

# VALIDITY, RELIABILITY AND THE QUALITY OF RESEARCH

Clive Seale

## Chapter Contents

<b>The scientific tradition</b>	<b>529</b>
Measurement validity	529
Internal validity	530
External validity	531
Reliability and replicability	532
<b>The interpretivist tradition</b>	<b>533</b>
Modifications	533
Radical conceptions	539
<b>Conclusion</b>	<b>540</b>

Discussions of the quality of social and cultural research often begin with the ideas of validity and reliability. These derive from the scientific (sometimes thought of as 'positivist') tradition. Thus **validity** refers to the truth-value of a research project; can we say whether the reported results are true? **Reliability**, on the other hand, concerns the consistency with which research procedures deliver their results (whether or not these are

true). Thus we can ask whether a particular questionnaire, if applied on two different occasions to the same person, would generate the same answers. When the concept of reliability is applied to whole research projects, we are asking questions about their **replicability**. That is to say, if we repeated the research project exactly, would we get the same result again? In the scientific tradition, replicable studies using reliable research instruments have

been considered essential preconditions for studies that produce valid or true knowledge. Many procedures and techniques have been devised in order to test validity and reliability and this chapter will demonstrate some of these.

However, the scientific discussion of validity and reliability makes assumptions that sit uncomfortably with many conceptions of qualitative social and cultural research. As was shown in Chapter 2, some researchers in the interpretivist tradition reject realism as an adequate basis for judging the value of research studies, substituting a variety of idealist philosophical conceptions, or indeed political conceptions, of the value of research. Scientific discussions of validity and reliability are firmly rooted in the realist tradition. Here, the task of the researcher is to find something out about the world and report findings in an objective, value-free manner. If, however, research knowledge itself is treated as a social construction, it is hard to sustain a commitment to realism and objectivity. Other criteria must then be used to judge the quality or value of a research study. Perhaps, for example, the quality of a study can be judged according to whether it promotes insight, understanding or dialogue, or in terms of whether it gives voice to particular social groups whose perspective has been hidden from public view. This chapter will first introduce you to scientific conceptions of validity, reliability and replicability and will then show you how a variety of qualitative researchers in the interpretive tradition have approached the issue of

judging the quality of their work. The chapter will conclude with some comments about how you may be able to use these discussions to inform your own research practice.

## The scientific tradition

In the scientific tradition, validity is understood to have various components. These are indicated in Box 30.1.

## Measurement validity

The **measurement validity** of questions in interviews and questionnaires can be improved by various methods (see Chapter 11 for an account of how to design questions for social surveys). The first and perhaps most common method is known as **face validity**, whereby the researcher thinks hard about whether the questions indicate the intended concept. The assessment of face validity may be helped by asking people with practical or professional knowledge of the area to assess how well questions indicate the concept, including their judgements of how comprehensively the various aspects of the concept have been covered. Thus a sequence of questions designed to indicate a person's health status might be assessed by a group of nurses or doctors (the term content validity is also sometimes used to describe such assessment by experts).

### BOX 30.1

#### THE COMPONENTS OF VALIDITY

- **Measurement validity:** the degree to which measures (e.g. questions on a questionnaire) successfully indicate concepts.
- **Internal validity:** the extent to which causal statements are supported by the study.
- **External validity:** the extent to which findings can be generalised to populations or to other settings.

Criterion validity involves comparing the results of questions with established indicators of the same concept. Criterion validity can be concurrent or predictive. Concurrent criterion validity might, for example, involve comparing the results of an interview survey of people's health status with the results of a doctor's examination of the same people done at around the same time. If the interview results differ from the doctor's assessment, the interview would be judged to have poor validity. Predictive criterion validity involves comparisons with what happens in the future. For example, the validity of examinations at school in measuring academic ability might be judged by seeing whether these are good at predicting eventual degree results.

Construct validity evaluates a measure according to how well it conforms to expectations derived from theory. Thus, if we have reason to believe that health status is related to social class, we would expect our measure of health status to give different results for people from different social classes. The construct validity of certain questions may only be established after a series of studies and analyses in which researchers build up a greater understanding of how the questions relate to other constructs.

None of these methods of improving measurement validity is perfect. Argument about the face validity of indicators often reveals disagreement about the meaning of concepts. For example, what do we mean by 'health'? Although our indicator may agree with some external criterion, who is to say that the external criterion is valid? Thus a doctor's judgement about health status is not infallible; sometimes people get poor degrees for reasons other than their academic ability. Construct validity depends both on a theory being correct and on other measures of other concepts in the theory being valid. If social class is not related to health, or if our measure of social class is itself not valid, then associations between

health and social class cannot show the validity of our measure of health.

### Internal validity

It is important to have valid measures if internal validity is to be sustained, but this is not the only necessary component. In order to prove that one thing (A) has caused another (B) three basic conditions must be met. First, A must precede B in time (the problem of time order). Second, A must be associated with B. That is to say, when the measure of A changes, the measure of B must also change. Third, the association must not be caused by some third factor C (the problem of spurious causation).

Thus, in examining the hypothesis that people with a higher educational level (A) therefore subsequently achieve higher income levels (B), it is no good if a person's income is assessed before their education is complete (a time order problem). Additionally, there is unlikely to be a causal relationship if people with a high educational achievement do not differ from those with a low educational achievement in their incomes (in which case we would say that there is no association between the variables). Most difficult to establish in social research, however, is the issue of whether some third variable – such as parental social class (C) – is associated with both educational achievement (e.g. rich parents send their children to private schools) and income (e.g. a private income from family wealth). In this case, an apparent relationship between education and income may be spurious since both educational achievement and income have been affected by the third variable. Ensuring that causal statements are valid is a matter of research design (see Chapter 8) and the adequacy of statistical analysis (see Chapters 18–20). The example in Box 30.2 will also help you understand the ideas involved.

### External validity

In social survey work external validity is ensured by representative sampling, techniques for which are described in Chapter 9. Since a researcher cannot study everyone in a population (unless they do a complete census of all members of that population), there is inevitably a degree of selection involved in choosing

people (or settings) to study. Representative sampling seeks to ensure that the people (or settings) studied are not unusual or atypical in any way, so that what is discovered about them may also hold true for others in the population. Usually, statistical inference is used to make a probabilistic estimate of the likelihood that a result in a randomly selected sample is a freak occurrence (see Chapter 19).

## BOX 30.2

### CONNECTICUT TRAFFIC FATALITIES

#### The causal proposition

In 1960 the Governor of Connecticut announced that a police crackdown on speeding and drunk drivers had resulted in a dramatic reduction from the alarmingly high rate of traffic fatalities that had been evident in 1955.

#### Threats to internal validity

Campbell (1969) listed a number of 'threats' to this as a causal claim, including:

- *History*: For example, the weather might have been better in later years resulting in fewer accidents.
- *Maturation*: Drivers may have been getting more careful anyway.
- *Instability and regression*: Traffic fatality rates go up and down from year to year anyway; 1956 just happened to have a high number of fatalities. In subsequent years a 'regression towards the mean' was therefore pretty likely.
- *Testing*: Perhaps publishing the high 1955 death rate made people more careful when driving.
- *Instrumentation*: Perhaps the method for estimating number of deaths changed. For example, in 1955 death could have been recorded according to whether people resided in Connecticut, whereas in 1956 death might have been recorded according to place of death (or vice versa).
- *Selection*: This would occur, for example, if the population of Connecticut had undergone a change. Perhaps an economic boom produced an influx of young male drivers with cheap cars during the high fatality year.
- *Experimental mortality*: Perhaps fewer counties in the state returned death statistics in one year compared to the other.

However, it is quite common for there to be shortcomings in the degree to which social and psychological research done from within the scientific tradition deals with external validity.

This can be particularly evident in experiments where people are recruited as volunteers. An experimenter may discover that a group of volunteers behave in a certain way under experimental

conditions, but if the volunteers are different from the people to whom the result is to be generalised, external validity may be poor. Additionally, an experimental situation may not be very good at mimicking the conditions of real life.

Consider the traffic fatalities example (Box 30.1). Imagine that the problems of internal validity were overcome. As a complete census of Connecticut drivers, there would be no problem then in drawing conclusions about police influence on drivers' behaviour in Connecticut. But if drivers' attitudes to the police, or indeed police behaviour during crackdowns were different in other places, a different impact on driver behaviour might be experienced.

### Reliability and replicability

A study can be reliable without being valid. Consider an archery target: arrows can strike it consistently (reliably) in the wrong place. Thus a measurement can be consistently wrong. At the same time, as the second target shows, a valid measure is not necessarily reliable if the object being measured is changing: perhaps the target is moving? Figure 30.1 shows these ideas in visual form. See if you can interpret the third and fourth targets.

In the realist, scientific tradition it is important to get consistent results when observations are being made, or questions are being asked. If different researchers use the same interview

schedule it is no good if they get different results with the same person (assuming that the person has not changed their views between interviews). Similarly, if researchers applying a coding scheme (see Chapter 21) for analysing data disagree amongst each other about how to assign codes, it is hard to place much faith in their objectivity. For this reason, questionnaires, interview schedules, measuring devices of various sorts and coding schemes are often subjected to tests of their reliability, sometimes involving *inter-rater reliability* tests (described in detail in Chapter 26). Thus a questionnaire designed to measure political preferences might be tested by being applied to the same group of respondents twice by different researchers. If the results are the same each time, even though different researchers have used it, the questionnaire is said to be reliable. If different researchers categorise the same qualitative answers from a survey in the same way, *inter-coder* reliability is said to be high.

More broadly, replicability is at stake when comparing different studies of the same problem that have used the same or similar methods. In the early days of scientific studies it was considered important to develop a style of research reporting so that other investigators could repeat studies and, hopefully, get the same results. This would then increase faith in the truth-value of the findings because they would be seen to have been replicated by other investigators. For this reason, accounts of method in research reports may be quite detailed.

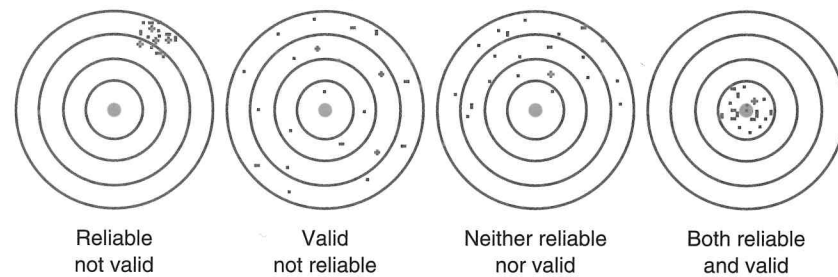


FIGURE 30.1 The relationship between reliability and validity (Trochim, 2003)

### The interpretivist tradition

It would be wrong to say that all qualitative, interpretivist approaches to research make a radical break with the conceptions of validity and reliability thus far outlined. Quite a lot of qualitative researchers pursue a broadly *realist* and scientific agenda, and so can often apply the ideas of internal and external validity and reliability to their work, though some modifications may be necessary. In other cases, though, particularly if associated with idealist and social constructionist perspectives, this is a greater problem and quite different notions of quality come into play. These reflect profoundly different conceptions about the purposes and status of the knowledge that researchers produce and ultimately relate to differing philosophical and political considerations. I will therefore first describe modifications and then radical breaks from the scientific tradition.

### Modifications

Realist qualitative or interpretivist research often involves intensive study of single settings (case studies) or a small number of people. In ethnography, for example, a researcher may spend a considerable amount of time participating in the everyday life of a particular social group so that it can be studied in considerable depth (see Chapter 14). The advantage of doing this is often claimed to be that of *naturalism*: the capacity to reveal how people behave as they ordinarily ('naturally') go about their lives. This is felt to contrast with less naturalistic methods (such as interviews) which temporarily extract people from their daily lives so that they can answer questions about events that they may not actually have to face in real life, or about which they may give misleading answers. Qualitative, exploratory interviews, though, are sometimes said to be superior to the structured ones favoured by survey

researchers in the scientific tradition in that they allow the perspectives and priorities of individuals to be revealed, without imposition of the pre-conceptions of the researcher (see Chapter 12).

Both ethnographic method and qualitative interviewing are very time-consuming, though, and can normally only be applied to very few cases, settings or people. In other words, the *breadth* of a social survey may be sacrificed for *depth*, meaning that representativeness and therefore external validity may be seen as questionable. At the same time, an exploratory approach can reveal phenomena that have not been predicted in advance. Thus it can be said that quantitative research often establishes the prevalence of things already known about, whereas in-depth case study research can find things that no one has ever noticed before. Originality and discovery, then, might be seen as indicators of the quality of qualitative research, with external validity being of lesser importance. Some people (e.g. Mitchell, 1983) have expressed this as *theoretical generalisation*, contrasting this with the 'empirical generalisation' of statistical studies. This is because, when a new phenomenon is discovered, its importance can only be judged by reference to its contribution to some existing body of knowledge, or 'theory'. Thus, discovering a black stone on a pebble beach may seem to be of no great importance in its own right, but if such a discovery is made in an area where geological conditions were thought to make black stones impossible, the finding acquires greater significance. This is why some qualitative researchers nowadays like to speak about 'theorising' an area of inquiry: only when a finding is placed in a relevant theoretical context can it acquire significance, so a knowledge of social theory may be particularly important for qualitative inquirers. Chapter 3 discusses a variety of ways in which social theory can be incorporated into research practice.



Additionally, internal validity may take on a different meaning in interpretivist research. Causal inquiry has got itself a bad name in some qualitative research circles, being associated with a deterministic model of human agency that denies the capacity of people to exercise free will and fails to explore meaning-making activities in social life. Yet causal statements are pretty much inevitable in any discussion of human social and cultural life. If you look

closely at research reports they will always contain implied causal mechanisms. Box 30.3 contains an illustration of this. Additionally, not all statistical work is devoted to proving causality, but instead is descriptive. It is true to say, though, that proving the existence of causality is only very rarely an interest of qualitative researchers, so scientific notions of internal validity are not much use in assessing the quality of such studies.

### BOX 30.3

#### IMPLICIT CAUSAL REASONING IN A QUALITATIVE STUDIES

The general region from which the immigrant came was also important in the organization of Cornerville life. The North Italians, who had greater economic and educational opportunities, always looked down upon the southerners, and the Sicilians occupied the lowest position of all. (Whyte, 1943/1981: xvii)

Whyte, in this passage of 'description', is proposing a causal relationship between region of origin and Cornerville pecking order. Further, he is suggesting that economic and educational differences between Italian regions influence this.

Measurement validity, which has been shown to be an essential precondition for the internal validity of statistical studies, might be seen quite straightforwardly as an important aspect of quality in qualitative research. Of course, measurement may not be something attempted by qualitative researchers (although they may sometimes count things), but the underlying issue in measurement validity is the adequacy of links between concepts and their indicators (*concept-indicator links*). These links are important in qualitative research too, and grounded theory is an approach that prioritises the creation of good concept-indicator links. Chapter 22 describes grounded theory in depth, but for the

moment you should note that it is based on creating new concepts and ideas and the relations between them (in other words, theory) from observations of social settings. This contrasts with an approach that starts with theory and then seeks empirical examples. As a result, research reports based on grounded theorising generally exhibit excellent links between concepts and the examples drawn from data. In this sense, qualitative researchers can be thought of as being concerned with a form of 'measurement validity'. A good qualitative report exemplifies concepts with good examples.

Yet there are recognisable difficulties in applying the scientific paradigm to qualitative

research work. Various authors have therefore proposed modified schemes. Lincoln and Guba's (1985) account of quality issues in what they call 'naturalistic inquiry' (drawing on the meaning of naturalism that refers to the study of people in their normal or 'natural' settings) is one such effort and is shown in Box 30.4.

These authors are critical of the notion of 'truth-value,' saying that it assumes a 'single tangible reality that an investigation is intended to unearth and display' (1985: 294), whereas the naturalistic researcher makes 'the assumption of

multiple constructed realities' (1985: 295). In this respect they reveal a dissatisfaction with crude realism and appear to be moving towards a social constructionist epistemological position (see Chapter 2). They argue, then, that *credibility* should replace 'truth-value'. Through prolonged engagement in the field, persistent observation and triangulation exercises, as well as exposure of the research report to criticism by other researchers and a search for negative instances that challenge emerging hypotheses and demand their reformulation, credibility is built up.

### BOX 30.4

#### LINCOLN AND GUBA'S TRANSLATION OF TERMS

Conventional inquiry	Naturalistic inquiry
Truth-value (Internal validity)	Credibility
Applicability (External validity)	Transferability
Consistency (Reliability)	Dependability
Neutrality (Objectivity)	Confirmability

*Triangulation* is a technique advocated by Denzin (1978) for validating observational data. Denzin outlines four types of triangulation:

- 1 *Data triangulation* involves using diverse sources of data, so that one seeks out instances of a phenomenon in several different settings, at different points in time or space. Richer descriptions of phenomena then result.
- 2 *Investigator triangulation* involves team research; with multiple observers in the field, engaging in continuing discussion of their points of difference and similarity, personal biases can be reduced.
- 3 *Theory triangulation* suggests that researchers approach data with several hypotheses in mind, to see how each fares in relation to the data.
- 4 *Methodological triangulation* is the most widely understood and applied approach. This, for Denzin, ideally involves a 'between-method' approach, which can take several forms but, classically, might be illustrated by a combination of ethnographic observation with interviews. Additionally, methodological triangulation is frequently cited as a rationale for mixing qualitative and quantitative methods in a study (see Chapter 27).

### BOX 30.5

#### AN EXAMPLE OF METHODOLOGICAL TRIANGULATION

Rossman and Wilson (1994) describe a project to investigate the impact on school organisation of state authorities' introduction of minimum competency tests in schools. This combined qualitative interviews with school teachers and other educationists in 12 school districts with a postal questionnaire of a larger sample. Analysis of the questionnaire results suggested that curricular adjustments were more common in school districts where teachers reported that their relationship with state educational authorities was 'positive'. The qualitative interviews sought and found corroboration of this. Thus, for example, in a district where no changes occurred in the curriculum, a local administrator said, 'The state has become someone we have to beat rather than a partner to work with' (1994: 320–321). The authors go on to say:

On the other extreme was a district that accepted the state's increased role in monitoring educational outcomes and worked hard to find creative instructional techniques to improve student performance. The qualitative descriptions of how these two districts responded to the state mandate corroborated and offered convergence to the quantitative findings. (1994: 321)

Negative instances are instances of data (sometimes also called 'deviant cases') that contradict emerging analyses, generalisations and theories. Discovery of these can have a variety of effects, sometimes leading to the abandonment of ideas,

but more often to a deeper analysis that accounts for a wider variety of circumstances. An example of a negative instance found in a research study that extended an initial analysis produced by another investigator is shown in Box 30.6.

### BOX 30.6

#### TYPIFICATIONS AND PERSONAL RESPONSIBILITY IN HOSPITAL CASUALTY DEPARTMENTS

##### The initial generalisation (Jeffery, 1979)

In hospital casualty departments, staff categorise patients as 'bad' if they have problems deemed to be trivial, or are drunks, tramps or victims of self-harm. On the other hand, if patients have problems which allow doctors to practise and learn new clinical skills, or test the professional knowledge of staff, they are categorised as 'good'.

##### The negative instance (Dingwall and Murray, 1983)

Children in casualty departments often exhibit the qualities identified by Jeffery as being those of the 'bad' adult patients, being uncooperative for example, or suffering from mild or self-inflicted injuries. Yet staff do not treat them harshly.

#### Reformulation of the generalisation by Dingwall and Murray, (1983)

Labels applied by staff depend on a prior assessment of whether patients are perceived as being able to make choices (children are not, adults are, on the whole). Children are therefore generally 'forgiven' behaviour that in adults would be deemed reprehensible on the grounds that children are understandably irresponsible. Additionally, staff assess whether the situation is such that patients are able to make choices. Thus, some adults might be categorised as being present in casualty inappropriately, rather than being 'bad' patients if the events that led them there are not their 'fault' (e.g. they had been given poor advice to go to casualty from a person in authority).

Returning to Lincoln and Guba (Box 30.4), these authors also advise researchers to ' earmark ' a portion of data to be excluded from the main analysis, returned to later once analysis has been done in order to check the applicability of concepts. But 'the most crucial technique for establishing credibility', they say, is through 'member checks' (1985: 314), showing materials such as interview transcripts and research reports to the people on whom the research has been done, so that they can indicate their agreement or disagreement with the way in which the researcher has represented them (this can also be called *member validation*).

'Applicability', in Lincoln and Guba's view, depends on generalising from a sample to a population on the untested assumption that the 'receiving' population is similar to that of the 'sending' sample. The naturalistic inquirer, on the other hand, would claim the potential uniqueness of *every* local context. This means study of both sending and receiving contexts so that *transferability* is established. This is clearly quite demanding and, apart from theoretical generalisation (see above), other conceptions of transferability in qualitative research are possible. For example, it can be argued that a very detailed or thick description of a setting can give a reader of a research report the vicarious experience of 'being there', in the same way as a good travel writer can facilitate armchair 'travelling' (Geertz, 1973, 1988). The reader is then well equipped to assess the similarity of the setting described in the research report to settings in

which she or he has personal experience (see Chapter 14 for a discussion of thick description).

To replace consistency, or reliability as conventionally conceived, Lincoln and Guba propose *dependability*, which can be achieved by a procedure they call *auditing*. This involves 'auditors' scrutinising the adequacy of an 'audit trail', consisting of the researchers' documentation of data, methods and decisions made during a project, as well as its end product. Auditing is also useful in establishing *confirmability*, Lincoln and Guba's fourth criterion, designed to replace the conventional criterion of neutrality or objectivity. Auditing is also an exercise in *reflexivity* which involves the provision of a methodologically self-critical account of how the research was done. The authors conclude by pointing out that the trustworthiness of a qualitative, naturalistic study is always negotiable and open-ended, not being a matter of final proof whereby readers are compelled to accept an account.

Lincoln and Guba's philosophical position is (at this stage in their writing) half-way between realism and idealism. As we saw, they are dissatisfied with the crude realism that they feel characterises the conventional, scientific view of validity and reliability and, at some points, speak of 'multiple realities,' something which is normally associated with a social constructionist, idealist view (see Chapter 2). Another way of describing such half-way positions is Hammersley's (1992b) term: *subtle realism*. Here, there is recognition of the existence of a

social world that exists independently of the researcher's mind, but also recognition of the impossibility of knowing this world in any final, certain sense. Research reports can only approach reality in various ways.

This subtle realist position, for Hammersley, leads to an emphasis on the *plausibility* and the credibility of research reports. In assessing the claims made in a research report, Hammersley argues that we should first assess how plausible these are in the light of what is already known about the subject. If a research study contradicts existing knowledge, we need quite compelling evidence in support of its claims. Credibility refers to the adequacy of the links between claims and evidence within the report. It is important to provide the strongest of evidence for the most important claims; lesser claims may need less stringent proof. Additionally, we may wish to assess the relevance of a research study for political, policy-related or practical concerns (see Chapter 4).

Hammersley has been described as a *post-positivist*, signalling his position as one who

modifies 'positivist' or 'scientific' conceptions of validity and reliability in order to apply somewhat similar thinking to qualitative, interpretivist research work. Also within this post-positivist tradition can be placed the work of Becker (1970) and Glaser and Strauss (1967). A later writer who adopts a somewhat scientific conception of the quality of qualitative research is Silverman (2001). These authors represent a tradition that has advocated a number of practical ways in which the quality of qualitative research may be enhanced, listed in Box 30.7 along with references to fuller discussions. Some of these have been mentioned in this chapter or are explained more fully elsewhere in this book. Not all of these authors would agree on all of these things, and their discussions of them contain many subtleties and reservations that cannot be discussed fully here, but you can use the items as a guide to further exploration of these issues and techniques.

### BOX 30.7

#### WAYS OF ENHANCING THE QUALITY OF QUALITATIVE RESEARCH

- *Triangulation* (Seale, 1999a: Ch. 5; see also Box 30.5)
- *Member validation* (Seale, 1999a: Ch. 5)
- Search and account for *negative instances* or deviant cases that contradict emerging ideas (Seale, 1999a: Ch. 6; see also Box 30.6)
- Produce *well-grounded theory* with good examples of concepts (Seale, 1999a: Ch. 7; Chapter 22 in this book)
- Demonstrate the *originality* of findings by relating these to current social issues or social theories (Seale, 1999a: Ch. 8; Mitchell, 1983; Chapters 3 and 4 in this book)
- *Combine* qualitative and quantitative methods (Seale, 1999a: chs 8 and 9; Chapter 27 in this book)
- Use *low inference descriptors* that show the reader a very full account of observations made, reducing the extent to which the researcher's interpretations are involved in recording raw data, as in conversation analytic transcriptions (Seale, 1999a: Ch. 10; Chapter 24 in this book)
- Present a *reflexive* account of the research process so that the reader can see where the ideas and claims come from (Seale, 1999a: Ch. 11; Chapter 5 in this book)

### Radical conceptions

Lincoln and Guba occupy an interesting position in these debates because, even in their 1985 book, they sat rather uneasily in the post-positivist or subtle realist camp. In later work (Guba and Lincoln, 1994) they reveal a more radical position and it is worth examining the shift in their thinking that occurs here since it gets to the heart of the difference between the modified and the radical views.

As we saw, at one point they referred to 'multiple constructed realities' lying at the heart of their position, thus revealing themselves to be, at the philosophical level, occupying a relativist or social constructionist position (see Chapter 2). In this respect they differ from post-positivists like Hammersley, though by 1985 it seems they had not fully worked through the implications of this for research practice. Relativism, if applied to the truth status of research reports themselves, suggests that these are humanly constructed 'versions' of the world, perhaps written out of a commitment to certain value positions or political interests. This contrasts with a view of research as an objective report on the world. Instead, research reports are really no more than 'representations' of the social and cultural world and should be assessed as 'partial truths' (Clifford and Marcus, 1986). Chapter 14 assesses the application of this view to ethnography, demonstrating that – particularly in the discipline of anthropology – a view of research as representation has led to a deeper understanding of the political uses of research knowledge. In the case of anthropology, for example, a view has emerged that is highly critical of the involvement of this (supposedly 'objective') discipline in supporting oppressive colonialist views.

Bauman (1987) is another writer who has thought deeply about the politics of research knowledge. He distinguishes between two positions on this, which can be broadly equated with those of post-positivism and postmodernism:

- One view of research knowledge, Bauman argues, is that it is an attempt to *legislate* on the truth, so that debates can be resolved once and for all. The researcher occupies a superior position, employing methods that provide a better, more authoritative view than those employed in everyday life.
- A second view, though, is that researchers are more like *interpreters*, who generate conversations between groups of people who may not yet have communicated. Thus a researcher or an intellectual, Bauman says, occupies a facilitative role in society, encouraging debate rather than ruling on the truth.

Research understood from this second perspective starts to lose its distinction from social commentary. The distinction between 'data' and 'theory' begins to break down, being revealed as a hangover from a past scientific age. Data, after all, is pre-constituted by the theories and values of the researcher so that it cannot be regarded as an objective account of reality (see Chapter 2). Rather than looking to the inner qualities of a research account in order to judge its quality, some say that it would therefore be better to examine the effects of a research study in society in order to see whether it is good or bad. In their 1994 book, Guba and Lincoln begin to outline this view by presenting a fifth criterion for judging the quality of naturalistic inquiry: *authenticity*.

In describing this, Guba and Lincoln reveal a sympathy for political conceptions of the role of research that goes several steps beyond Hammersley's concern with political and practical relevance (see earlier). Authenticity, they say, is demonstrated if researchers can show that they have represented a range of different realities ('fairness'). Research should also help people develop 'more sophisticated' understandings of the phenomenon being studied ('ontological authenticity'), be shown to have helped people appreciate the viewpoints of people other than themselves ('educative authenticity'), to have stimulated some form of action ('catalytic authenticity') and to have empowered people to act ('tactical authenticity').



Of course, the view that fairness, sophistication, mutual understanding and empowerment are generally desirable is itself a value-laden position. It represents an attempt to pull back from the relativist abyss by founding research practice on a bedrock of political values. Attempts to implement 'democratic' values like this are not always appreciated by people who prefer to organise their lives and political systems according to alternative values. But it can be seen that Guba and Lincoln have travelled on a path beginning with a rejection of positivist criteria and the substitution of interpretivist alternatives. Dissatisfied with the limitations of these, constructionism has been embraced, introducing an element of relativism. Political versions of the value of research have then been imported to save facing the logical implications of relativism, which might end in a nihilistic vision and abandonment of the research enterprise.

This is a path that other qualitative researchers have trodden. Working together with Yvonna Lincoln, Norman Denzin has been influential in promoting political conceptions of the research enterprise, arguing that qualitative research has reached a moment in its development where postmodernist and constructionist influences have resulted in a 'crisis of legitimation'. They argue that

[t]he qualitative researcher is not an objective, authoritative, politically neutral observer standing outside and above the text ... Qualitative inquiry is properly conceptualized as a civic, participatory, collaborative project. This joins the researcher and the researched in an ongoing moral dialogue. (Denzin and Lincoln, 2000: 1049)

This follows on from an earlier statement in which they say that a central commitment of qualitative researchers remains

in the humanistic commitment of the qualitative researcher to study the world always from the perspective of the interacting

individual. From this simple commitment flow the liberal and radical politics of qualitative research. Action, feminist, clinical, constructivist, ethnic, critical and cultural studies researchers are all united on this point. They all share the belief that a politics of liberation must always begin with the perspectives, desires, and dreams of those individuals and groups who have been oppressed by the larger ideological, economic, and political forces of a society, or a historical moment. (Denzin and Lincoln, 1994: 575)

As a criterion for judging the quality of research it is immediately obvious that this is open to dispute. It is not difficult to imagine a well-conducted study that enabled people in positions of power to achieve their aims. The vision of society as no more than a system inhabited by oppressors and oppressed also seems naive (see also Hammersley, 1995a). Research can at times be more relevant to direct political projects, at others less relevant, but its quality is an issue somewhat independent of this.

## Conclusion

As a practising researcher you may be wondering which of these conceptions suits you best. Are you going to commit yourself to a scientific vision in which you prioritise objectivity and replicability, or to a post-positivist position in which you retain some of this commitment in a modified form, or will you reject these in favour of a political conception of the research process? Clearly, the views of the various authors are often incompatible. It seems wrong to develop a measuring instrument that can be judged reliable and valid if the measuring instrument is really no more than an imposition of a particular, value-laden vision of the world on oppressed people. It seems foolish to assess a research report solely according to its political consequences if its findings and claims are poorly supported with

evidence, or if the analysis of evidence is clearly influenced by the researcher's values.

In these disputatious circumstances many researchers seem to feel that they must belong to one camp or another, to identify themselves as 'scientists', 'subtle realists' or 'radical constructionists' before they begin their research activities. In my view this is a mistake. Many of the disputes that exist at the level of methodological debate are simply not resolvable by further discussion, but are a matter of preference. Depending on the actual topic of the research and the problems that are seen to be central, certain considerations will always be more important than others. The personal biographical situation and local circumstances of researchers and their

likely audiences are the main influences on how projects proceed and quality is judged. Exposure to methodological discussions such as the ones outlined in this chapter can help in producing generalised methodological awareness that can be helpful when actually carrying out a research project or intellectual inquiry. Thus a researcher who is aware of these debates is more likely than one who is not to produce a research study that is sophisticated. That is to say, it will be a study that is sensitive to a variety of ways in which it is possible to proceed, show awareness of the consequences of particular decisions made during the course of the study, and the eventual report will demonstrate to a variety of potential audiences that something of value has been created.

## FURTHER READING

This chapter is a condensed and simplified version of a book on the quality of qualitative research which I wrote (Seale, 1999a) and which is the best place to start in expanding your knowledge of this area. The book contains an account of validity and reliability in the quantitative tradition as well. There are relevant chapters by Gobo, Flybjerg and Seale in *Qualitative Research Practice* (Seale et al., 2004).

### Student Reader (Seale, 2004b): relevant readings

- 6 Thomas D. Cook and Donald T. Campbell: 'Validity'
  - 25 R.C. Lewontin: 'Sex lies and social science'
  - 35 Martyn Hammersley: 'Some reflections on ethnography and validity'
  - 48 Anssi Peräkylä: 'Reliability and validity in research based on tapes and transcripts'
  - 62 Zygmunt Bauman: 'Intellectuals: from modern legislators to post-modern interpreters'
  - 65 Patti Lather: 'Fertile obsession: validity after poststructuralism'
  - 66 Thomas A. Schwandt: 'Farewell to criteriology'
  - 79 Maureen Cain and Janet Finch: 'Towards a rehabilitation of data'
- See also Chapter 27, 'Quality in qualitative research' by Clive Seale in Seale et al. (2004).

### Journal articles discussing the issues raised in this chapter

- Seale, C. (1999b) 'Quality in qualitative research', *Qualitative Inquiry*, 5: 465–478.
- Meyrick, J. (2006) 'What is good qualitative research?: a first step towards a comprehensive approach to judging rigour/quality', *Journal of Health Psychology*, 11: 799–808.

(Continued)

(Continued)

Decker, S.H. and Pyrooz, D.C. (2010) 'On the validity and reliability of gang homicide: a comparison of disparate sources', *Homicide Studies*, 14: 359–376.

Walwyn, R. and Roberts, C. (2010) 'Therapist variation within randomised trials of psychotherapy: implications for precision, internal and external validity', *Statistical Methods in Medical Research*, 19: 291–315.

### Web links

Research methods knowledge base: [www.socialresearchmethods.net](http://www.socialresearchmethods.net)

Validity and reliability in quantitative research: <http://allpsych.com/researchmethods/variablesvalidityreliability.html>

A Framework for Assessing the Quality of Qualitative Research: [www.nationalschool.gov.uk/policyhub/news\\_item/qual\\_framework.asp](http://www.nationalschool.gov.uk/policyhub/news_item/qual_framework.asp)

Methods@manchester – What is quality in qualitative research?: [www.methods.manchester.ac.uk/methods/qualityinqualii/index.shtml](http://www.methods.manchester.ac.uk/methods/qualityinqualii/index.shtml)

e-Source – Chapter 7 on 'Observational studies' by Richard Berk: [www.esourceresearch.org](http://www.esourceresearch.org)

## KEY CONCEPTS FOR REVIEW

**Advice:** Use these, along with the review questions in the next section, to test your knowledge of the contents of this chapter. Try to define each of the key concepts listed here; if you have understood this chapter you should be able to do this. Check your definitions against the definition in the glossary at the end of the book.

Auditing	Naturalism
Authenticity	Negative instances
Concept–indicator links	Plausibility
Confirmability	Post-positivism
Construct validity	Reflexivity
Credibility	Reliability
Criterion validity (concurrent and predictive)	Replicability
Dependability	Spurious causation
External validity	Subtle realism
Face validity	Theoretical generalisation
Inter-coder/inter-rater reliability	Time order
Internal validity	Transferability
Measurement validity	Triangulation
Member validation	Validity

## Review questions

- 1 What is the difference between validity, reliability and replicability?
- 2 Outline different ways of improving (a) measurement validity, (b) internal validity, and (c) external validity.
- 3 Describe what is meant by each of the following terms: credibility, transferability, dependability and confirmability. How do they differ from more scientific conceptions of reliability and validity?
- 4 How can triangulation and searching for negative instance help improve the quality of qualitative research?
- 5 What is authenticity and how might it be achieved in a research study?

## Workshop and discussion exercises

- 1 How would you design a study of the causal influence of police crackdowns on driving behaviour that overcame the threats to internal validity listed in Box 30.2 and the threat to external validity mentioned later in the chapter?
- 2 Seek out and read two studies that represent different 'moments' in the history of qualitative research. For example, choose a study that involves grounded theorising and another where the author situates him- or herself within postmodernism. How do the studies differ in their conception of what makes a good research study? How might each author apply these criteria to the other's work?
- 3 Choose a research study in an area of work where you have some knowledge of existing literature and assess it in the light of the following questions:
  - (a) How consistent are the findings with what is already known?
  - (b) What evidence is supplied to support the credibility of the conclusions and how persuasive is this?
  - (c) What relevance might the study have for political or practical affairs?
- 4 In relation to a specific study, consider whether its quality would be improved by attention to the issues raised under the 'positivist' headings of measurement validity, internal and external validity and reliability. To what extent could the modified interpretivist criteria outlined in the chapter be applied to the study? Do these lead you to consider different issues from those raised under the 'positivist' headings?
- 5 This exercise requires you to work with others on some qualitative data, such as some interview transcripts.
  - Without discussing your ideas with others in your group, read one part of the data transcript (e.g. a single interview) and draw up a list of key themes you perceive in the data.
  - Compare the themes you have identified with those of others in your group. What are the similarities and differences?
  - Take four or five themes from those identified by members of the group and, working individually again, apply them to some new data (e.g. a second interview) by marking parts of the transcript which you believe exemplify each theme.
  - Compare what you have done with others in the group. What difficulties are there in consistently applying the themes? Does inconsistency matter?