

AS Sociology For AQA
[2nd Edition]

Unit 2:
Sociological Methods



Chris. Livesey and Tony Lawson



Unit 2: Sociological Methods

Contents

- | | |
|--|-----------|
| 1. The distinction between primary and secondary data, and between quantitative and qualitative data. | 2 |
| 2. Sources of data, including questionnaires, interviews, participant and non-participant observation, experiments, documents, and official statistics; the strengths and limitations of these sources. | 9 |
| 3. The relationship between positivism, interpretivism and sociological methods; the nature of social facts. | 32 |
| 4. Quantitative and qualitative methods of research; their strengths and limitations; research design. | 45 |
| 5. The theoretical, practical and ethical considerations influencing choice of topic, choice of method(s) and the conduct of research. | 63 |

1. The distinction between primary and secondary data, and between quantitative and qualitative data.

Sociological Research: Introduction

The idea we floated in the Introductory Chapter (and which is implicit throughout the whole textbook) is that the knowledge produced by sociologists is both *different to* - and has greater *validity* than - “common sense” or “everyday” knowledge. This claim is based on the idea that sociological knowledge is not just the expression of someone’s *opinion*; rather it represents data that has been *systematically* collected, analysed and interpreted through a research process. The key difference between sociological and common-sense knowledge, therefore, is that with the former some attempt has been made to verify (or check) its accuracy. If this is a crucial difference between the two types of knowledge it follows that we need to explore the sociological research process in more detail and, in this respect, we can initially note that it has two main components:

1. Research Methods: These are the various ways sociologists collect data – some you may be familiar with (such as questionnaires) and others you may never have heard of before (such as Creative Visual methods).



2. Methodology: The ability to collect data systematically, although a *necessary* part of the research process, isn’t the full story. The decision to use certain methods (but not others) or collect certain types of data (but not others) is surrounded by **beliefs** – and these involve, for example, ideas about the nature of the social world, the ability of different research methods to study that world and the capacity for different types of data to capture and accurately reflect that world. In other words, sociological research and data collection is *always* surrounded by

methodological questions that have to be posed and answered by the researcher.

Although the distinction between **methods** (*what* you do) and **methodology** (*why* you do it) is in some ways a forced or artificial one – collecting data (using a research method) would be a fairly pointless exercise if the reasons for such collection (methodology) weren’t clear to us - it is nonetheless a useful one for our current purpose, for a couple of reasons:

Firstly, it allows us to ease our way into the study of the sociological research process by looking. Initially, at some basic concepts (such as the distinction between *primary* and *secondary* data) and then by outlining and evaluating a range of possible data sources and research methods.

Secondly, once we’ve familiarised ourselves with these ideas we can move up a gear to consider a range of *methodological questions* (such as outlining two different types of research methodology – **Positivism** and **Interpretivism**) and looking at the research process more systematically (in terms of different explanations about the organisation of sociological research). In the final section of this Chapter we can examine a range of *practical, theoretical* and *ethical* considerations that surround the research process as a whole – from choosing a topic, through choice of method to the overall conduct of the research process.

Sociological Methods: Observations

In this Section we can introduce and examine some “basic research concepts”, the general understanding of which will help you come to terms with the various aspects of the research process introduced and examined throughout the remainder of the Chapter. In this respect we can begin to think about the information sociologists collect as belonging to one of two basic types:

1. Primary data involves information collected *personally* by a sociologist - who, therefore, knows exactly how the data was collected, by whom and for what purpose (you don’t, for example, have to trust other people collected their data accurately). As we will see, sociologists use a range of research methods (such as questionnaires, interviews and observational studies) as *sources* of primary data.

2. Secondary data involves information *not* personally collected by the researcher, but used by them in their

research. Sources of secondary data include newspaper articles, books, magazines, personal documents (such as letters and diaries), official documents (such as government reports and statistics) and even the research of other sociologists. In turn, each of the above can be further subdivided into either of two types:

1. Quantitative data represents an attempt to *quantify behaviour* - to express it *statistically* or *numerically*. For example, we could count the number of people in the UK who wear glasses (which is probably not that useful unless you happen to manufacture spectacles) or the number of people who commit crimes each year (which is probably a little more useful, in the general scheme of things). Quantitative data is usually expressed in one of three main ways. As a:

- **Number:** For example, the number of people who live in poverty.
- **Percentage** (the number of people **per 100** in a population). For example, 30% of voters in Britain regularly vote Conservative.
- **Rate** (sociologically, this is defined as the number of people **per 1000** in a population). For example, if the birth rate in Britain is 1 (it's not, by the way) this means for every 1000 people in a population, one baby is born each year.

Although "raw numbers" can be useful (for example, knowing the number of children who will be starting school in 10 years time allows the government to plan for the number of people who will need to be trained to teach them), data is often expressed as a **rate** or **percentage** because it allows:

Comparisons *between* and *within* groups and societies. For example, when comparing levels of unemployment between Britain and America, expressing unemployment as a simple or *raw* number wouldn't tell us very much, since the population of America is roughly 5 times that of Britain. Expressing unemployment as a *percentage* or *rate* allows us to compare "**like with like**", in the sense we're taking into account the fact one society has substantially more people than the other (so we might expect the larger society to, *numerically*, have more people unemployed - even though their unemployment *rates* might be broadly similar).

Module Link

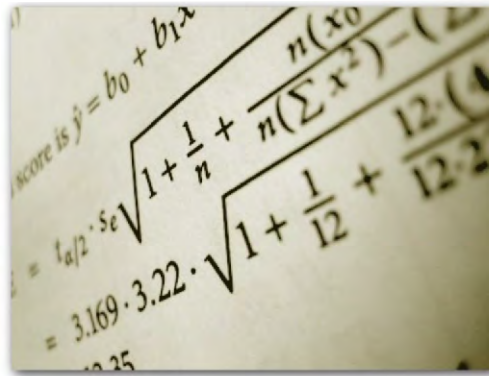
Education

Statistical data about a range of things – from gender differences in the choice of A-level subjects / degrees, through educational achievement to the ethnic backgrounds of those excluded from school – are routinely collected and produced by the government.

2. Qualitative data, on the other hand, tries to capture something of the *quality* of people's behaviour (what they *feel*, for example, about a sociologist asking them if they wear glasses). Such data, therefore, says

something about how people *experience* the social world; it's also used to understand the:

Meanings people give to their own behaviour and that of others. **Boyle** (1977), for example, studied a juvenile gang from the viewpoint of its members while **Goffman** (1961) tried to understand the experiences of patients in an American mental institution. Both, in their different ways, were trying to capture and express the *quality* of people's behaviour, albeit in different situations.



An example of some Very Complicated Statistics. Be Afraid. Be Very Afraid...

Although these distinctions are important – and necessary to understand - research methods, as we've suggested don't simply involve thinking about **data types** (qualitative and quantitative) and **sources** (primary and secondary); we also need to think about our **reasons** for choosing particular types and sources in our research - something that involves considering sociological **methodology**.

Methodological Concepts

For the moment there are *four* main methodological concepts we need to initially outline:

1. Data Reliability relates to the "nuts-and-bolts" of actually doing research; in other words, it mainly refers to the methods of data collection we use (such as interviews) and, more specifically, to the *consistency* of the data we collect. Data reliability is important because it suggests we can *check* the data we get from our research by **repeating** that research to see if we get the same, or very similar results (we may have to allow for possible individual changes over time). If a researcher, for example, needs to know someone's age this is something that will change over time, depending on the gap between two surveys. In general, therefore, we can say data is *reliable* if similar results are gained by different researchers (or the same researcher at different times) asking the same questions to similar people.

True
True
True
True
True
True
True
True

The ability to replicate research results is a good indication of data reliability.

A simple (in the sense of not being particularly realistic – it's just for explanatory purposes - example here might be a researcher trying to *cross-check* the reliability of a response within a questionnaire by asking the same question in a different way:

- How old are you?
- When were you born?

If they get two different answers, it's likely the data is *unreliable*.

2. Data Validity refers to the extent to which data gives an accurate measurement or description of whatever it is the researcher is trying to measure or describe. Data, it could be argued, is only useful if it actually *measures* or *describes* what it claims to be measuring or describing. For example, if we were interested in the extent of crime in our society, we could use *official crime statistics* (a *secondary* data source published by the government). We would need to be aware, however, that the validity of these statistics may be limited since they only record *reported* crimes - and people may not report the fact they have been a victim (for many possible reasons - such as a fear of reprisal from the criminal or the belief the police will not be able to trace the perpetrator, to name but two).

Module Link

Crime and Deviance

Notwithstanding the fact that we have to be careful about the validity of official crime statistics they still represent an important source of data about crime in our society. We should also note that not all crimes are underreported in our society, Car theft statistics, for example, have a high level of validity because insurance companies insist on the theft being reported to the police. Murder statistics – for rather different reasons (it's actually quite difficult, so we've been told, to dispose completely of a human body – also tend towards high validity.

This example also raises questions relating to:

3. Representativeness: Whatever type of data we use (primary or secondary, quantitative or qualitative), an important question to always consider is the extent to which the data accurately *represents* what it claims to represent - something we can think about in two basic ways:

Data representativeness refers to the idea that any information we collect through our research is sufficiently comprehensive to accurately represent whatever the research claims to represent. Using the crime statistics example introduced above it can be argued these statistics are *unrepresentative* of all crimes committed in our society; anything we say, therefore, about "crime" in our society on the basis of this data source needs to be *qualified* by saying that some types of criminal behaviour may not be fully represented in the statistics.

Group representativeness refers to the use of **samples** (explained in more detail in Section 4 of this Chapter) in our research. In basic terms, if we're researching a small group (of students, for example) and, on the basis

of this research, want to be able to say something about *all students*, we need to ensure that the characteristics of the group we study (our *sample*) **exactly match** those of the larger group. For example, if the gender distribution of *all students* is in the ratio 1:1 (for every male student there is 1 female student – this isn't the case, but it does illustrate the basic point) then the *same* must be true of our sample if it is to be representative. In other words, we can use one, small, group to *represent* a much larger group - an idea that leads to the related concept of:

4. Generalisation: If data can be *generalised* it means information we collect about a relatively small group (the *sample group*) can be applied to larger groups who share the same general characteristics as the sample. In other words, if the sample group is representative of the larger group anything we discover about the former can be generalised to the latter. The usefulness of these two concepts - representativeness and generalisation - will become clearer when we consider them in more detail in the context of *sampling techniques* (Section 4).

Sociological Methods: Explanations

The different data types we've just identified each have their different **advantages** and **disadvantages**.

Primary Data

The ability to generate this type of data has some clear **advantages** for the sociologist:

Data Control: Because the researcher is responsible for collecting data they have complete control over such areas as how much data is collected, how and from whom it's collected and so forth.

Reliability, validity and representativeness: Simply because you can exercise some measure of control over how data is collected doesn't, of course, guarantee its reliability, validity or representativeness - a badly designed piece of research can be unreliable, invalid and unrepresentative. However, it's much easier for the researcher to consider and control these concepts when they design and carry out the research themselves.

This type of data also has a few potential **disadvantages:**

Resources: Primary data collection can be:

- **Time-consuming** - to design, construct and carry-out. If the group being studied is large and involves something like interviewing each group member individually this is going to take a great deal of time and resources.

- **Expensive** - as in the above example, the cost of a researcher's time (amongst other things) may be a factor in the design of the research.



AS Sociology For AQA

Access: Having designed a piece of primary research, you need access to the people you want to study – and your plan to interview the 10 richest people in the UK, for example, may come to nothing if they refuse to be interviewed.

Availability: Sometimes it's just *impossible* to collect primary data. In the above instance, for example, it's impossible because the people you want to research do not make themselves available to you. In another (admittedly more extreme) example, if you wanted to research why people commit suicide this would be difficult because your potential subjects refuse to answer your questions because they're *dead*. In this case, one way around the problem of availability is to use *secondary data*. **Durkheim** (1897), for example, used **official statistics** to test whether suicide rates varied within and between societies. By so doing, he argued social factors, such as religious belief, were significant in the explanation of why people took their own life. This leads us neatly to consider:

Secondary Data

In terms of **advantages** we can note the following:

Resources: Because secondary data already exists (someone else has done the work of collecting it) there are advantages in terms of *time* and *money* – collecting primary data on national crime or unemployment statistics, for example, would be a potentially daunting task. In some instances, *access* to data is much easier, although the researcher does, of course, have to rely on the availability / existence of such data.



Module Link

Education

Secondary data – in the form of official statistics – are useful for tracking a range of educational issues on a national (and international) basis, from levels of absence, through examination results to class sizes at primary and secondary level. A useful source of secondary data here is something like **Social Trends**, a digest of official government statistics published annually on a wide range of topics (family life, work, education and so forth).

Reliability: Some (but not all) forms of secondary data can be highly reliable – *official statistics* (those produced by the UK government, for example) are a good case in point – for a couple of reasons:

1. They are collected regularly and consistently in the same way from the same sources. Educational statistics, for example, are regularly collated by the **Office for National Statistics** from a variety of government sources and surveys.
2. They generally measure the same things each time they are collected so that any **comparisons** made between different years are comparing "like with like". For example, official statistics measuring educational achievement at GCSE consistently use the same

Sociological Methods

definition of achievement (grades A* - C). This isn't to say, of course, that definitions do not change over time; at A-level, for example, the current (2007) pass grades (A - E) will change in 2008 to

A* - E pass grades.

However, if the researcher is made aware of a definitional change (as is normally the case with official statistics) it is possible to adjust the research to take account of this potential threat to reliability.

Validity: Again, while it's not always easy to make generalisations, some forms of secondary data (*biographies* and *personal documents* such as *diaries* for example) provide

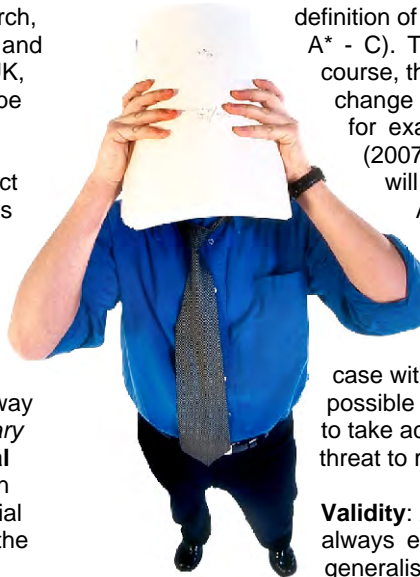
highly valid data because they give detailed insights into people's thoughts and behaviour – something that may be especially important and / or useful in circumstances where individuals are dead or have written contemporary accounts of historical events. Although it may, in some circumstances, be possible to generate primary data from such people (presupposing they are still alive...) validity may be lowered if the researcher is asking people to remember events that happened many years previous to the interview.

Representativeness: Where data is produced on a national level, by the government for example, there is normally a high level of representativeness because the level of **resources** (such as funding, number of researchers and so forth) available to governments means that large samples can be constructed. The **Census** (a **questionnaire** distributed to every household in the UK every 10 years), for example, is a *highly representative* sample of the UK population (its *reliability* is also high because it must, by law, be completed by every recipient).

In terms of some **disadvantages** of secondary data, however, we can suggest:

Data Control: This may be difficult because secondary data is not always produced with the needs of sociologists in mind. The data's creator will have their own reasons for producing it and these may not coincide with sociological concerns, interests and agendas. The way governments, for example, measure **social class** may be different to sociological ways of measuring class.

Reliability: The range and variety of secondary data available to the researcher makes generalisations about reliability difficult – some sources, such as official statistics, may be reliable whereas others, such as a diary or newspaper article may be potentially unreliable. In this instance to access the reliability of secondary data we always need to keep in mind questions like



Surprising as it may seem, not everyone welcomes being studied by sociologists...



who produced it, how it was produced and the reasons for its production.

Validity and Representativeness: An important consideration with secondary data is the extent to which it simply represents the viewpoint of one individual or a much wider range of views. Newspaper articles, for example, can be the personal, unsupported and unrepresentative view of a single journalist. Similarly, historical documents may reflect the views of particular social classes (because it was generally the upper classes in Britain who, until quite recently perhaps, recorded their particular view of the world). Conversely, the only surviving record of something may provide a valid insight into that event, but without supporting evidence (a question of *reliability*) we can't be certain of either its validity or *representativeness*. In addition, the *authenticity* (has the data been faked?) and *credibility* (who produced it and for what reasons?) of secondary data may be difficult to check.

As with *reliability*, the range and scope of secondary data makes it difficult to generalise about its validity – some forms (such as eyewitness descriptions of an event) may have greater validity than official statistics that simply focus on quantifying something.

Quantitative Data

This type of data has a number of distinct **advantages** for sociological researchers:

Quantification: The ability to express relationships *statistically* can be advantageous if the researcher doesn't particularly need or want to explore the *reasons* for people's behaviour. For example, if you simply need to know the *number* of murders committed each year or the *number* of students absent from the classroom in any given month then quantitative data satisfies this purpose more than adequately.

Social changes: Following from the above, quantitative data gives us an easy, manageable, way of tracking *social changes* over time. For example, statistics on

educational achievement over the past 25 years can show us changes in relative levels of achievement between different genders, ethnic groups and social classes.

Module Link

Education

Changes in the relative levels of educational achievement are explored in the Section **"Differential educational achievement of social groups"**

Comparisons: Similarly, if we want to compare differences between two or more things, (such as middle-class and working-class family size within our society), quantitative data makes this relatively easy. Alternatively, *cross-cultural* comparisons (crime rates in different countries, for example) are similarly made possible through the use of quantitative data. In addition:

"Before" and "after" studies are a further type of comparison we can make using quantitative data. For example, we could examine, using statistical data, the effect changes in the law have had on patterns of divorce in our society by quantifying the number of divorces *before* and *after* a change in the divorce law..

Module Link

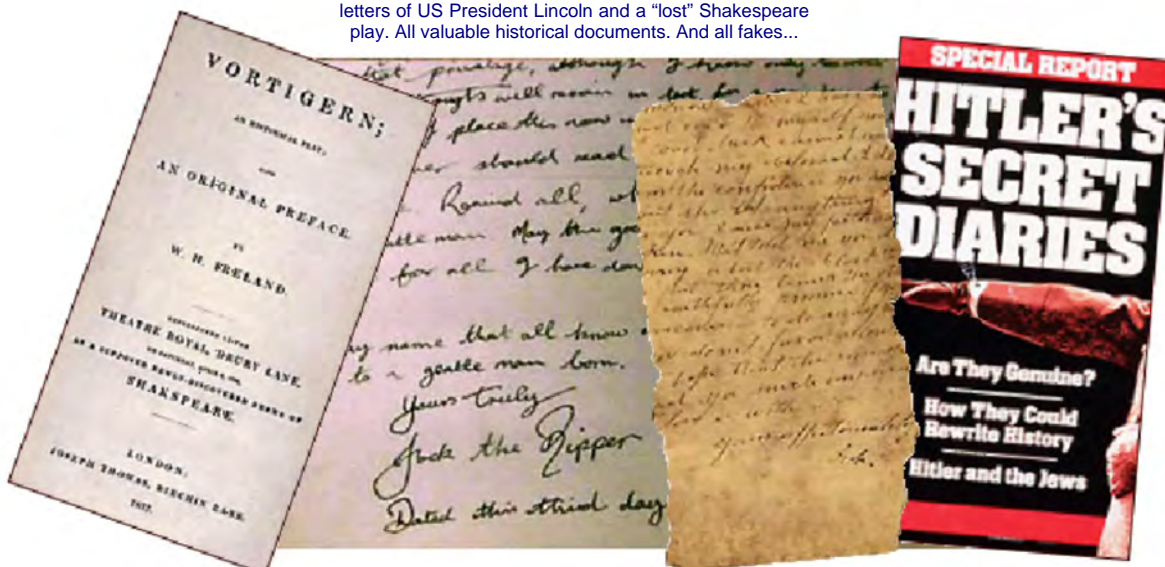
Families and Households

The relationship between divorce and legal change is explored in more-detail in the Section **"Changing patterns of marriage, cohabitation, separation, divorce, child-bearing and the life-course"**.

Reliability: As a general rule, quantitative data tends to be more reliable than qualitative data because it's easier to *replicate* (repeat) the collection of such data. This is because *standardised questions* (questions that don't change) can be asked to different groups (or the same group at different times).

Enabling studies: Although we have, for the sake of clarity, discussed quantitative and qualitative data

The diaries of Jack the Ripper and Adolf Hitler, the love letters of US President Lincoln and a "lost" Shakespeare play. All valuable historical documents. And all fakes...



separately (as if the two are mutually exclusive) there are occasions when a researcher may want to combine the two types of data. This may, for example, involve collecting statistics about educational achievement or the number of people who visit their doctor each year alongside qualitative data that seeks to explore the satisfaction levels of pupils or patients.

Alternatively, quantitative data is sometimes collected as a *prelude* to qualitative research. For example, a researcher looking at reasons for school truancy in their locality may firstly carry-out a quantitative analysis to discover whether or not pupils are actually absent from the classroom. In this respect a quantitative *enabling study* can be used to establish whether or not there is anything for the researcher to qualitatively investigate...



Quantitative data does, of course, have **disadvantages**, a couple of which involve:

Validity: This type of data can't be easily used to explore issues in any great *depth*; as we've suggested,

knowing the number of thefts in our society doesn't tell us anything about *why* people commit steal. Similarly, the knowledge that working class boys have lower levels of educational achievement than middle class girls doesn't tell us anything about the possible reasons for this situation (although it may, as we've suggested, **enable** the sociologist to identify a sociological problem to research).

Meanings: Related to the above, quantitative data isn't designed to tell sociologists much - if anything - about how people *interpret* and *understand* social behaviour. For example while it might be possible to quantify "the *fear of crime*" (counting the percentage of people who fear being a victim, for example), quantitative data tells us nothing about *why* people may be fearful of victimisation.

Qualitative Data



In terms of **advantages** we can note:

Validity: Because this type of data encourages *depth* and *detail* (in an interview, for example, people may be

encouraged to talk at great length about themselves and their beliefs) we are more likely to gain a *complete picture* of whatever we are researching or measuring.

Meanings: Qualitative data allows sociologists to explore the meanings people give to events and behaviour. While we can represent divorce statistically, for example, qualitative data allows us to explore how people feel and react to this situation. The same, of course, is true for areas like education and health.

In sign language this gesture means "I can smell something disgusting on my fingers". Possibly..

Imposition: If your research objective is to *understand* the *meaning* of people's behaviour, it follows you must allow people the scope to talk freely about that behaviour. If a researcher *imposes* their interpretation on a situation (by asking direct, quantifiable, questions for example) then data *validity* will be affected because you are restricting people's ability to talk at length and in depth about what they believe. Qualitative data may avoid this type of problem (although it may create a different kind of *imposition problem* which we'll examine in more detail when we consider different research methods).

Some **disadvantages** of qualitative data we can note are:



Reliability: Qualitative research is, by its very nature, difficult (if not impossible) to **replicate** (think, for example, about how difficult it would be to exactly repeat even a very recent conversation you've had with somebody). In addition, with something like *historical* data we may have no reliable way of knowing if our data source is *representative* of anything more than the views of a single individual.

Data Overload: Qualitative research tends to produce masses of data, much of which will be largely *irrelevant* in terms of achieving the research objective. With something like an interview, the problem of how to *interpret* or represent the data may also occur. Do you as a researcher report *everything* someone says or do you edit the data (and risk *imposing* your interpretation on the information)? A similar, if slightly different, problem is presented by observational forms of research – these too produce masses of data, the relevance of which has to be interpreted by the researcher (and may involve making difficult decisions about what to include or exclude as part of the research analysis).

Comparisons: Qualitative data makes *measuring* and comparing behaviour very difficult, mainly because the data can't be easily *standardised*. It's very difficult, for example, to ensure that you're comparing "like with like"; if you were interviewing people about their attitudes to something like fear or crime how difficult would it be to ensure that everyone in your sample thinks about (interprets) "fear" in the same way?

Reliability



Data Reliability is an important research consideration since, if data is *unreliable*, any conclusions we draw from it are going to be fairly limited (if not useless). For example, if I attempt to draw conclusions about the state of education in Britain on the basis of a couple of interviews I conducted "down the pub" with whoever happened to be present at the time, it's probable such data will

not be very reliable as a guide to what is actually happening in the educational system. In

general terms, therefore, data reliability is affected by such things as:

Bias: Are there opportunities for the researcher (consciously or unconsciously) to distort the data collection process?

Standardisation: Is everyone in the group you are researching asked the same questions in the same way? If they're not, how easy would it be to check data reliability by repeating this research?

Consistency: Will, for example, the same question asked of the same person in similar circumstances, produce the same answer?

Replication: If another sociologist attempted to repeat my "down the pub" research would similar results be achieved? If not, then my research would not be very reliable...

Tried and Tested

- (a) Explain what is meant by the term "primary data" (2 marks).
- (b) Suggest two reasons why sociologists might want to collect quantitative rather than qualitative data (4 marks).
- (c) Suggest two reasons why sociologists might use quantitative data (4 marks).
- (d) Examine the problems sociologists may find when considering the reliability and validity of their research. (20 marks).

Validity

Data Validity is a useful concept because it reminds us to think about the *accuracy* – or otherwise – of different data types (primary, secondary, qualitative and quantitative). While some forms of data (such as official statistics) may be reliable, their validity may be questionable for two reasons:

Representativeness: They may not apply to everyone in a particular group. In the UK, for example, "unemployment statistics" only represent those who are registered for unemployment benefit with the government - not everyone who doesn't have a job.

Depth: They may lack the depth and detail required to accurately represent the views of a particular individual or group (and so measure what they aim – or claim - to measure).

In both these respects, therefore, when evaluating the validity of a particular research method, data type or data source we need to always keep in mind the question of whether these actually measure what they claim to be measuring; if they do (however, limited their scope may be), then they are valid. If they don't then validity is likely to be both compromised and low.

References

Boyle, Jimmy (1977) "A Sense of Freedom": Pan Books.

Durkheim, Emile (1951: first published 1897) "Suicide: A Study In Sociology": The Free Press

Goffman, Erving (1968: first published 1961) "Asylums": Penguin.