4

Principles of social research

This chapter

- **focuses on the basic rules of research**
- **presents the main types of measurement in detail**
- **introduces the nature of validity and reliability in research**
- **considers the role of representativeness and generalization**
- **examines the relevance of these principles for qualitative research.**

Key headings

Introduction

Regardless of its diverse and pluralistic nature, structure and process, social research is expected to adhere to certain standards and principles. The nature of these standards and principles may vary, but their presence and necessity are taken for granted. For quantitative researchers, for instance, this is a reflection of the nature of social research, which requires it to rest on sound and reliable criteria. A list of their eight standard principles are listed in Box 4.1, a number of which are accepted by qualitative researchers, although in a different format.

In this chapter we shall explore and discuss these research principles and demonstrate how they are employed and justified by quantitative and qualitative researchers. Many of

these principles will be discussed in more detail in other chapters directly and/or indirectly, and more extensively.

Box 4.1

Principles of quantitative research

- Precision in measurement
- Replication
- Validity
- Reliability
- Objectivity
- Ethics
- Representativeness
- Generalizability

1 Measurement

1.1 Introduction

The research process usually begins with the theoretical preparation or formulation of the research topic. This aims to establish the foundations for the remaining parts of the study, deserves special consideration and will be discussed in this chapter. Before we enter this discussion, however, we need to consider another general and fundamental research procedure, namely measurement. The concept, nature, major characteristics and types of measurement will be introduced next.

1.2 Nature of measurement

Social research, irrespective of its type and nature, entails a degree of measurement. This involves categorizing and/or assigning values to concepts, and is diverse in nature and level of operation. It is also a very useful procedure because it serves to ensure high quality in social research. Most of all, measurement is undertaken to facilitate adequacy, uniformity, comparisons, consistency, accuracy and precision in describing and assessing concepts.

Box 4.2

Measurement facilitates:

- adequacy in description and assessment, offering a full account of the concept
- uniformity in description and assessment, over time and among researchers
- comparisons between complex concepts, enabling the identification of fine distinctions
- consistency in the assessment of concepts, over time and among researchers
- accuracy and precision in procedures, by taking into consideration all aspects of concepts
- replicability in social research, by the same or different researchers, in the same or different contexts.

Generally, measurement may be quantitative or qualitative. Quantitative measurement concentrates on numerical values and attributes. Qualitative measurement refers to labels, names and qualities. Qualitative measurement describes attributes by using common concepts or symbols or introducing new ones; a common procedure involves description

of categories and classifications. The classification of 'residence' into 'urban' and 'rural', for instance, is a qualitative measurement. In the view of some writers, qualitative measurement does not qualify as 'measurement' since it does not demonstrate the main criteria of measurement, such as precision, reliability and validity; rather, it is a process of labelling, classification and description. Nevertheless, as we shall see soon, this view is not fully accepted.

1.3 Variables

Variables are empirical constructs that take more than one value or intensity; for example, sex (male, female), marital status (single, married, divorced, widowed, deserted), age and education are variables. The opposite of variables are constants. Constants take only one value or intensity. The researcher determines at the outset of the study which concepts will act as constants and which as variables.

The construction of variables follows a systematic procedure that adheres to the rules of measurement. Failure to meet these requirements will result in distortions and inaccuracies in measurement, and hence in false results. Two very important rules are that variables must *relate to one concept only* and *must be measurable*.

There are many types of variables. These types vary according to a number of criteria, such as their nature (geographic variables, demographic variables etc.) and their position within the research context (dependent variables, independent variables etc.). The following types are most common:

- *Dependent and independent variables*: An independent variable (IV) is a variable that is set to cause changes in or explain another; a dependent variable (DV) is a variable that is set to be affected or explained by another variable. For instance, in a research study of 'family status and scholastic achievement', the independent variable can be family status, and the dependent variable scholastic achievement. This distinction is not associated with the nature of the concepts but rather with the nature of the research design. The same variable (e.g. scholastic achievement) can be an independent variable in one project (when studying the question: does scholastic achievement affect alcohol consumption?), but a dependent variable in another (when studying the question: does gender affect scholastic achievement?).
- *Extraneous variables*: These are variables which are 'outside' the research question, argument or hypothesis; they are distinct from the dependent or independent variable. For instance, in a study investigating the validity of the theory that height (IV) is associated with scholastic achievement (DV) (e.g. with short students doing far better than tall students), the real reason for the changes in scholastic achievement may be not height (IV) but income and prejudice, which were unexpected and unplanned, and not included in the original research. In this case, income and prejudice are extraneous variables. Hence, looking for the presence of possible extraneous variables is always a logical and necessary procedure in social research, before the relationship between IV and DV is confirmed.

- *Discrete and continuous variables*: Discrete and continuous variables differ from each other in terms of scale continuity; the former are not continuous but use whole units only, whereas the latter are continuous and can be fractioned indefinitely. In discrete variables (also called binomial variables), measurement uses whole units, with no possible values between adjacent units. For instance, marital status is a discrete variable; it can be 'single', married', 'divorced', 'widowed' or 'cohabiting'. Similarly, gender is a discrete variable. Gender can be either 'male' or 'female'. In both examples, there is nothing between the values. In contrast, 'weight' is a continuous variable; it can use smaller increments of units, for example it can be 73.2, 78.1 or 85.6 kg. Discrete variables are counted, not measured; continuous variables are measured, not counted. Examples of discrete variables are ethnicity, race, sex, marital status, cause of death or blood type. Examples of continuous variables are height, distance, time, age, temperature or IQ scores.
- *Demographic variables*: Demographic variables deal with demographic data such as age, residence, religion, marital status, family size, race, education and sexual preference. Remember, a demographic variable can be dependent or independent, discrete or continuous, depending on the nature of the research design.

There are many more distinctions in variables, most of which are common when advanced statistics are used. For example, there are 'quantitative' and 'qualitative variables'. Qualitative variables use nominal scale measurement; racial origin, ethnic origin, religious affiliation or sex are qualitative variables. Quantitative variables use either ordinal or metric scales. There is also often a distinction between conceptual (or nominal) and operational variables.

1.4 Levels of measurement

Measurement can be performed at four levels, which among other things vary with regard to the degree to which they match the characteristics of the real-number system. The four levels of measurement and four corresponding scales are: the nominal, the ordinal, the interval and the ratio level. Nominal-level measurement has the lowest and ratio-level measurement the highest match with the real-number system.

Box 4.3

Nominal-level measurement:

- Involves nominal categories and is essentially a qualitative and a non-mathematical measurement; it actually names and classifies data into categories.
- Does not have a zero point.
- Cannot be ordered in a continuum of low–high.
- Produces nominal or categorical data.
- Assumes no equal units of measurement.
- Assumes the principle of equivalence: all units of a particular group are taken to be the same.

1.4.1 Nominal-level measurement

This is the simplest, the lowest and the most primitive type of measurement. At this level, measurement involves classification of events into categories that must be distinct, unidi-

mensional, mutually exclusive and exhaustive; the resulting scales are 'naming' scales. Such a measure indicates that there is a difference between the categories considered.

Such differences refer to nature but not to magnitude. Thus, dormitory No. 10 is not twice as large as dormitory No. 5. In a similar fashion, numbers assigned to categories have no mathematical meaning, are used only for identification and cannot be added, subtracted, multiplied, divided or otherwise manipulated mathematically. Classifying the respondents in categories such as male–female, black–white, young–old, single, married, cohabiting, separated, divorced, remarried or widowed, or Catholic, Protestant, Anglican or Orthodox is based on nominal measurement. Classifying respondents according to their place of birth, religious affiliation, political affiliation, car type and place of residence are additional examples. Further examples of nominal measurement are: nationality, type of shoes, skin colour, type of music and brands of drinks.

It must be noted that only statistical measures designed for nominal measurement can be employed in this context.

1.4.2 Ordinal-level measurement

Measurement at the ordinal level involves not only categorizing elements into groups but also ordering data and ranking variables in a continuum ranging according to magnitude, that is, from the lowest to the highest point (transitive relationship). Here, numbers offer more information since they not only indicate differences between categories but also rank them; however, they do not allow mathematical operations such as addition or subtraction.

Box 4.4

Ordinal measurement:

- refers to ranks based on a clear order of magnitude of low and high signifying that some elements have more value than others
- assigns numbers actual mathematical meaning as well as identification properties
- is essentially a quantitative measurement
- shows a relative order of magnitude.

With regard to the last point, order of magnitude allows categories to be ranked (who is first, second, last) but does not indicate the amount of difference between the groups (how much above or below a certain category neighbouring categories are). So the difference between the first and second may be different from that between the sixth and seventh categories. The intervals are not necessarily equal.

Examples of such forms of continuum employed in ordinal measurement are: status (low, middle, high); size (smallest, small, big, biggest); quality (poor, good, very good, excellent); class (low, middle, high); achievement (poor, moderate, high); income (low, middle, high). Ranking occupations is another example.

1.4.3 Interval-level measurement

This level of measurement, as well as demonstrating the properties of ordinal-level measurement, provides information about the distance between the values, and contains equal

intervals, ordering subjects into them. This method allows the researcher to assess differences between respondents and to obtain more detailed information about the research topic.

Box 4.5

Interval-level measurement:

- includes equal units
- is essentially a quantitative measurement
- facilitates differentiation and classification
- incorporates ordering of subjects
- specifies the numerical distance between the categories.

Interval-level measurement allows the researcher, first, to establish whether two values are the same or different (as in nominal measurement), second, to determine whether the one is greater or smaller than the other (as in ordinal measurement), and third, to ascertain the degree of difference between them. Nevertheless, it does not have a true zero point, and if a zero is used it is set arbitrarily, is done so for convenience and does not mean absence of the variable.

For example, if the IQ of two students is 105 and 125 respectively, in nominal terms this means that they have a different IQ; in ordinal terms that the first student has a lower IQ than the second; and in interval terms, that the IQ of the second student is 20 points higher than that of the first student, but not, say, one-fifth greater than the other student.

In mathematical terms, at this level numbers assigned to categories are used to count and rank, but can also be added to and subtracted from each other. This indicates that interval-level measurement is superior to the other two. However, given that there is no true zero, they cannot be multiplied or divided. Statistical measures for nominal, rank and interval data can be used. Examples of this type of measurement are degrees of temperature, calendar time (day, week, month), attitude scales and IQ scores.

1.4.4 Ratio-level measurement

Measurement at this level includes all the attributes of the other three forms, plus the option of an absolute true zero (0) as its lowest value, which in essence indicates the absence of the variable in question. Simply, it is an interval-level measurement with an added true zero. Hence all attributes of interval-level measurement also apply here. Ratio-level measurement allows the researcher to make statements about proportions and ratios, that is, to relate one value to another. For instance, comparing the speed of two students' response to a stimulus – say, 10 seconds and 20 seconds – allows the researcher to conclude that the first is twice as fast as the second.

In social sciences this level of measurement is employed mainly when measuring demographic variables; however, it is considered inappropriate for measuring attitudes and opinions. This is because a zero (0) option in an attitude scale means no attitude, or no opinion, which is misleading, because in the context of a questionnaire even having 'no opinion' is in itself an opinion.

Table 4.1 Levels of measurement: A Summary

Criteria	Nominal	Ordinal	Interval	Ratio
Properties of measurement	Naming	Naming and ranking	Naming, ranking and equal intervals	Naming, ranking equal intervals and zero point
Nature of measurement	Categorical	Ranking	Scoring	Scoring
Mathematical functions	None	None	Addition and subtraction	All four functions
Relevant statistical tests	Lambda test χ^2 test	Spearman's ρ M-W U test Sign test	Pearson's r t-test; ANOVA	Pearson's r t-test; ANOVA
Nature of underlying construct	Discrete	Discrete or continuous	Continuous	Continuous
Examples	Marital status, gender, race, residence, ethnicity	Income, status, achievement, social class, size	Temperature, calendar time, IQ scores, attitude scales	Length, weight, distance, number of children, age
Typical answers to questions	Male, Female Single, married, Irish.	Always; often; sometimes; never.	Scores Likert scales Degrees	Years Kilograms Kilometres

In mathematical terms, numbers arrived at through ratio ordering indicate counting as well as ranking, and can also be added, subtracted, multiplied or divided. Examples of this type of measurement are those given for the interval level above, with the addition of a 0 point in the continuum. Other examples could, for instance, come from the following areas: number of family members, weight, length, distance, number of books that subjects own, reaction time and number of products produced per hour.

1.4.5 Measuring variables

Variables are not measured at one specific level only. Whether a variable will be measured one way or another depends very much on how it is conceptualized and on what type of indicators have been used during measurement. The same variable can be measured in a variety of ways. Age, for instance, can be measured nominally, if it is defined in broad and discrete categories, such as infancy, adolescence, adulthood, middle age and old age; or as young and old. It can be measured also at the ordinal level, when respondents are ranked according to age from the oldest to the youngest.

Age can also be measured at the interval level, given that units are equal, and that we can determine how many units of difference there are between age levels. Interval-level measurement tells us not only whose age is higher (as in ordinal-level measurement) but

also how much higher it is. Age can, finally, be measured at the ratio level, since it has an absolute (non-arbitrary) zero. One cannot be younger than 0; and a 20-year-old person is twice as old as a 10-year-old person.

Despite the degree of freedom researchers enjoy when measuring variables, there is a rule of thumb according to which variables are measured at the highest level possible. Overall, discrete variables are measured at the nominal or ordinal level, and continuous variables at the interval or ratio level.

Box 4.6

Arbitrary and true zeros

The use of true zero as the distinguishing characteristic of ratio scales has caused some confusion. This is due to the fact that it often is difficult to distinguish true zeros from arbitrary zeros. Zeros are not always 'true'. True zeros are meaningful; arbitrary zeros are not. For instance, when we measure temperature, a zero degree reading does not mean no temperature at all! And in measuring attitudes, a zero does not mean no attitude at all (having no opinion on an issue is an opinion!). These zeros are not true zeros, they are arbitrary zeros. However, when measuring income, number of cars, or number of children, a zero indicates no presence of these criteria: it means no children, no income, no cars. These are true zeros; and only measurement using these true zeros can be conducted at the ratio level.

1.4.6 Putting it together

All levels of measurement are effective and useful in their own context and in terms of the purpose for which they have been developed. However, nominal-level measures are the least precise, followed by ordinal-level measures, and then by interval-level measures, with ratio-level measures offering the highest degree of precision. Measuring at the interval and ratio level has many advantages, but not all variables can be measured at these levels (Wang and Mahoney, 1991).

Measurement is a very important and relatively complicated process, but it is associated with many problems and errors. It is limited by the nature of our social world and the variables in question, the perceptions of researchers and their personal bias. Its value depends on the accuracy of the instruments used and the model of operationalization employed.

2 Replication

Replication is a systematic repetition of a baseline study employing different respondents, location, time or method with the purpose to re-examine research topic and establish whether the findings of the original study can be generalized (King, 2003; Freeze, 2007; Gleditsch, 2003). There are many types of replicative studies varying according to four basic research features such as method, participants, time and place (Bahr et al., 1983). Beyond this, replication is thought to provide verification and disconfirmation functions for the scholarly fields (Muma, 1993).

Although the research procedures of replication studies do not differ from those employed in other contexts, there are some issues that require special consideration. Bahr et al. (1983: 243) referring to community research lists five guidelines that are thought to be considered appropriate in such situations; these regulations are listed below.

(a) clearly establish the baseline to be replicated;
(b) resist the temptation to expand the replication to include topics absent in the baseline study;
(c) where extensions are justified, try to include longitudinal linkages using techniques of retrospective or life-history reconstruction;
(d control the urge to update the data collection instruments; and
(e) pay close attention to the interplay between community-level variables and data on individual attitudes and experiences.

Replication is applied in quantitative research, where studies are generally conducted so that they can be repeated by other researchers, in order to facilitate validity checking and more comparisons. This principle is supposed to guarantee the absence of subjective influence by the researcher as well as full objectivity in the procedure. The results here are expected to reflect the views of the respondents fully, so that the same outcomes are produced each time the study is repeated. Apart from this, researchers who conduct the same studies in the same or in a different context should be in a position to use the research instruments of the previous study without difficulty, so that the full study can be replicated. This permits valid comparisons and more legitimate generalizations.

Although replication is a common procedure in quantitative research, in qualitative studies it is seen differently. In general, most qualitative researchers do not employ replication, seeing it as irrelevant and impossible. The nature of qualitative research – the lack of strong interest in representativeness, the use of ad hoc sampling arrangements, and the focus on the views of the subjects rather than of the whole community – places replication outside the realms of this type of research. The common view is that, qualitative research and replication are incompatible.

3 Scales and indexes

3.1 Introduction

Scales are techniques employed by social scientists in a variety of contexts, particularly in the area of attitude measurement. They are included in surveys and consist of a number of items (statements or questions) and a set of quantified response categories. Each item is chosen so that people with different points of view about it react to it in a different way. Scales are employed because they offer (see Benini, 2000):

- *High coverage*. Scales allow a complete coverage of all significant aspects of the concept.

- *High precision and reliability*. Scales allow a high degree of precision and reliability.
- *High comparability*. The use of scales permits detailed and accurate comparisons between sets of data.
- *Simplicity*. Scales help to simplify collection and analysis of the data.

Scales are constructed in a series of complex steps and then statistically tested. Construction and statistical testing are very involving and demanding tasks; they are therefore not easily accessible to the novice, and are certainly beyond the scope of this text. Even experienced researchers prefer to employ already existing and well-tested scales rather than develop new ones. It is worth referring however to a few basic guidelines of scale construction, as handed down to us by two experts Edwards (1957), and Likert (1932). These are shown in Box 4.7.

Box 4.7

Guidelines for scale construction

- Language must be simple, clear and direct.
- Items must be brief (up to 20 words) and contain one issue only.
- Complex sentences must be avoided.
- Items referring to past events and factual items must be avoided.
- Ambiguous and irrelevant items must be avoided.
- Items that may be accepted or rejected by all respondents must be avoided.
- Words such as all, always, no one, never, only, exactly, almost should be avoided.
- Use of professional jargon and double negations should be avoided.
- Response categories must be mutually exclusive, exhaustive and unidimensional (i.e. measuring one single construct).

3.2 Examples of scales

One type of scales is Likert scales which present items in a continuum that covers the whole range of possible responses, allowing respondents to choose the answer that fits their opinion. The following is an example of the type of questions employed in Likert scales.

> Qu. 57. Do you agree that gay marriage is as good as heterosexual marriage? (Please circle the number in front of the answer of your choice).
> 1. Strongly agree
> 2. Agree
> 3. Undecided
> 4. Disagree
> 5. Strongly disagree

Another example is the Bogardus Social Distance Scale, which helps to test how close people allow others, for example strangers, to come to them. The content of this scale is shown in Box 4.8.

Box 4.8

Bogardus Social Distance Scale

Would you perceive an asylum seeker to be as close to you as a(n)

- ☐ relative
- ☐ personal friend
- ☐ neighbour
- ☐ colleague at work
- ☐ speaking acquaintance only
- ☐ unwanted visitor to the country
- ☐ person to be kept out of the country?

(Please tick the box in front of the answer of your choice)

Over the years, researchers have constructed an armoury of scales that are used to test topics of interest. These scales are made available to researchers for a small charge to use when studying the relevant topic. There are books of several volumes that contain such scales, details about their creators, their fields of application and other useful information.

3.3 Indexes

An index is a measure containing a combination of items, the values of which are summed up to provide a numerical score. Indexes are used to describe and measure global concepts accurately by considering a number of specific and representative aspects of the concepts. They represent a summary figure and a composite measure in which each item measures one element of the concept, and provides information on this element or part.

An example is the *Quality of Life* index for the city of Vienna. Such an index may include the following items: employment opportunities, recreation opportunities, weather, pollution level, medical services, educational opportunities, childcare services, safety, crime rate and racial problems. These items will be transformed into questions/statements and the index presented for evaluation. Each question will be scored and the total will present a single measure.

The items of an index can be given the same weight (unweighted index) or different values (weighted index). The latter option is taken when, for instance, some index items are thought to be more important than others. In the example given above, employment rates, safety and crime rates may be considered more important for the quality of life in a country town than the weather or childcare, and may, therefore, be given a higher value (and a higher score) than the other items. In other cases, the unweighted index is employed.

Indexes are useful measures and can be employed in every aspect of life, such as the economy, politics, education, social life, teaching and religious observance. They are constructed using theoretical principles or mathematical formulae. In either case, they are compound measures, and as such they do not differ greatly from scales (Pfeifer, 2000).

4 Validity

4.1 Validity in quantitative research

Validity is the property of a research instrument that measures its relevance, precision and accuracy. Validity tells the researcher whether an instrument measures what it is supposed to measure, and whether this measurement is accurate and precise. Hence, it is a measure of the quality of the process of measurement, and one that reflects the essential value of a study, and which is accepted, respected, and indeed expected by the researchers and users of research.

Box 4.9

What is validity?

- a measure of precision, accuracy and relevance
- it reflects the quality of indicators and instruments
- it refers to the ability to produce findings that are in agreement with theoretical or conceptual values
- it answers the question: Do the instruments/indicators measure what they are supposed to measure?

In general, a measure is expected to be relevant, accurate and precise.

- *Relevance*: An instrument is considered to have absolute validity when it measures what it is supposed to measure and nothing else. If a researcher wanted to know the distance between two cities, kilometres or miles would be a relevant instrument. Likewise, a scale of kilos is a relevant instrument when measuring a person's weight, but is not relevant if used to estimate the person's intelligence.
- *Accuracy*: Validity also entails a degree of accuracy. Accuracy refers to the ability to identify the true value of the item in question. For instance, if you step on your bathroom scales (which measure whole kilos only), and you obtain a reading of 70 kilos (which is your real weight), the scale is accurate. However, if the reading were 68 kilos, the scale would have been inaccurate (and hence invalid).
- *Precision*: Validity requires also that a measure is precise. Precision implies accuracy, but in addition it requires that measurements employ the smallest possible measure. For instance, for a dietician who wants to measure the weekly weight gain or loss of a patient undergoing a special medical treatment, scales that read whole kilos only (68, 69, 70 kilos, and so on) are not precise enough. They are required to read fractions of a kilo.

4.2 Testing validity

In quantitative research, there are two ways of checking the validity of an instrument; these are empirical validation and theoretical validation. In this context, tests of internal

and external validity are employed. In the former, the validity of a measure is checked against empirical evidence. In the latter, the validity of an instrument is ascertained through theoretical or conceptual constructs. In both cases, validity is claimed if the test results are acceptable.

4.2.1 Empirical validation

Empirical validation tests *pragmatic* or *criterion validity*. If an instrument has, for instance, produced results indicating that students involved in student union activities do better in their exams, and if this is supported by available data, the instrument in question has pragmatic validity. Again, validity here is assumed if the findings are supported by already existing empirical evidence. In this case the validity is *concurrent validity*.

If new findings support the predictions of the measure in question, this measure is said to be valid. For example, if a study found that an eventual introduction of advanced statistics into the social sciences degree would result in a significant drop-out of older students, and if meanwhile this prediction is supported by new findings, the measure has validity. This is known as *predictive validity*.

4.2.2 Theoretical validation

Theoretical or conceptual validation is employed when empirical confirmation of validity is difficult or impossible. A measure is taken to have theoretical validity if its findings comply with the theoretical principles of the discipline, that is, if they do not contradict already established rules of the discipline. There are several types of theoretical validity.

Face validity

An instrument has face validity if, 'on the face of it', it measures what it is expected to measure. For example, a questionnaire aimed at studying sex discrimination has face validity if its questions refer to discrimination experienced by people because they are male or female. The standards of judgement here are based on general theoretical standards and principles, and on what other researchers consider to be the case.

It should be noted that when there are no common standards and principles, and when there is disagreement as to what is generally right to expect, this instrument has no face validity. An instrument employed within a study aiming to establish whether people are religious or not that contains only questions related to the respondents' smoking habits is not valid, because it is theoretically known that smoking is not associated with religiosity.

Content validity

A measure is considered to have content validity if it covers all possible dimensions of the research topic. If a researcher in a study of religiosity employs a questionnaire that contains questions only on 'church attendance', this research instrument has no content validity. This is because it focuses only on religious practice and neglects other parts of the concept, such as religious beliefs and religious commitment.

Construct validity

A measure can claim construct validity if its theoretical construct is valid, in other words, if it measures the constructs it is supposed to measure. Validation concentrates here on the validity of the theoretical construct. For example, if an instrument tests the attitudes of two groups of students known to have different views on the issue in question, and this instrument finds them to be different – that is, it verifies the known difference – this instrument is said to have construct validity.

4.2.3 Internal validity

Internal validity refers to the extent to which the research design impacts on the research outcomes. Internal validity checks ensure that the findings of the research have not been affected by instruments or procedures, and that they are the results of the independent variable. Examples of factors that can threaten internal validity, for example in experimental research, panel studies or trend studies, are given below (see Farber, 2001):

- Unexpected structural changes might occur during the course of the study, subjecting respondents to different conditions.
- Normal developmental changes are to be expected in longitudinal studies where data collection occurs in, say, five-year intervals.
- Diverse methods may be used over the course of the study, subjecting respondents to different research instruments.
- Different sampling procedures may be employed during the course of the study, leading to selection problems.
- There may be diverse personnel in the study, with different levels of different competence, experience, knowledge and attitude.
- Changes or alterations in recording techniques may lead to inconsistent records.

In such cases, the respondents are exposed to factors that can affect the information collected in the study.

4.2.4 External validity

External validity refers to the extent to which research findings can be generalized, and is mostly relevant to explanatory studies. The following are a few examples of how the way of conducting the research can threaten external validity (see Farber, 2001).

- *Testing*. Being chosen to take part in the study can stimulate respondents to become more familiar with the study object and hence become more knowledgeable than the average population.
- *Sampling*. Inadequate or biased selection may lead to unrepresentative samples.
- *Multiple exposure*. Exposure to a variety of research instruments might cause an interaction effect and associated problems.
- *Measures*. Inappropriate measures may produce unrealistic responses.

Regardless of the type of validity, its contribution to research is most important, and researchers make a concerted effort to include relevant tests in their studies.

4.3 Validity in qualitative research

Validity is a methodological practice not only of quantitative but also of qualitative research (Lancy, 1993; Maxwell, 1992; Miles and Huberman, 1994; Steinke, 2000). Qualitative researchers aim to achieve validity, which they consider to be a strength of their research, since it frees data from interference and contamination, control or variable manipulation (LeCompte and Goetz, 1982); this is facilitated in a number of ways, particularly through their orientation towards, and study of, the empirical world (Blumer, 1979a: 49), through construction of appropriate methods of data collection and analysis (Volmerg, 1983: 124) or through specific measures such as communicative, cumulative, ecological or argumentative validity (see Köckeis-Stangl, 1980).

Validity is an integral part of qualitative research (Lancy, 1993; Steinke, 2000; Volcott, 1990) although it often appears under a different name. Some speak of 'credibility', 'trustworthiness' and 'authenticity' instead; others use concepts such as objectivity, reliability, credibility, transferability, confirmability, verification, quality, standards dependability, corroboration, referential adequacy, truth, and honesty; while others employ the original name but qualify it in some way. They speak, for instance, of intersubjective validity, supplemental validity, and paralogic validity, or even trustworthiness, ironic validity (Lather, 1993: 677), paralogic/neo-pragmatic validity (Lyotard, 1984), rhizomatic validity (Derrida, 1976), and sensual validity, or situated validity (Lather, 1993). In some contexts, the quantitative researchers' question 'Does the research instrument study what it is supposed to study' is replaced in qualitative research by the question 'Do the researchers see what they think they see?'

To guarantee validity in their work, qualitative researchers apply a number of measures. These vary from case to case, with some researchers proposing one set of measures and others suggesting another. Some of these types of validation, referred to by a number of writers (see Drew et al., 1996; Lamnek, 1993; Pfeifer, 2000; Terhardt, 1981: 789), are presented below.

- *Cumulative validation*. A study is validated if its findings are supported by other studies. The researcher can compare the various findings and make a judgement about the validity of the studies.
- *Communicative validation*. This form of validation entails the involvement of the participants – by checking accuracy of data, evaluation of project process, change of goals etc. (in the Delphi format), by employing expert external audits, and by using triangulation – in order to achieve a multiple perspective (Kardorff, 2000: 245–6), and to confirm authenticity.
- *Argumentative validation*. This form of validity is established through presentation of the findings in such a way that conclusions can be followed and tested.
- *Ecological validation*. A study is held to be valid if carried out in the natural environment of the subjects, using suitable methods and taking into consideration the life and conditions of the researched.

4.4 Other 'tactics'

In a different manner, Miles and Huberman (1994) suggest 'tactics' for testing or confirming findings which, although not direct forms of validity, have a similar function. Some are similar to those presented above (e.g. cumulative or communicative validation); others are close to the form of validation employed by quantitative researchers. Proposals for an integrated effort to safeguard validity in qualitative research have been offered by some writers (Lincoln and Guba, 1985; Drew et al., 1996: 169–71).

The important point here is that validity is not a criterion of quantitative research but a common basis for most types of research. A view supported by many workers in this area is that investigators do not need to demonstrate validity but rather methodological excellence, that is, research performance in a professional, accurate and systematic manner.

These types of validation are considered to be as effective as those employed in quantitative research. For some writers they are even more effective. Lamnek (1993: 154–9), justifies this point on the ground that, in qualitative research:

- The data are closer to the research field than in quantitative research.
- The collection of information is not determined by research screens and directives.
- The data are closer to reality than in quantitative research.
- The opinions and views of the researched are considered.
- The methods are more open and more flexible than in quantitative research.
- There is a communicative basis that is not available in quantitative research.
- A successive expansion of data is possible.

Observing the developments in the area of validity in social research, it becomes obvious that in qualitative research quality assurance has been given increasingly high priority (see for example Lincoln, 1995; Seale, 1999; Steinke, 1999). Important innovations have been introduced to ensure validity. Some central points in this area have been those described above, but apart from this the notion of an open and clear description of the procedures of data collection and interpretation, the presentation of relevant materials, the reproduction of transcripts, field notes, and even an emphasis on replication have been highly significant (Matt, 2000: 585).

Box 4.10

Validity in ethnographic research

'(1)The researcher should refrain from talking in the field but rather should listen as much as possible. He or she should (2) produce notes that are as exact as possible (3) begin to write early, and in a way (4) which allows readers of his or her notes and reports to see for themselves. This means providing enough data for readers to make their own inferences and follow those of the researcher. The report should be (5) as complete and (6) as candid as possible. (7) The researcher should seek feedback for his or her findings and presentations in the field or from his or her colleagues. (8) Presentations should be characterised by balance between the subjects and (9) by accuracy in writing.'

(Wollcott, 1990: 127–8)

5 Reliability

Reliability refers to the capacity of measurement to produce consistent results. Reliability is equivalent to *consistency*. Thus, a method is reliable if it produces the same results whenever it is repeated, and is not sensitive to the researcher, the research conditions or the respondents. Reliability is also characterized by precision and objectivity (see Box 4.11). As in validity, so in reliability there are two major aspects of interest in this context; these are *internal* reliability and *external* reliability. Internal reliability means consistency of results within the site, and that data are plausible within that site. External reliability refers to consistency and replicability of data across sites.

The purpose of reliability testing is to ensure that the instruments in question are robust and not sensitive to changes of the researcher, the respondent or the research condition. This, apart from implying that the instrument allows replicability, demonstrates that reliability is concerned with objectivity, accuracy, precision, consistency and stability. These criteria are employed as in validity.

Box 4.11

What is reliability?

- a measure of objectivity, stability, consistency, and precision
- a measure of the quality of indicators and instruments
- it refers to the ability to produce the same findings every time the procedure is repeated
- it answers the questions: does the instrument/indicator produce consistent results? Is the instrument free of bias associated with the researcher, the subject or the research conditions?

5.1 Reliability in quantitative research

There are at least three types of reliability, all of which are considered by social researchers. These are:

- *Stability reliability*, relating to reliability across time. Here the question is whether a measure produces reliable findings when it is employed at different points in time.
- *Representative reliability*, which relates to reliability across groups of subjects. The question here is whether the measure is reliable when employed in groups other than the original group of subjects.
- *Equivalence reliability*, which relates to reliability across indicators and to multiple indicators in operationalization procedures. The question here is: Will the measure in question produce consistent results across indicators?

There are also several methods for testing reliability of an instrument. The most common methods are the following:

- *Test-retest method*: The same subjects are tested and retested with the same instrument. If the same results are obtained the instrument is reliable.
- *Split-half method*: Responses to the items of an instrument are divided into two groups (e.g. odd/even questions) and the scores correlated. The type and degree of correlation indicate the degree of reliability of the measurement.
- *Inter-item test and item-scale test*. Inter-item correlations or item-scale correlations indicate the degree of reliability of the instrument.
- *Alternate-form reliability*: Reliability is tested by administering two similar instruments in one session, and is assessed by the degree of correlation between the scores of the two groups.

These tests are regularly used and entail a considerable amount of statistical analysis and interpretation. Instruments are tested before they are put to use, and the results are normally disclosed to the academic community every time these instruments are referred to and their findings published. The advent of computer-based statistical analysis has made this task easier, more accurate and more enjoyable than before, and is being consistently used by quantitative researchers.

A closer analysis of reliability shows that it is related to validity. Actually, reliability without validity is of little use. Even the most reliable instrument is useless if it is not valid. For instance, a scale showing that a student's weight is exactly 65 kg every time the student steps on it is of no value if the student's real weight is known to be 60 kg. Hence, it is useful to measure and interpret reliability results together with validity scores.

5.2 Reliability in qualitative research

Qualitative researchers do consider reliability an important parameter of research but, to adhere to it, they employ methods that are radically different from those employed in quantitative research (see Box 4.12). They use measures of reliability such as increasing the variability of perspectives in research, or setting up a list of possible errors or distortions which they aim to avoid (McCall, 1979). Overall, qualitative researchers strive for rigour but employ different methods to achieve it. In the majority of cases, they avoid the use of the concept 'reliability'; instead they use concepts such as *credibility* and *applicability*, or *auditability*. Objectivity is replaced by: *confirmability* (Guba and Lincoln, 1989); *coherence*, that is, the extent to which methods meet the research goals; *openness*, the degree to which otherwise suitable methods are allowed to be used; and *discourse*, that is, the extent to which researchers are allowed to discuss the researched data and interpret them together and evaluate the consequences of such findings (Bogumil and Immerfall, 1985: 71). References to *trustworthiness, dependability, credibility, transferability* and *confirmability* also seem to be popular (Flick, 1998: 231–2).

Box 4.12

Think critically

Qualitative researchers argue that in their quest to achieve a high reliability, quantitative researchers:

- control the environment
- employ high levels of measurement and standardization
- restrict the researcher-researched relationship
- create artificial situations which are different from those they intend to study
- alienate the researcher from the research environment, which is counterproductive.

What is your opinion?

Overall, the quality of qualitative research is assessed in more general terms than that of quantitative research. It is done by such means as demarcating statements of the subjects and interpretations of the researcher; following procedures that would guarantee that multiple researchers produce comparable results (e.g. through appropriate training); and increasing the documentation of the results (Flick, 1998: 224).

Both aspects of reliability, the internal and the external, are considered in qualitative research. How these dimensions of reliability are addressed in practice varies from case to case. One common view (Flick, 1998: 231–2) proposes the following paths:

- prolonged engagement and persistent observation
- peer review or debriefing
- analysis of negative cases
- checking 'the appropriateness of the terms of reference of interpretations and their assessment'
- member checks (communicative validation)
- external auditing.

In a similar fashion, Drew and associates (1996: 169) suggest that the following steps should be followed if internal reliability is to be achieved:

- Use low inference descriptors.
- Use multiple researchers whenever possible.
- Create a careful audit trail (a detailed record of data that can be used by other scholars to check internal validity).
- Use mechanical recording devices where possible (and with permission).
- Use participant researchers or informants to check the accuracy or congruence of perceptions.

With regard to external reliability, the same authors propose the following five steps:

- Specify the researchers' status or position clearly so that readers know exactly what point of view drove the data collection.

- State the identity of the informants (or what role they play in the natural context) and how and why they were selected (while maintaining confidentiality).
- Delineate the context or set boundaries and characteristics carefully so that the reader can make judgements about similar circumstances or settings.
- Define the analytic constructs that guide the study (describe specific conceptual frameworks used in design and deductive analysis).
- Specify the data collection and analysis procedures meticulously.

Regardless of the significance of these approaches to reliability, a number of writers argue that qualitative research does not provide as high a degree of reliability as quantitative research. However, this view is not shared by others who argue that both models of reliability testing are correct in their contexts. They are different but they serve their particular purposes effectively.

5.3 Validity and reliability

Validity and reliability are both quality measures of research instruments, and although they are quite different in their nature and purpose, some students find it difficult to distinguish between the two. An example may help to clarify the difference. If a male student weighs himself 20 times and every time he receives a reading of 65 kg (which is also his true weight), the scale is both reliable and valid. If all recorded readings were 40 kg, the scale is reliable but not valid. And if he obtains 20 different readings (40 kg, 45 kg, 63 kg etc.), the scale is neither valid nor reliable.

The validity and reliability of a measure are closely interrelated. Nevertheless, the one cannot predict the other. A reliable instrument is not necessarily valid. Even the most reliable instrument can be invalid. As noted above, reliability alone cannot assess the quality of an instrument fully. The validity score is the most important.

Box 4.13

Criteria of validity and reliability

Validity …	**Reliability …**
■ is a measure of the quality of measurement ■ tests the quality of indicators and research instruments ■ measures relevance, precision, and accuracy ■ tests the ability to produce findings that are in agreement with theoretical or conceptual values	■ is a measure of the quality of measurement ■ tests the quality of indicators and research instruments ■ measures objectivity, stability, consistency, and precision ■ tests consistency, i.e. the ability to produce the same findings every time the procedure is repeated
ASKS: Does the instrument measure what it is supposed to measure?	**ASKS**: Does the instrument produce the same results, every time it is employed

6 Objectivity

6.1 The debate

Objectivity is the research principle that requires that all personal values and views of the investigator must be kept out of the research process. The purpose of this is to minimize personal prejudice and bias, and to guarantee that social reality will be presented as it is, and not as the investigator interprets it, imagines it or wants it to be.

Although this is a long-standing principle, the question of whether social inquiry ought to be objective or not has not been answered uniformly by the academic community. Over the years academic views on this issue have been divided, with two lines of thought occupying the two extremes of the argument. The one is known as *value neutrality*, and the other as *normativism*. The former was the position of quantitative researchers, and the latter the stance of qualitative and other researchers.

Box 4.14

What is objectivity?

Objectivity is the empiricist doctrine that the research process and design must be free of personal bias and prejudice. It rests on the belief that facts and values should be kept apart, and that research should focus on what really is and not on what ought to be. Objectivity reflects value neutrality.

6.1.1 Value neutrality

The notion of value neutrality reflects the requirement that investigators ought to minimize the effects of their own biases. Social researchers are seen as 'technicians' or consultants and not as reformers; or better, as neutral observers and analysts and not as philosophers or moralists. The researcher's personal views and value judgements are to be kept out of research. In a more general context, objectivity subscribes to a number of principles and convictions, three of which are the following:

- Social sciences are value free; their goal is to study what is and not what ought to be. Research should aim to achieve the highest possible degree of objectivity.
- Social scientists should be value free; they should rule out value judgements, subjective views, personal bias and personal convictions.
- Value judgements should be reserved for policy makers, and not for social scientists.

6.1.2 Normativism

Normativism is critical of the value and usefulness of objectivity and proposes that value-neutrality is not justified. More specifically it proposes that (Abercrombie et al., 1988; Fay, 1980; Wadsworth, 1984):

- Objectivity is unattainable, unnecessary and undesirable.
- Social science is normative; its goal is to study what ought to be and not only what is.

- People's orientation is based on and constructed with values, that direct thinking and action, and cannot be neutralized, isolated or ignored.
- Being normative and disclosing the inevitable bias or personal beliefs is less dangerous than pretending to be value free.

6.2 Objectivity in quantitative research

Quantitative research accepts and supports objectivity, and considers it one of the most important principles of social inquiry. This is obvious, given that quantitative research operates within an objectivist epistemology. For quantitative researchers, objectivity is regarded as a virtue that every social researcher should try to achieve. Although they are aware that it is difficult to reach a high degree of objectivity, the endeavour to reach the highest possible level is taken for granted.

Box 4.15

The logic of objectivity

- The purpose of research is to discover objective truths.
- Objective truths can be verified only when contrasted with objective reality.
- The task of verification is completed by researchers.
- It is important that verification is conducted objectively and focuses on objective reality.
- The subjective views and personal values of the researcher can only distort the process of verification, and cannot enhance the objectivity of truths.
- Hence, subjectivity distorts the process of discovery of objective truths and must be excluded from research.

The logic of objectivity rests here on the notion that research is expected to capture and present reality as it is and not as it is interpreted, imagined or wanted to be by the investigator. If one is interested in why female students get higher grades in science courses, one expects the researcher to set aside personal views and try to establish the real reasons. If one disregards objectivity and constructs a research design that is biased by personal beliefs and ideologies, the sampling procedures will be affected as much as the choice of methods of data collection and analysis and, obviously, the findings and conclusions. The results will be flawed and will not correspond to reality. Five such researchers may produce five different sets of answers to the same question. Objectivity serves to restrict the influence of such personal biases and prejudices in the research process and to allow reality to come forward as it is, without manipulation.

6.3 Objectivity in qualitative research

Qualitative researchers fundamentally reject the notion of objectivity. Given that qualitative research rests within the parameters of an interpretivist epistemology, this is self-evident. Hence, involving personal views and interpretation in the research process is not only acceptable but advisable, and is considered an advantage.

The logic behind this position is that there is no objective reality to begin with. Hence, it is not possible to capture objective reality in the way that objectivism proposes.

Researchers capture one aspect of reality – their reality – and this is what they can describe and present. Apart from this, value neutrality is considered to be unattainable, unnecessary and undesirable. One cannot consider intrinsic evaluation, feelings, beliefs and standards as insignificant or uninfluential. Social scientists ought to have a standpoint on social issues, and they must produce value judgements if they wish to solve social problems. Qualitative research encourages intersubjectivity, closeness between the elements of the research, and involvement of the researcher in the whole research process (Becker, 1989; Stergios, 1991).

Box 4.16

Where and how objectivity is practised

Research parameters	How is objectivity practised?
Reality	By perceiving reality in objective terms, as an objective reality, which must be reproduced as it is, without distortions of any kind.
The researcher	By respecting value neutrality, i.e. be free of personal values, bias, and prejudice, and to study 'what is' and not 'what ought to be'.
Research topic	By conceptualizing the research topic in an objective manner; selection of indicators and definitions should be free of personal bias, experiences and views.
Methodology	By choosing the appropriate methodology in a process free of personal preferences, ideologies or bias.
Design construction	By constructing the design using professional standards, avoiding personal bias, and assuring compliance to ethical standards throughout the study.
Sampling	By choosing the sample by complying with research standards and practices, excluding personal bias and preference.
Data collection	By choosing relevant methods and by gathering data using professional standards, and by focusing on 'what is' and not on 'what ought to be'.
Administration	By guiding arrangements towards facilitating the completion of the study, using fair and professionally acceptable standards and not personal or ideological preferences.
Data analysis	By describing relevant methods clearly and by conducting analysis in a manner reflecting professionalism, and avoiding personal biases and preferences.
Interpretation	By including clear and detailed justification of conclusions, revealing personal perceptions and interpretations of data and reality for verification.
Reporting	By constructing the report in a manner that outlines clearly data and personal interpretations of the writer. Personal views etc. should be made clear.

Further, adherence to objectivity is taken to support the status quo. Hence believing in and trusting what is objective means believing in what people learned to consider objective, and failing to challenge it. Objectivity legitimizes beliefs and practices which

people take for granted. Women for instance learned to take the world of males as objective, and came to accept it (see Keller, 1985; Reinharz, 1992a). It follows that abandoning objectivity will free thinking from 'inappropriate' constraints and 'unconscious' mythologies. Through disengaging thinking from notions of what is generally considered to be objective, one can capture reality more effectively.

This does not mean that qualitative researchers abandon the academic requirement to be responsible, truthful and transparent in their work. It is often argued that in qualitative research objectivity is 'emergent'; that is, it evolves out of the subjectivity of the parties of interaction. Qualitative research uses an intersubjective concept of objectivity in as far as its aim is to break away from subjectivity through generalization.

Box 4.17

Think critically

- If adherence to objectivity is taken to support the status quo, isn't it logical to argue that – vice versa – adherence to status quo supports objectivity?
- If objectivity legitimizes beliefs and practices, isn't it logical to argue that – vice versa – the establishment and acceptance of beliefs and practices legitimize objectivity?
- If believing in and trusting what is objective means believing in what people have learned to consider correct and right, and failing to challenge it, isn't it logical to argue that – vice versa – believing in what people have learned to consider correct and right, and failing to challenge it makes them believe and trust it as objective?
- If women learned to take the world of males as objective, and came to accept it, failing to challenge it, isn't it logical to argue that – vice versa – accepting and failing to challenge their new status makes them learn to take the post-male world as objective?

Some researchers, instead of speaking of objectivity, speak of transparency: openly stating the course and elements of the research process, and letting others judge its quality. Others (Rorty, 1985: 10; J. K. Smith, 1992a: 101) consider objectivity as equivalent to solidarity, namely the degree of agreement among colleagues and researchers. Simply, objectivity is achieved if colleagues agree with the research process, accept the results and praise the researchers for their achievements.

In this sense, a form of objectivity is accepted and practised, although under different labels and within the parameters of the overarching paradigm. Guba and Lincoln (1989), for instance, confirm that any qualitative study of rigour is expected to contain what they call *truth value, applicability, consistency* and *neutrality*. The latter is another word for objectivity and is expected to provide confirmability. For many qualitative researchers, some form of neutrality is a central element of qualitative research, although the nature and extent of that 'objectivity' are different from that of quantitative researchers.

Finally, in feminist research, some forms of objectivity are accepted and practised. Examples of objectivity are: *dynamic objectivity* (Keller, 1985); *openness*, where all facts are made known to the respondent, *highlighting all contingencies* of representation (Harding, 1991); *democratic discussion*, where all parameters are set in the open, which encourages cooperation among all researchers as well as criticism from all points of view, and is based on equality of intellectual authority (Longino, 1990); and *anti-sexism*.

7 Representativeness

Representativeness has always been central to social research where it is aspired to collect data which would produce results to speak for the whole target population. As we shall see later, one of the most important aspects of sampling is to ensure representativeness, and researchers have devised many methods to ensure that social research will meet this requirement. Apart from this, sampling procedures are considered proper if they are representative for the target population.

Representativeness is associated with the principles of the underlying methodology, and hence its meaning, structure, underlying procedure and interpretation depend on the parameters of the nature of the methodology.

Box 4.18

What is representativeness?

Representativeness is a research principle that reflects the capacity of social research to produce findings that are consistent with (representative of) what appears in the target population; this is a property of sampling. The aim of representativeness is to ensure that all groups of the target population are adequately represented in the research sample. The degree of representativeness determines the extent to which the findings of a study can be generalized

7.1 Quantitative research

Representativeness has a central place in, and is one of the aims of, quantitative research. Several procedures have been introduced to ensure that the principles of representativeness are being applied. Most of the relevant procedures deal with the nature of sample selection as well as with sample size and composition. Statistical techniques have been developed to assist with this process. Standard errors, for instance, are calculated, and techniques used that can assist in achieving a sample size that will allow the study to claim representativeness.

The application of representativeness is guided by procedures which provide suitable conditions for constructing a representative study. This entails not only relevant standards and principles but also guides as to how to prevent errors in the design and/or execution of the research, some being accidental errors (caused by uncontrollable mistakes), while others are systematic errors. These are caused by faulty designs that allow a disproportionate representation of some parts of the population in the sample.

7.2 Qualitative research

In qualitative research, representativeness, as we introduced it above, is considered irrelevant and insignificant. First, it is not consistent with the principles of the qualitative paradigm, and second, the size of the sample and the nature of qualitative sampling procedures are not compatible with the claims for representativeness. Nevertheless, this does not mean that qualitative researchers are not interested in representativeness. There are researchers who consider it to be an indispensable element of qualitative research and take precautions to ensure that it is employed in their research (Stergios, 1991).

Such precautions include, for instance, avoidance of 'sampling non-representative informants, e.g. by over-relying on accessible or elite respondents', and of 'generalizing from unrepresentative events or activities', or 'drawing inferences from non-representative processes'. Miles and Huberman (1994: 265) advise that weak non-representative cases should be expanded, and suggest employing the following steps:

- Increase the number of cases.
- Search purposely for contrasting cases.
- Sort the cases systematically and fill out weakly sampled case types.
- Sample randomly, within the total universe of people and phenomena under study.

Although the last suggestion brings qualitative research as near to quantitative research as one can get, it also takes it further away from the very principles of qualitative research. It is also a controversial proposition indeed, which most qualitative researchers will not accept.

8 Generalizability

Literally, generalizability is the ability to generalize something. In social research it means to generalize the findings beyond the boundaries of the research sample. It is the other side of representativeness: high representativeness is associated with high generalizability, and vice versa. Given that researchers always aspire to produce results that can be applicable in populations as large as possible, generalizability becomes a central issue of social research.

Box 4.19

What is generalizability?

In social research, generalizability refers to the capacity of a study to extrapolate the relevance of its findings beyond the boundaries of the sample. In other words it reflects the extent to which a study is able to generalize its findings from the sample to the whole population. Obviously, the higher the generalizability, the higher the value of the study.

There are several types of generalization; *scientific* (*inductive*) generalization and *naturalistic* generalization are two representative examples. The former refers to the extrapolation of the validity of the findings of a study of representative cases to the whole population. Naturalistic generalizations are more diverse and include several variations. *Analytical generalizations*, *exemplar generalizations*, and *case-to-case transfer* (transferability) (Firestone, 1993) are a few examples. These are basically theory-related generalizations. They can also rest on the argument that the typical cases studied are representative of a species, and hence findings concerning these typical cases can be considered applicable within this species. The question here is how one could verify that certain cases are typical before one has sufficient information about the cases.

Box 4.20

Think critically

- Can a qualitative study of 15 divorced women produce generalizable findings?
- Can a qualitative study of 15 divorced women be considered representative?
- Can a quantitative study of 3,500 divorced women selected using area sampling produce generalizable findings?
- Can a quantitative study of 3,500 divorced women selected using area sampling be considered representative?
- Are the justifications offered by qualitative researchers regarding representativeness and generalizability acceptable and convincing?
- Are the justifications offered by quantitative researchers regarding representativeness and generalizability acceptable and convincing?

Which research model offers the best approach to generalizability and representativeness? The qualitative model, the quantitative model, or both models?

Quantitative researchers mostly employ scientific or inductive generalizations, using probability theory to construct the samples, and statistical methods and techniques to estimate the level of generalizability. Qualitative researchers employ naturalistic generalizations, and in order to achieve this, they use typical cases as their sample. The manner in which generalizability is justified within qualitative research varies considerably. Some writers employ *multi-site research* (i.e. sample triangulation) to achieve generalizability. Obviously, choosing typical subjects from a variety of backgrounds ensures representativeness and hence generalizations.

Other writers explain generalizability by means of criteria such as *conceptual power, fittingness* and *comparability* (e.g. Schofield, 1993). Fittingness relates to the degree of fit between the case studied and the case to which researchers want to generalize their findings (ibid.: 211). Focusing on this point is a way of 'claiming' generalizability. Comparability refers to the assessment of the relevance of the findings by checking similarities and differences between the studied site and the sites to which the findings are to be generalized. The question here is whether the site studied is *typical* of the others. Multi-site research entails the study of the same issue in more than one site, thus establishing the relative representativeness of the sites.

Main points

- The principles of research are precision in measurement, accuracy, validity, reliability, objectivity, replication, representativeness and generalizability.
- Some form and degree of measurement is included in all types of research.
- There are four levels of measurement: nominal, ordinal, interval and ratio levels.
- Variables are measured at the highest level possible.
- Validity is the ability to produce accurate results and to measure what is supposed to be measured. It is an attribute of quantitative and qualitative research.

- Quantitative research employs many types of validation: for example, empirical, theoretical, face, content and construct validity.
- In qualitative research, validation takes the form of cumulative, communicative, argumentative or ecological validation.
- Reliability is the capacity to produce consistent results. It is an attribute of both quantitative and qualitative research.
- Objectivity excludes personal values from research, and is valued in quantitative research.
- Representativeness is an important characteristic of social research that is closely adhered to in both quantitative and qualitative research.

? Where to from here?

Before you leave this chapter, visit the companion website for the fourth edition of *Social Research* at http://www.palgrave.com/sociology/sarantakos4e to review the main concepts introduced in this chapter and to test yourself on the major issues discussed.

🕮 Further reading

Bahr, H. M., Caplow, T. and Chadwick, B. A. (1983) 'Middletown III: Problems of Replication, Longitudinal Measurement, and Triangulation.' *Annual Review of Sociology,* 9: 243–64

Bowling, A. (1997) *Measuring Health: A Review of Quality of Life Measurement Scales*. Buckingham: Open University Press.

Crano, W. D. and Brewer, M. B. (2002) *Principles and Methods of Social Research*. Mahwah, N.J.: Lawrence Erlbaum Ass. Inc.

Freese, J. (2007) 'Replication Standards for Quantitative Social Science Why Not Sociology?' *Sociological Methods & Research*, 36(2): 153–72.

Golafshani, N. (2003) 'Understanding Reliability and Validity in qualitative research.' *The Qualitative Report,* 8(4):597–607.

Lather, P. (1993) 'Fertile Obsession: Validity after Post-Structuralism'. *Sociological Quarterly*, 35: 673–93.

Lloyd, E. (1995b) 'Objectivity and the Double Standard for Feminist Epistemologies'. *Synthese*, 104: 351–81.

Maxwell, J. A. (1992) 'Understanding and Validity in Qualitative Research'. *Harvard Educational Review*, 62(3): 279–300.

Miller, D. C. and Salkind, N. J. (2002) *Handbook of Research Design and Social Measurement*. London: Sage.

Wang, M. and Mahoney, B. (1991) 'Scales and Measurement Revisited'. *Health Values*, 15: 52–6.