypothesis and instead theory acts loosely as a set of concerns in relation to which the social researcher collects data. The specification of hypotheses to be tested is particularly likely to be found in experimental research. Other research designs sometimes entail the testing of hypotheses. In Chapter 1, a cross-sectional design using social survey research instruments that was used as an example (see Box 1.5) involved hypothesis testing. However, as a rule, we tend to find that step 2 is more likely to be found in experimental research.

The next step entails the selection of a research design, a topic that was explored in Chapter 2. As we have seen, the selection of research design has implications for a variety of issues, such as the external validity of findings and researchers' ability to impute causality to their findings. Step 4 entails devising measures of the concepts in which the researcher is interested. This process is often referred to as operationalization, a term that originally derives from physics to refer to the operations by which a concept (such as temperature or velocity) is measured (Bridgman 1927). Aspects of this issue will be explored below in this chapter.

The next two steps entail the selection of a research site or sites and then the selection of subjects/respondents. (Experimental researchers tend to call the people on whom they conduct research 'subjects', whereas social survey researchers typically call them 'respondents.') Thus, in social survey research, an investigator must first be concerned to establish an appropriate setting for his or her research. A number of decisions may be involved. The well-known Affluent Worker research undertaken by Goldthorpe et al. (1968: 2-5) involved two decisions about a research site or setting. First, the researchers needed a community that would be appropriate for the testing of the 'embourgeoisement' thesis (the idea that affluent workers were becoming more middle class in their attitudes and lifestyles). As a result of this consideration, Luton was selected. Secondly, in order to come up with a sample of 'affluent workers' (Step 6), it was decided that people working for three of Luton's leading employers should be interviewed. Moreover, the researchers wanted the firms selected to cover a range of production technologies, because of evidence at that time that technologies had implications for workers' attitudes and behaviour. As a result of these considerations, the three firms were selected. Industrial workers were then sampled, also in terms of selected criteria that were to do with the researchers' interests in embourgeoisement and in the implications of technology for work attitudes and behaviour.

Box 3.1 provides a more recent example of research that involved similar deliberations about selecting research sites and sampling respondents. In experimental research, these two steps are likely to include the assignment of subjects into control and treatment groups.

Step 7 involves the administration of the research instruments. In experimental research, this is likely to entail pre-testing subjects, manipulating the independent variable for the experimental group and post-testing respondents. In cross-sectional research using survey research instruments, it will involve interviewing the sample members by structured interview schedule or distributing a self-completion questionnaire. In research using
Box 3.1 Selecting research sites and sampling respondents: The Social Change and Economic Life Initiative

The Social Change and Economic Life Initiative (SCELI) involved research in six labour markets: Aberdeen, Coventry, Kirkaldy, Northampton, Rochdale, and Swindon. These labour markets were chosen to reflect contrasting patterns of economic change in the early to mid-1980s and in the then recent past. Within each locality, three main surveys were carried out.

- The Work Attitudes/Histories Survey. Across the six localities a random sample of 6,111 individuals was interviewed using a structured interview schedule. Each interview comprised questions about the individual’s work history and about a range of attitudes.
- The Household and Community Survey. A further survey was conducted on roughly one-third of those interviewed for the Work Attitudes/Histories Survey. Respondents and their partners were interviewed by structured interview schedule and each person also completed a self-completion questionnaire. This survey was concerned with such areas as the domestic division of labour, leisure activities, and attitudes to the welfare state.
- The Baseline Employers Survey. Each individual in each locality interviewed for the Work Attitudes/Histories Survey was asked to provide details of his or her employer (if appropriate). A sample of these employers was then interviewed by structured interview schedule. The interview schedules covered such areas as the gender distribution of jobs, the introduction of new technologies, and relationships with trade unions.

The bulk of the results was published in a series of volumes, including Penn et al. (1994) and A. M. Scott (1994). This example shows clearly the ways in which researchers are involved in decisions about selecting both research sites and respondents.

Concepts and their measurement

What is a concept?

Concepts are the building blocks of theory and represent the points around which social research is conducted. Just think of the numerous concepts that have already been mentioned in relation to research examples cited so far in this book:

- structure, agency, social class, job search method, deskilling, occupational satisfaction, religious orthodoxy, religious orientation, preservation of self, informal social control, negotiated order, culture, academic achievement, teacher expectations, charismatic leadership, healthy lifestyle, convention.
- Each represents a label that we give to elements of the social world that seem to have common features and that strike us as significant. As Bulmer succinctly puts it, concepts ‘are categories for the organisation of ideas and observations’ (1984: 43). One item mentioned in Chapter 2 but omitted from the list of concepts above is IQ. It has been omitted because it is not a concept. It is a measure of a concept—namely, intelligence. This is a rare case of a social scientific measure that has become so well known that the measure and the concept are almost synonymous as representations of the centroid of Fahrenheit scales, or as length and the metric scale. The concept of intelligence has arisen as a result of noticing
There are three main reasons for the preoccupation with measurement in quantitative research.

- Measurement allows us to discern fine differences between people in terms of the characteristic in question. This is very useful, since, although we can often distinguish between people in terms of extreme categories, finer distinctions are much more difficult to recognize. We can detect clear variations in levels of job satisfaction — people who love their jobs and people who hate their jobs — but small differences are much more difficult to detect.

- Measurement gives us a consistent device or yardstick for making such distinctions. A measurement device provides a consistent instrument for gauging differences. This consistency relates to two things: our ability to be consistent over time and our ability to be consistent with other researchers.

- In other words, a measure should be something that is influenced neither by the timing of its administration nor by the person who administers it. Obviously, saying that the measure is not influenced by timing is not meant to indicate that measurement readings do not change: they are bound to be influenced by the process of social change. What it means is that the measure should generate consistent results, other than those that occur as a result of natural changes. Whether a measure actually possesses this quality has to do with the issue of reliability, which was introduced in Chapter 2 and which will be examined again below.

- Measurement provides the basis for more precise estimates of the degree of relationship between concepts (for example, through correlation analysis, which will be examined in Chapter 13). Thus, if we measure both job satisfaction and the things with which it might be related, such as stress-related illness, we will be able to produce more precise estimates of how closely they are related than if we had not proceeded in this way.

### Why measure?

There are three main reasons for the preoccupation with measurement in quantitative research.

1. Measurement allows us to delineate fine differences between people in terms of the characteristic in question. This is very useful, since, although we can often distinguish between people in terms of extreme categories, finer distinctions are much more difficult to recognize. We can detect clear variations in levels of job satisfaction — people who love their jobs and people who hate their jobs — but small differences are much more difficult to detect.

2. Measurement gives us a consistent device or yardstick for making such distinctions. A measurement device provides a consistent instrument for gauging differences. This consistency relates to two things: our ability to be consistent over time and our ability to be consistent with other researchers.

3. In other words, a measure should be something that is influenced neither by the timing of its administration nor by the person who administers it. Obviously, saying that the measure is not influenced by timing is not meant to indicate that measurement readings do not change: they are bound to be influenced by the process of social change. What it means is that the measure should generate consistent results, other than those that occur as a result of natural changes. Whether a measure actually possesses this quality has to do with the issue of reliability, which was introduced in Chapter 2 and which will be examined again below.

4. Measurement provides the basis for more precise estimates of the degree of relationship between concepts (for example, through correlation analysis, which will be examined in Chapter 13). Thus, if we measure both job satisfaction and the things with which it might be related, such as stress-related illness, we will be able to produce more precise estimates of how closely they are related than if we had not proceeded in this way.

### Indicators

In order to provide a measure of a concept (often referred to as an operational definition, a term deriving from the idea of operationalization), it is necessary to have an indicator or indicators that will stand for the concept (see Box 3.2). There are a number of ways in which indicators can be devised:

- through a question (or series of questions) that is part of a structured interview schedule or self-completion questionnaire. The question(s) could be concerned with the respondents' report of an attitude (e.g. job satisfaction) or their social situation (e.g. poverty) or a report of their behaviour (e.g. leisure pursuits);

- through the recording of individuals' behaviour using a structured observation schedule (e.g. pupil behaviour in a classroom);

- through official statistics, such as the use of Home Office crime statistics to measure criminal behaviour;

- through an examination of mass media content through content analysis, for example, to determine changes in the salience of an issue, such as AIDS, in the mass media (Beharrell 1993).

Indicators, then, can be derived from a wide variety of different sources and methods. Very often the researcher has to consider whether one indicator of a concept will be sufficient. This consideration is frequently a focus for social survey researchers. Rather than have just a single indicator of a concept, the researcher may feel that it may be preferable to ask a number of questions in the course of a structured interview or a self-completion questionnaire that tap a certain concept (see Boxes 3.3 and 3.5 for examples).
Box 3.4 What is a Likert scale?

The investigation of attitudes is a prominent area in much survey research. One of the most common methods for conducting such an investigation is the Likert scale, named after Rensis Likert who developed the method. The Likert scale is essentially a multiple-indicator or item measure of a set of attributes relating to a particular area. The goal of the Likert scale is to measure intensity of feelings about the area in question. In its most common format, it comprises a series of statements (known as "items") which focus on a certain issue or theme. Each respondent is then asked to indicate his or her level of agreement with the statement. Usually, the format for indicating level of agreement is a 5-point scale going from 'strongly agree' to 'strongly disagree', but 7-point and other formats are used too. There is usually a middle position of 'neither agree nor disagree' or 'undecided', indicating neutrality on the issue. Each respondent's reply on each item is scored and then the scores for each item are aggregated to form an overall score. Normally, since the scale measures intensity, the scoring is carried out so that a high level of intensity of feelings in connection with each indicator receives a high score (for example, on a 5-point scale, a score of 5 for very strong positive feelings about an issue and a score of 1 for very negative feelings). The measure of commitment to work referred to in Box 3.3, some items imply a positive view of work itself, and others a negative one. Thus, in the example in Box 3.3, some items imply a positive view of the phenomenon of interest and others a negative one. This is useful to vary the phrasing so that some items imply a positive view of the phenomenon of interest and others a negative one. This variation is advised in order to identify respondents who exhibit 'response sets' (see the sections on response sets in Chapters 5 and 6).

Box 3.5 A multiple-indicator measure of another concept

In Kelley and De Graaf's (1997) research on religious beliefs, one of the main concepts in which they were interested—national religiosity and family religious orientation—were each measured by a single indicator (see Box 3.5). However, religious orthodoxly was measured by four survey questions, answers to which were aggregated for each respondent to form a 'score' for that person. Answers to each of the four questions were given a score and then aggregated to form a religious belief score. The four questions were as follows:

- Please indicate which statement below comes closest to expressing what you believe about God:
  - I don't believe in God.
  - I don't know whether there is a God and I don't believe there is any way to find out.
  - I believe in a personal God, but I don't believe in a higher power of some kind.
  - I find myself believing in God some of the time, but not at others.

Dimensions of concepts

One elaboration of the general approach to measuring social concepts is to consider the possibility that the concept in which you are interested comprises different dimensions. This view is particularly associated with Lazarsfeld (1958). The idea behind this approach is that, when the researcher is seeking to develop a measure of a concept, the different aspects or components of that concept should be considered. This specification of the dimensions of a concept would be undertaken with reference to theory and research associated with that concept. Examples of this kind of approach can be discerned in Seeman's (1959) delineation of five dimensions of alienation (powerlessness, meaninglessness, normlessness, isolation, and self-estrangement). Bryman and Cramer (2001) demonstrate the operation of this approach with reference to the concept of
Reliability and validity

Although the terms reliability and validity seem to be almost like synonyms, they have quite different meanings in relation to the evaluation of measures of concepts, as was seen in Chapter 2.

Reliability

As Box 3.7 suggests, reliability is fundamentally concerned with issues of consistency of measures. There are at least three different meanings of the term. These are outlined in Box 3.7 and elaborated upon below.

Stability

The most obvious way of testing for the stability of a measure is the test-retest method. This involves administering a test or measure on one occasion and then readministering it to the same sample on another occasion, i.e.: $T_1$, $T_2$, $O_{b1}$, and $O_{b2}$.

We should expect to find a high correlation between $O_{b1}$ and $O_{b2}$. Correlation is a measure of the strength of the relationship between two variables. This topic will be covered in Chapter 11 in the context of a discussion about quantitative data analysis. Let us imagine that we develop a multiple-indicator measure that is supposed to tap a concept we might call ‘designerism’ (a preference for buying goods and especially clothing with ‘designer’ labels). We would ‘administer the measure to a sample of respondents, and then readminister it, there will be little variation in the results obtained.

In other words, the aim would be to establish whether respondents’ answers cannot be relied upon.

However, there are a number of problems with testing internal reliability. One of these is the splitting of the data into halves using the odd-even basis. This can arise in a number of contexts, for example in content analysis where decisions have to be made about how to code categories, or in structured observation where observers have to decide how to classify subjects’ behaviour.

Internal reliability

This meaning of reliability applies to multiple-indicator measures like those examined in Boxes 3.3 and 3.5. When you have a multiple-indicator measure in which each respondent’s answers to each question are aggregated to form an overall score, the possibility is raised that the indicators do not relate to the same thing; in other words, they lack coherence. We need to be sure that all our designerism indicators are related to each other. If they are not, some of the items may actually be unrelated to designerism and therefore indicative of something else.

One way of testing internal reliability is the split-half method. We can take the commitment to work measures developed by Westergaard et al. (1990) in an example (see Box 3.3). The ten indicators would be divided into two halves with five in each group. The indicators would be allocated on a random or an odd-even basis. The degree of correlation between scores on two halves would then be calculated. In other words, the aim would be to establish whether

Box 3.6 Specifying dimensions of a concept: the case of deskilling

This example is taken from a social survey research primarily concerned with social class in Britain by Marshall et al. (1988). The research was based on structured interviews with a national, random sample of individuals. One of the researchers’ areas of interest was Braverman’s (1974) deskilling thesis (see Box 1.2). Based on a reading of the literature on this topic at the time, the authors argued that two important components or dimensions of deskilling on which they were able to shed light were ‘skill as complexity and skill as freedom’, which are central to the thesis that work is being proletarianized through the deskilling of tasks (Marshall et al., 1988: 116). ‘Skill as complexity’ was measured by a single interview question asking respondents whether their current jobs required more, less, or about the same amount of skill as when they first started. ‘Skill as freedom’ was measured by seven indicators that were treated separately and not aggregated. The questions entailed asking respondents about such things as whether they were able to reduce the pace of their work or to initiate new tasks in their work. Neither dimension comprised measures that offered significant support for the deskilling thesis.

Box 3.7 What is reliability?

Reliability refers to the consistency of a measure of a concept. The following are three prominent factors involved in considering whether a measure is reliable:

- Stability. This consideration entails asking whether a measure is stable over time, so that we can be confident that the results relating to that measure for a sample of respondents do not fluctuate. This means that, if we administer a measure to a group and then readminister it, there will be little variation over time in the results obtained.

- Internal reliability. The key issue is whether the indicators make up the scale or index are consistent—in other words, whether respondents’ scores on any one indicator tend to be related to their scores on the other indicators.

- Inter-observer consistency. When a great deal of subjective judgement is involved in such activities as the recording of observations or the translation of data into categories and where more than one ‘observer’ is involved in such activities, there is the possibility that there is a lack of consistency in their decisions. This can arise in a number of contexts, for example in content analysis where decisions have to be made about how to categorize media items; when answers to open-ended questions have to be categorized; or in structured observation where observers have to decide how to classify subjects’ behaviour.
The calculation of the correlation will yield a figure, also scored high on the other group of indicators. Known as a coefficient, that varies between 0 (no internal consistency) and 1 (perfect correlation and therefore complete internal consistency). It is usually expected that a result of 0.8 and above implies an acceptable level of correlation and therefore no internal consistency.

Nowadays, most researchers use a test of internal consistency. Box 3.7 outlines the issues involved and gives a brief overview of the topic. The chief point to carry away with you at this stage is that the Correlation establishes how closely two aspects of indicators are related.

Results on the 15-item measure varied between 0.79 and 0.95 for each of the fifteen scales. As noted in Chapter 2, the issue of measurement validity has to do with whether a measure of a concept really measures that concept. When people argue about whether a person’s IQ score really measures or reflects that person’s level of intelligence, they are raising questions about the measurement validity of the IQ test in relation to the concept of intelligence. Similarly, one often hears people say that they do not believe that the Retail Price Index really reflects inflation and the rise in the cost of living. Again, a query is being raised in such comments about measurement validity. And whenever students or lecturers debate whether formal examinations are assessing the ability of educational achievement, they too are raising questions about measurement validity.

Validity

As noted in Chapter 2, the issue of measurement validity has to do with whether a measure of a concept really measures that concept (see Box 3.9). When people argue about whether a person’s IQ score really measures or reflects that person’s level of intelligence, they are raising questions about the measurement validity of the IQ test in relation to the concept of intelligence. Similarly, one often hears people say that they do not believe that the Retail Price Index really reflects inflation and the rise in the cost of living. Again, a query is being raised in such comments about measurement validity. And whenever students or lecturers debate whether formal examinations are assessing the ability of educational achievement, they too are raising questions about measurement validity.

Accuracy refers to the issue of whether an indicator (set of indicators) that is devised to gauge a concept really measures that concept. Several ways of establishing accuracy are explored in the text: face validity; concurrent validity; predictive validity; construct validity; and convergent validity. Here the issue is being used as a shorthand for what was referred to as measurement validity in Chapter 2. Accuracy should therefore be distinguished from the other terms introduced in Chapter 2: internal validity; external validity; and ecological validity.

Box 3.8

What is Cronbach’s alpha?

To a very large extent we are leaping ahead too much here, but it is important to appreciate the basic features of what this widely used test of internal reliability is. It essentially calculates the average of all possible split-half reliability coefficients. A computed alpha coefficient will vary between 1 (denoting perfect internal reliability) and 0 (denoting no internal reliability). The figure 0.80 is typically employed as a rule of thumb to denote an acceptable level of internal reliability, though many writers work with a slightly lower figure. In the case of the commitment to work scale devised by Westerling et al., alpha was 0.70, which they refer to as ‘a satisfactory level’ (1989: 93). In the case of Kelley and De Graaf’s (1997) measure of religious orthodoxy, which comprised four indicators, alpha was 0.93. The alpha levels varied between 0.79 and 0.95 for each of the fifteen national samples that make up the data (see Boxes 1.3 and 3.4 for more information about this research). Berthoud (2000: 169) writes that a minimum level of 0.60 is ‘good’ and cites the case of an index of ill-health used in the BHS which achieved a level of 0.77.

Box 3.9

What is validity?

Validity refers to the issue of whether an indicator (set of indicators) that is devised to gauge a concept really measures that concept. Several ways of establishing accuracy are explored in the text: face validity; concurrent validity; predictive validity; construct validity; and convergent validity. Here the term is being used as a shorthand for what was referred to as measurement validity in Chapter 2. Accuracy should therefore be distinguished from the other terms introduced in Chapter 2: internal validity; external validity; and ecological validity.

Predictive validity

Another possible test for the validity of a new measure is predictive validity, whereby the researcher uses a future criterion measure, rather than a contemporary one, as the case of concurrent validity. With predictive validity, the researcher would take future levels of absenteeism as the criterion against which the validity of a new measure of job satisfaction would be examined. The difference from concurrent validity is that a future rather than a simultaneous criterion measure is employed.

Box 3.10

The British Crime Survey

Validity refers to the issue of whether an indicator (set of indicators) that is devised to gauge a concept really measures that concept. Several ways of establishing accuracy are explored in the text: face validity; concurrent validity; predictive validity; construct validity; and convergent validity. Here the term is being used as a shorthand for what was referred to as measurement validity in Chapter 2. Accuracy should therefore be distinguished from the other terms introduced in Chapter 2: internal validity; external validity; and ecological validity.

Construct validity

Some writers advocate that the researcher should also estimate the construct validity of a measure. Here, the researcher is encouraged to deduce hypotheses from a theory that is relevant to the concept. For example, drawing upon ideas about the concept of technology on the experience of work, the researcher might anticipate that people who are satisfied with their jobs are less likely to work on routine jobs, whereas those who are not satisfied are more likely to work on routine jobs. Accordingly, we could investigate this theoretical deduction by examining the relationship between job satisfaction and job routine. However, some caution is required in interpreting the absence of a relationship between job satisfaction and job routine. The researcher may be making a deduction that is made from it might be misguided. Accordingly, the measure of job routine could be an invalid measure of that concept.
The main preoccupations of quantitative researchers

Both quantitative and qualitative research can be viewed as exhibiting a set of distinctive but contrasting preoccupations. These preoccupations reflect epistemologically grounded beliefs about what constitutes acceptable knowledge. In this section, four distinctive preoccupations that can be discerned in quantitative research will be outlined and examined: measurement, causality, generalization, and replication.

Reflections on reliability and validity

There are, then, a number of different ways of investigating the merit of measures that are designed to represent social scientific concepts. However, the discussion of reliability and validity is potentially misleading, because it would be wrong to think that all new measures of concepts are subjected to the rigorous described above. In fact, most typically, measurement is undertaken within a context that Cicourel (1954) described as ‘measurement by fiat’. By the term ‘fiat’, Cicourel was referring not to a well-known Italian car manufacturer but to the notion of ‘decree’. He meant that most measures are simply asserted. Fairly straightforward, but minimal steps may be taken to ensure that a measure is reliable and/or valid, such as testing for internal reliability when a multiple-indicator measure has been devised and examined face validity. But in many, if not the majority of cases in which a concept is measured, no further testing takes place. This point can be made with respect to each of the three criteria of reliability that have been discussed. If the measure is not stable over time, it simply cannot be providing a valid measure. The measure could not be tapping the concept it is supposed to be related to if the measure fluctuated. If the measure fluctuates, it may be measuring different things on different occasions. If a measure lacks internal reliability, it means that a multiple-indicator measure is actually measuring two or more different things. Therefore, the measure cannot be valid. Finally, if there is a lack of inter-observer consistency, it means that observers cannot agree on the meaning of what they are observing, which in turn means that a valid measure cannot be in operation.
The most obvious preoccupation is with measurement, a feature that is scarcely surprising in the light of much of the discussion in the present chapter so far. From the position of quantitative research, measurement carries a number of advantages that were previously outlined. It is not surprising, therefore, that issues of reliability and validity are a concern for quantitative researchers, though this is not always manifested in research practice.

Causality

There is a very strong concern in most quantitative research with explanation. Quantitative researchers are rarely concerned merely to describe how things are, but are keen to say why things are the way they are. This emphasis is also often taken to be a feature of the ways in which the natural sciences proceed. Thus, researchers are often not only interested in a phenomenon like racial prejudice as something to be described, for example, in terms of how much prejudice exists, but also in explaining why it exists, therefore, that issues of reliability and validity are a concern for quantitative researchers, though this is not always manifested in research practice.

Measurement

The simple answer is that it will be the most obvious preoccupation is with measurement, a feature that is scarcely surprising in the light of much of the discussion in the present chapter so far. From the position of quantitative research, measurement carries a number of advantages that were previously outlined. It is not surprising, therefore, that issues of reliability and validity are a concern for quantitative researchers, though this is not always manifested in research practice.

Causality

There is a very strong concern in most quantitative research with explanation. Quantitative researchers are rarely concerned merely to describe how things are, but are keen to say why things are the way they are. This emphasis is also often taken to be a feature of the ways in which the natural sciences proceed. Thus, researchers are often not only interested in a phenomenon like racial prejudice as something to be described, for example, in terms of how much prejudice exists, but also in explaining why it exists, therefore, that issues of reliability and validity are a concern for quantitative researchers, though this is not always manifested in research practice.

Generalization

In quantitative research the researcher is usually concerned to be able to say that his or her findings can be generalized beyond the confines of the particular context in which the research was conducted. Thus, if a study of racial prejudice is carried out by a questionnaire with a number of people who answer the questions, we often want to say that the results are relevant to all similar cities. We should not make inferences beyond the population from which the sample was selected, but researchers frequently do so. The concern to be able to generalize is often so deeply ingrained that the limits to the generalizability of findings are frequently forgotten or sidestepped.

Replication

The natural sciences are often depicted as wishing to reduce to a bare minimum the contaminating influence of the scientist's biases and values. The results of a piece of research should be unaffected by the researcher's special characteristics or expectations or whatever. If biases and lack of objectivity were pervasive, the claims of the natural sciences to provide a definitive picture of the world would be seriously undermined. As a check upon the influence of these potentially damaging problems, scientists often attempt to be highly explicit about their procedures so that an experiment is capable of replication. Likewise, quantitative researchers in the social sciences often regard replication, or more precisely the ability to replicate, as an important ingredient of their activity. It is easy to see why: the possibility of a lack of objectivity and of the intrusion of the researcher's values would appear to be much greater when examining the social world than when the natural scientist investigates the natural order. Consequently, it is often regarded as important that the researcher spells out clearly his or her procedures so that they can be replicated, even if the research does not end up being replicated.

The concern with generalizability or external validity is particularly strong among quantitative researchers using cross-sectional and longitudinal designs. There is a concern about generalizability among experimental research, as the discussion of external validity in Chapter 2 suggested, but users of this research design usually give greater attention to internal validity issues.
The critique of quantitative research

Over the years, quantitative research along with its epistemological and ontological foundations has been the focus of a great deal of criticism, particularly from proponents and spokespersons of qualitative research. To a very large extent, it is difficult to distinguish between different kinds of criticism when reflecting on the different critical points that have been proffered. These include: criticisms of quantitative research in general as a research strategy; criticisms of the epistemological and ontological foundations of quantitative research; and criticisms of specific methods and research designs with which quantitative research is associated.

Criticisms of quantitative research

To give a flavour of the critique of quantitative research, four criticisms will be covered briefly.

- Quantitative researchers fail to distinguish people and social institutions from the ‘world of nature’. The phrase ‘the world of nature’ is from the writings of Schutz and the specific quotation from which it has been taken can be found on p. 13. Schutz and other phenomenologists charge social scientists who employ a natural science model with treating the social world as if it were no different from the natural order. In so doing, they draw attention to one of positivism’s central tenets—namely, that the principles of the scientific method can and should be applied to all phenomena that are the focus of investigation. As Schutz argues, this tactic is essentially to imply that this means turning a blind eye to the difference between the social and natural world. More particularly, as was observed in Chapter 1, it therefore means ignoring and riding roughshod over the fact that people interpret the world around them, whereas this capacity for self-reflection cannot be found among the objects of the natural sciences (‘molecules, atoms, and electrons’, as Schutz put it).

- The measurement process possesses an artificial and spurious sense of precision and accuracy. There are a number of aspects to this criticism. For one thing, it has been argued that the connection between the measures developed by social scientists and the concepts they are supposed to be revealing is assumed rather than real; hence, Cicourel’s (1964) concept of ‘measurement by fiat’. Testing for validity in the manner described in the previous section cannot really address this problem, because the very tests themselves entail measurement by fiat. A further way in which the measurement process is regarded by writers like Cicourel as flawed is that it presumes that when, for example, members of a sample respond to a question on a questionnaire (which is itself taken to be an indicator of a concept), they interpret the key terms in the question similarly. For many writers, respondents simply do not interpret such terms similarly. An often used reaction to this problem is to use questions with fixed-choice answers, but this approach merely provides ‘a solution to the problem of meaning by simply ignoring it’ (Cicourel 1964: 108).

- The reliance on instruments and procedures hinders the connection between research and everyday life. This issue relates to the question of ecological validity that was raised in Chapter 2. Many methods of quantitative research rely heavily on administering research instruments to subjects (such as structured interviews and self-completion questionnaires) or on controlling situations to determine their effects (such as in experiments). However, as Cicourel (1982) asks, how do we know if survey respondents have the requisite knowledge to answer a question or whether they are similar in their sense of the topic being important to them in their everyday lives? Thus, if respondents answer a set of questions designed to measure racial prejudice, can we be sure that they are equally aware of what it is and its manifestations and can we be sure that it is of equal concern to them in the ways in which it connects with everyday life? One can go even further and ask how well their answers relate to their everyday lives. People may answer a question designed to measure racial prejudice, but respondents’ actual behaviour may be at variance with their answers (La Piere 1934).

- The analysis of relationships between variables creates a static view of social life that is independent of people’s lives. Blumer argued that studies that aim to bring out the relationships between variables omit ‘the process of interpretation or definition that goes on in human groups’ (1956: 685). This means that we do not know how what appears to be a relationship between two or more variables has been produced by the people to whom it applies. This criticism incorporates the first and third criticisms that have been referred to—that the meaning of events to individuals is ignored and that we do not know how such findings connect to everyday contexts—but adds a further element—namely, that it creates a sense of a static social world that is separate from the individuals who make it up. In other words, quantitative research is seen as carrying an objectivist ontology that refutes the social world.

We can see in these criticisms the application of a set of concerns associated with a qualitative research strategy that reveals the combination of an interpretivist epistemological orientation (an emphasis on meaning from the individual’s point of view) and a constructionist ontology (an emphasis on viewing the social world as the product of individuals rather than as something beyond them). The criticisms may appear very damaging, but, as we will see in Chapter 13, quantitative researchers have a powerful battery of criticisms of qualitative research in their arsenal as well.

Is it always like this?

One of the problems with characterizing any research strategy, research design, or research method is that to a certain extent one is always outlining an ideal-typical approach. In other words, one tends to create something that represents that strategy, design, or method, but that may not be reflected in its entirety in research practice. This gap between the ideal type and actual practice can arise as a result of at least two major considerations. First, it arises because those of us who write about and teach research methods cannot cover every eventuality that can arise in the process of social research, so that we tend to provide accounts of the research process that draw upon common features. Thus, a model of the process of quantitative research, such as that provided in Figure 3.1, should be thought of as a general tendency rather than as a definitive description of all quantitative research. A second reason why the gap can arise is that, to a very large extent when writing about and teaching research methods, we are essentially providing an account of good practice. The fact of the matter is that these practices are often not followed in the published research that students are likely to encounter in the substantive courses that they will be taking. This failure to follow the procedures associated with
Reverse operationism

As an example of the first source of the gap between the ideal type and actual research practice we can take the case of something that I have referred to as ‘reverse operationism’ (Bryman, 1988a: 28). The model of the process of quantitative research in Box 3.1 implies that concepts are specified and measures are then provided for them. As we have noted, this means that indicators must be devised. This is the basis of the idea of ‘operationism’ or ‘operationalism’, a term that derives from physics (Bridgman 1927), and that implies a deductive view of how research should proceed. However, this view of research neglects the fact that measurement can entail much more of an inductive element than Box 3.1 implies. Sometimes, measures are developed that in turn lead to conceptualization. One way in which this can occur is when a statistical technique known as factor analysis is employed. This technique is employed in relation to multiple-indicator measures to determine whether groups of indicators tend to bunch together to form distinct clusters, referred to as factors. In order to measure the concept of ‘charismatic leadership’, a term that owes a great deal to Weber’s (1947) notion of charismatic authority, Conger and Kanungo (1998) generated twenty-five items to provide a multiple-indicator measure of the concept. These items derived from their reading of existing theory and research on the subject, particularly in connection with charismatic leadership in organizations. When the items were administered to a sample of respondents and the results were factor analysed, it was found that the items bunched around six factors, each of which to all intents and purposes represents a dimension of the concept of charismatic leadership:

- strategic vision and articulation behaviour;
- personal risk;
- sensitivity to organizational members’ needs;
- action orientation away from the maintenance of the status quo.

The point to note is that these six dimensions were not specified at the outset: the link between conceptualization and measurement was an inductive one. Nor is this an unusual situation so far as research is concerned (Bryman 1988a: 25-6).

Reliability and validity testing

The second reason why the gap between the ideal type and actual research practice can arise is because researchers do not follow some of the recommended practices. This is a classic case of this tendency is that, while, as in the present chapter, much time and effort are expended on the articulation of the ways in which the reliability and validity of measures should be determined, a great deal of the time these procedures are not followed. There is evidence from analyses of published quantitative research in organization studies (Podsakoff and Dalton 1987), a field that draws extensively on ideas and methods used in the social sciences, that writers rarely report tests of the stability of their measures and even more rarely report evidence of validity (only 3 per cent of articles provided information about measurement validity). A large proportion of articles used Cronbach’s alpha, but, since this device is relevant only to multiple-item measures, because it gauges internal consistency, the stability and validity of many measures that are employed in the field of organization studies are unknown. This is not to say that the measures are necessarily unstable and invalid, but that we simply do not know. The reasons why the procedures for determining stability and validity are rarely used are almost certainly the cost and time that are likely to be involved. Researchers tend to be concerned with substantive issues and are less than enthusiastic about engaging in the kind of development work that would be required for a thoroughgoing determination of measurement quality.

However, what this means is that Cicourel’s (1964) previous critics of measurement in sociology being ‘measurement by fiat’ has considerable weight.

The remarks on the lack of assessment of the quality of measurement should not be taken as a justification for readers to neglect this phase in their work. My aim is merely to draw attention to some of the ways in which practices described in this book are not always followed and to suggest some reasons why they are not followed.

Sampling

A similar point can be made in relation to sampling, which will be covered in the next chapter. As we will see, good practice is strongly associated with random or probability sampling. However, quite a lot of research is based on non-probability samples—that is, samples that have not been selected in terms of the principles of probability sampling to be discussed in Chapter 4. Sometimes the use of non-probability samples will be due to the impossibility or extreme difficulty of obtaining probability samples. Yet another reason is that the time and cost involved in securing a probability sample are too great relative to the level of resources available. And yet a third reason is that sometimes the opportunity to study a certain group presents itself and represents too good an opportunity to miss. Again, such considerations should not be viewed as a justification and hence a set of reasons for ignoring the principles of probability sampling. The remarks on the lack of assessment of the quality of measurement should not be taken as a justification for readers to neglect this phase in their work. My aim is merely to draw attention to some of the ways in which practices described in this book are not always followed and to suggest some reasons why they are not followed.

The remarks on the lack of assessment of the quality of measurement should not be taken as a justification for readers to neglect this phase in their work. My aim is merely to draw attention to some of the ways in which practices described in this book are not always followed and to suggest some reasons why they are not followed.

**Questions for Review**

The main steps of quantitative research

- What are the main steps in quantitative research?
- To what extent do the main steps follow a strict sequence?
This chapter and the three that follow it are very much concerned with principles and practices associated with social survey research. Sampling principles are not the emphasis will be on sampling in connection with the selection of people who would be asked questions by interviewing or experimentation. The chapter explores: