The distinctions between primary and secondary data, and between quantitative and qualitative data;

© Chris. Livesey 2006: www.sociology.org.uk
Types of Data

Primary

Secondary

Quantitative

Qualitative

Information collected **personally** by a sociologist - who, therefore, knows exactly how the data was collected, by whom and for what purpose. A range of research methods (such as questionnaires, interviews and observational studies) can be primary data sources.

Information **not** personally collected by the researcher, but used by them in their research. Sources include newspaper articles, books, magazines, personal documents, official documents and the research of other sociologists.

Expressing data **statistically** or **numerically**. For example, the number of people who commit crimes each year.

Capture the **quality** of people’s behaviour. Qualitative data says something about the way people experience the social world and can be used to understand the meanings people give to behaviour.

- **Number**
  - For example, the number of people who live in poverty.

- **Percentage**
  - The number of people per 100 in a population.
  - Example, 30% of voters in Britain regularly vote Labour.

- **Rate**
  - Sociologically, the number of people per 1000 in a population.
  - For example, if the suicide rate was 1, for every 1000 people in a population, 1 person commits suicide each year.

- **Comparisons**
  - Example: Levels of unemployment or crime between countries. Percentage / rate allows us to compare “like with like”.

Boyle (1977) studied the behaviour of a juvenile gang from the viewpoint of its members.

Goffman (1968) tried to understand the experiences of patients in an American mental institution.
We can check the data we get from our research by repeating that research to see if we get the same results. Data is reliable if similar results are gained by different researchers asking the same questions to similar people. Example: A researcher may try to cross-check the reliability of a response within a questionnaire by asking the same question in a different way - if they get two different answers, it’s likely the data is unreliable.

Reliability

Validity

Methodology

Representativeness

Generalisability

We can check the data we get from our research by repeating that research to see if we get the same results. Data is reliable if similar results are gained by different researchers asking the same questions to similar people. Example: A researcher may try to cross-check the reliability of a response within a questionnaire by asking the same question in a different way - if they get two different answers, it’s likely the data is unreliable.

Reliability

Validity

Methodology

Representativeness

Generalisability

We can check the data we get from our research by repeating that research to see if we get the same results. Data is reliable if similar results are gained by different researchers asking the same questions to similar people. Example: A researcher may try to cross-check the reliability of a response within a questionnaire by asking the same question in a different way - if they get two different answers, it’s likely the data is unreliable.

Reliability

Validity

Methodology

Representativeness

Generalisability

We can check the data we get from our research by repeating that research to see if we get the same results. Data is reliable if similar results are gained by different researchers asking the same questions to similar people. Example: A researcher may try to cross-check the reliability of a response within a questionnaire by asking the same question in a different way - if they get two different answers, it’s likely the data is unreliable.

Reliability

Validity

Methodology

Representativeness

Generalisability

We can check the data we get from our research by repeating that research to see if we get the same results. Data is reliable if similar results are gained by different researchers asking the same questions to similar people. Example: A researcher may try to cross-check the reliability of a response within a questionnaire by asking the same question in a different way - if they get two different answers, it’s likely the data is unreliable.

Reliability

Validity

Methodology

Representativeness

Generalisability

We can check the data we get from our research by repeating that research to see if we get the same results. Data is reliable if similar results are gained by different researchers asking the same questions to similar people. Example: A researcher may try to cross-check the reliability of a response within a questionnaire by asking the same question in a different way - if they get two different answers, it’s likely the data is unreliable.

Reliability

Validity

Methodology

Representativeness

Generalisability

We can check the data we get from our research by repeating that research to see if we get the same results. Data is reliable if similar results are gained by different researchers asking the same questions to similar people. Example: A researcher may try to cross-check the reliability of a response within a questionnaire by asking the same question in a different way - if they get two different answers, it’s likely the data is unreliable.

Reliability

Validity

Methodology

Representativeness

Generalisability

We can check the data we get from our research by repeating that research to see if we get the same results. Data is reliable if similar results are gained by different researchers asking the same questions to similar people. Example: A researcher may try to cross-check the reliability of a response within a questionnaire by asking the same question in a different way - if they get two different answers, it’s likely the data is unreliable.

Reliability

Validity

Methodology

Representativeness

Generalisability

We can check the data we get from our research by repeating that research to see if we get the same results. Data is reliable if similar results are gained by different researchers asking the same questions to similar people. Example: A researcher may try to cross-check the reliability of a response within a questionnaire by asking the same question in a different way - if they get two different answers, it’s likely the data is unreliable.

Reliability

Validity

Methodology

Representativeness

Generalisability

We can check the data we get from our research by repeating that research to see if we get the same results. Data is reliable if similar results are gained by different researchers asking the same questions to similar people. Example: A researcher may try to cross-check the reliability of a response within a questionnaire by asking the same question in a different way - if they get two different answers, it’s likely the data is unreliable.

Reliability

Validity

Methodology

Representativeness

Generalisability

We can check the data we get from our research by repeating that research to see if we get the same results. Data is reliable if similar results are gained by different researchers asking the same questions to similar people. Example: A researcher may try to cross-check the reliability of a response within a questionnaire by asking the same question in a different way - if they get two different answers, it’s likely the data is unreliable.

Reliability

Validity

Methodology

Representativeness

Generalisability

We can check the data we get from our research by repeating that research to see if we get the same results. Data is reliable if similar results are gained by different researchers asking the same questions to similar people. Example: A researcher may try to cross-check the reliability of a response within a questionnaire by asking the same question in a different way - if they get two different answers, it’s likely the data is unreliable.

Reliability

Validity

Methodology

Representativeness

Generalisability

We can check the data we get from our research by repeating that research to see if we get the same results. Data is reliable if similar results are gained by different researchers asking the same questions to similar people. Example: A researcher may try to cross-check the reliability of a response within a questionnaire by asking the same question in a different way - if they get two different answers, it’s likely the data is unreliable.

Reliability

Validity

Methodology

Representativeness

Generalisability

We can check the data we get from our research by repeating that research to see if we get the same results. Data is reliable if similar results are gained by different researchers asking the same questions to similar people. Example: A researcher may try to cross-check the reliability of a response within a questionnaire by asking the same question in a different way - if they get two different answers, it’s likely the data is unreliable.

Reliability

Validity

Methodology

Representativeness

Generalisability

We can check the data we get from our research by repeating that research to see if we get the same results. Data is reliable if similar results are gained by different researchers asking the same questions to similar people. Example: A researcher may try to cross-check the reliability of a response within a questionnaire by asking the same question in a different way - if they get two different answers, it’s likely the data is unreliable.
**Example:** If another sociologist attempted to repeat the "pub research", would similar results be achieved? If not, then the research is unreliable…

**Examples:** Will the same question, asked of the same person in similar circumstances, produce the same answer? Is it possible for different people (or the same person at different times) to observe exactly the same things?

Is everyone in the group being researched asked the same questions in the same way? If not, how easy would it be to check data reliability by repeating this research?

Do opportunities exist for the researcher (consciously or unconsciously) to distort the data collection process?

If data is unreliable conclusions we draw from it are going to be of limited use. For example, if I attempt to draw conclusions about the state of education in Britain on the basis of an interview with someone in a bar, such data will be unreliable as a guide to what is really happening in the educational system.

Data validity encourages 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.

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

They may lack the depth and detail required to accurately represent the views of a particular individual or group.
Revision Mapping

AS Sociology For AQA

Control

Methodology

Advantages

Control over how data is collected doesn’t guarantee its reliability, validity or representativeness, but it’s much easier to consider these concepts when personally designing and carrying out research.

Disadvantages

Time-consuming

To design, construct and carry-out. If the group you’re researching is large and you’re interviewing them individually, this is going to take a great deal of time and resources.

Expense

The cost of a researcher’s time may be a factor in the design of the research.

Access

The group you want to study may not grant you access (or it may not be possible to construct a representative sample).

Availability

Sometimes it’s impossible to collect primary data - people may deny you access or the behaviour you want to research is in the past and the participants are no-longer alive.

Example: Durkheim (“Suicide: A Study In Sociology”, 1897).

Primary

Resources

There are advantages in terms of time and money – collecting primary data on national crime or unemployment statistics, for example, would be a daunting task. In some instances, access to data is much easier.

Secondary

Advantages

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.

Example: Government definitions of social class may be different to sociological definitions.

Disadvantages

Control

Methodology

Reliability

Some forms of secondary data (such as official statistics) can be highly reliable.

Validity

Some forms (biographies and personal documents such as diaries for example) give detailed insights into behaviour.

Representativeness

Where data is produced on a national level representativeness is normally high.

Expense

The cost of a researcher’s time may be a factor in the design of the research.

Availability

Sometimes it’s impossible to collect primary data - people may deny you access or the behaviour you want to research is in the past and the participants are no-longer alive.

Example: Durkheim (“Suicide: A Study In Sociology”, 1897).

Is secondary data representative of a 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.

Conversely, the only surviving record of an event provides a (valid) insight into that event, but without supporting evidence (a question of reliability) we can’t be certain of its representativeness. The authenticity (has the data been faked?) and credibility (who produced it and for what reasons?) of secondary data may be difficult to check.
Quantitative (statistical) data makes this relatively easy to compare differences between two or more things, (such as middle-class and working-class family size). Alternatively, cross-cultural comparisons (crime rates in different countries, for example) are made possible through the use of quantitative data.

Quantitative data gives us an easy, manageable, way of identifying, tracking and examining trends and changes over time. **Example:** Statistics on educational achievement over the past 25 years highlight changes in relative levels of achievement between (and within) the sexes.

The ability to express relationships statistically can be useful if you don’t need to explore the reasons for people’s behaviour (for example, if you only need to know the number of murders committed each year).

Quantitative data tends to be more reliable than qualitative data because it’s easier to replicate (repeat) the collection of such data. Standardised questions (questions that don’t change) can be asked to different groups (or the same group at different times).

Quantitative data can’t be easily used to explore issues in any great depth (knowing the number of thefts in our society doesn’t tell us anything about why people commit this crime).

Quantitative data isn’t designed to tell sociologists about how people interpret and understand social behaviour; that is, in terms of the various meanings they give to both their own behaviour and that of others. **Example:** While it might be possible to quantify “the fear of crime” (counting the percentage of people who fear being a victim, for example), this type of data tells us nothing about why people may fear victimisation.

“Before” and “after” studies are a further type of comparison we can make using quantitative data. **Example:** Examining the effect legal changes have had on patterns of divorce in our society by noting the number of divorces before and after a change in the law.
If the research objective is to understand the meaning of people's behaviour, people must be given 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 people's ability to talk at length and in depth about what they believe is artificially restricted.

Because qualitative 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, true-to-life, picture of whatever we are researching.

Qualitative data allows sociologists to explore the meanings people give to events and behaviour. Example: While we can represent divorce statistically, qualitative data allows us to explore how people feel and react to this situation.

Qualitative research produces a mass of data, much of which will be irrelevant in terms of achieving the research objective.

Example: With something like an interview or participant observation there is the problem of how to interpret or represent data. Does the researcher report everything someone says or does or is it permissible to edit the data (and risk imposing a single interpretation on the information)?

Qualitative research is difficult (if not impossible) to exactly repeat - how difficult would it 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 source is representative of anything more than the views of a single individual.

Qualitative data makes measuring and comparing behaviour very difficult, mainly because the data can't be easily standardised.
Exam Questions

8 marks
Identify and briefly explain two advantages to sociologists in carrying out their own primary research “rather than relying on secondary sources of evidence”.

20 marks
Examine the problems some sociologists may find when using secondary data in their research.

20 marks
Examine the uses of different kinds of secondary data in sociological research.

© Chris. Livesey 2006