## Surveys, longitudinal, cross-sectional and trend studies

CHAPTER 17

Cohen, L., Manion, L., & Morrison, K. (2017). Research methods in education. Retrieved from http://ebookcentral.proquest.com Created from gasouthern on 2020-03-05 19:17:31.

There are many different kinds of survey; each has its own characteristics and key issues. We set these out in this chapter, addressing such matters as:

- what is a survey?
- some preliminary considerations
- planning a survey
- low response and non-response, and how to reduce them
- survey sampling
- longitudinal, cross-sectional and trend studies
- strengths and weaknesses of longitudinal, cohort and cross-sectional studies
- postal, interview and telephone surveys
- comparing methods of data collection in surveys

#### **17.1 Introduction**

We advise readers to take this chapter in conjunction with Chapters 12 (sampling), 24 (questionnaires), 25 (interviews) and data-analysis techniques (Part 5). Many researchers reading this book will probably be studying for higher degrees within a fixed and maybe short time frame; that may render longitudinal study out of the question for them. Nevertheless longitudinal study is an important type of research, and we introduce it here. More likely, researchers for higher degrees will find cross-sectional survey research appropriate, and it is widely used in higher degree research.

In many quarters, Internet surveys are becoming the predominant method of surveys, through email (with a questionnaire as an attachment, or embedded in the email, or with a hyperlink link to a website, social networking site, special interest group, listserv, discussion group etc.), with companies providing free or low-cost software to design questionnaires and, indeed, to conduct the survey and collect data for researchers. Given the rise and widespread usage of Internet surveys, we devote an entire, separate chapter (Chapter 18) to this. However, we include reference to Internet surveys in Table 17.3 in this chapter, for purposes of comparison with other means of survey design and conduct.

#### 17.2 What is a survey?

Many educational research methods are descriptive; that is, they set out to describe and to interpret what is. Such studies look at individuals, groups, institutions, methods and materials in order to describe, compare, contrast, classify, analyse and interpret the entities and the events that constitute their various fields of enquiry. We deal here with several types of survey research, including longitudinal, cross-sectional and trend or prediction studies.

Typically, surveys gather data at a particular point in time with the intention of describing the nature of existing conditions, or identifying standards against which existing conditions can be compared, or determining the relationships that exist between specific events. They may vary in their levels of complexity, from those which provide simple frequency counts to those which present relational analysis.

Surveys may be further differentiated in terms of their scope and complexity. A study of contemporary developments in post-secondary education, for example, might encompass the whole of Europe; a study of subject choice, on the other hand, might be confined to one secondary school.

#### 17.3 Advantages of surveys

A survey has several characteristics and several claimed attractions; typically it is used to scan a wide field of issues, populations, programmes, people etc. in order to measure or describe any generalized features. It is useful (OECD, 2012; Dillman *et al.*, 2014) in that it often:

- gathers data on a one-shot basis and hence is economical and efficient;
- represents a wide target population (hence there is a need for careful sampling, see Chapter 12);
- generates numerical data;
- provides descriptive, inferential and explanatory information;
- manipulates key factors and variables to derive frequencies (e.g. the numbers registering a particular opinion or test score);

- gathers standardized information (i.e. using the same instruments and questions for all participants);
- ascertains correlations (e.g. to find out if there is any relationship between gender and test scores);
- presents material which is uncluttered by specific contextual factors;
- captures data from multiple-choice, closed questions, test scores or observation schedules;
- supports or refutes hypotheses about the target population;
- generates accurate instruments through piloting and revision;
- makes generalizations about, and observes patterns of response in, the targets of focus;
- gathers data which can be processed statistically;
- uses large-scale data gathered from a wide population in order to enable generalizations to be made about given factors or variables.

Examples of surveys are:

- test scores (e.g. from students nationally, internationally, locally);
- students' preferences for particular courses, for example, humanities, sciences;
- attitudes to, and opinions of, quality of teaching;
- surveys of groups of people's values over time;
- surveys of factors (e.g. income levels, social class membership, inequality) over time;
- opinion polls;
- reading and mathematics performance surveys.

Surveys in education often use test results, selfcompletion questionnaires and attitude scales. Here a researcher may be seeking to gather large-scale data from as representative a sample as possible in order to say with a measure of statistical confidence that certain observed characteristics occur with a degree of regularity, or that certain factors cluster together (see Chapter 43) or that they correlate with each other (correlation and covariance), or that they change over time and location (e.g. results of test scores used to ascertain the 'value-added' dimension of education, maybe using regression analysis and analysis of residuals to determine the difference between a predicted and an observed score), or regression analysis to use data from one variable to predict an outcome on another variable.

Surveys can be *exploratory*, in which no assumptions or models are postulated, and in which relationships and patterns are explored (e.g. through correlation, regression, stepwise regression and factor analysis). They can also be *confirmatory*, in which a model, causal relationship or hypothesis is tested (see

the discussion of exploratory and confirmatory analysis in Chapter 43). Surveys can be descriptive or analytic (e.g. to examine relationships). Descriptive surveys simply describe data on variables of interest, whilst analytic surveys operate with hypothesized predictor or explanatory variables that are tested for their influence on dependent variables or relationships between variables.

Many surveys combine nominal data on participants' backgrounds and relevant personal details with other data (e.g. attitude scales, data from ordinal, interval and ratio measures) (see Chapter 38). Surveys are useful for gathering factual information, data on attitudes and preferences, beliefs and predictions, opinions, behaviour and experiences – both past and present (Weisberg *et al.*, 1996; Aldridge and Levine, 2001; Dillman *et al.*, 2014). Their attraction lies in their appeal to generalizability or universality within given parameters, their ability to make statements which are supported by large data and their ability to establish the degree of confidence which can be placed in a set of findings.

On the other hand, if a researcher is concerned to catch local, institutional or small-scale factors and variables - to portray the specificity of a situation, its uniqueness and particular complexity, its interpersonal dynamics, and to provide explanations of why a situation occurred or why a person or group of people returned a particular set of results or behaved in a particular way in a situation, or how a programme changes and develops over time – then a survey approach may be unsuitable. Its explanatory potential or fine detail is limited; it is lost to broad-brush, often descriptive generalizations which are free of temporal, spatial or local contexts. Williams et al. (2016) note that having a twophase process of a postal survey – an initial screening survey followed by the topic-based survey sent to eligible people - is also a useful device for obtaining more in-depth data. In a survey the individual instance is sacrificed to the aggregated response (which has the attraction of anonymity, non-traceability and confidentiality for respondents and opportunity for trends and patterns to be discovered).

Surveys typically, though by no means exclusively, rely on large-scale data, for example, from questionnaires, test scores, attendance rates, results of public examinations etc., all of which enable comparisons to be made over time and between groups. This is not to say that surveys cannot be undertaken on a small-scale basis, as indeed they can; rather it is to say that the generalizability of such small-scale data will be slight. In surveys the researcher is usually an outsider; indeed questions of reliability and possible bias can attach themselves to researchers conducting survey research on their own subjects, for example, participants in a course that they have been running. Further, it is critical that attention is paid to rigorous sampling, otherwise the basis of the survey's applicability to wider contexts is seriously undermined. Non-probability samples tend to be avoided in surveys if generalizability is sought; probability sampling will tend to lead to generalizability of the data collected.

### 17.4 Some preliminary considerations

A fundamental decision by the researcher is whether a survey is the appropriate means of answering the research purposes and research questions (Magee *et al.*, 2013). Assuming that it is, three prerequisites to the design of any survey are: the specification of the exact purpose of the enquiry; the population and issues on which it is to focus; and the resources that are available. Hoinville and Jowell's (1978) consideration of each of these key factors in survey planning can be illustrated in relation to the design of an educational enquiry.

#### The purpose of the enquiry

First, a survey's general purpose must be translated into a specific central aim. Thus, 'to explore teachers' views about in-service work' is somewhat nebulous, whereas 'to obtain a detailed description of primary and secondary teachers' priorities in the provision of in-service education courses' is reasonably specific.

Having decided upon and specified the primary objective of the survey, the second phase of the planning involves the identification and itemizing of research questions which will enable the objective to be addressed. The third phase, usually driven by the research questions, is to identify subsidiary topics that relate to its central purpose. In our example, subsidiary issues might well include: the types of courses required; the content of courses; the location of courses; the timing of courses; the design of courses; and the financing of courses.

The fourth phase follows the identification and itemization of subsidiary topics and involves formulating specific information requirements relating to each of these issues. For example, with respect to the type of courses required, detailed information would be needed about the duration of courses (one meeting, several meetings, a week, a month, a term or a year), the status of courses (non-award bearing, award bearing, with certificate, diploma, degree granted by college or university), the orientation of courses (theoretically oriented involving lectures, readings, etc., or practically oriented involving workshops and the production of curriculum materials).

As these details unfold, consideration has to be given to the most appropriate ways of collecting items of information (interviews with selected teachers, postal questionnaires to selected schools, online questionnaires etc.).

### The population upon which the survey is focused

The second prerequisite to survey design, the specification of the population (e.g. people, issues) to which the enquiry is addressed, affects decisions that researchers must make both about sampling and resources. In our hypothetical survey of in-service requirements, for example, we might specify the population as 'those primary and secondary teachers employed in schools within a thirty-mile radius of Loughborough University'. In this case, the population is readily identifiable and, given sufficient resources to contact every member of the designated group, sampling decisions do not arise.

Things are rarely so straightforward, however. Often the criteria by which populations are specified ('severely challenged', 'under-achievers', 'intending teachers' or 'highly anxious') are difficult to operationalize. Populations, moreover, vary considerably in their accessibility; students and student teachers are relatively easy to survey, travellers' children and headteachers are more elusive. More importantly, in a large survey, researchers usually draw a sample from the population to be studied; rarely do they attempt to contact every member. We deal with the question of sampling shortly.

#### The resources available

Resources are not simply financial. For example, survey design can be costly in terms of time, and consideration of resources has to include human, material, financial, administrative, temporal, geographical, technical (e.g. computer-related) costs. An important factor in designing and planning a survey is financial cost. Sample surveys are labour-intensive, the largest single expenditure being fieldwork, where costs arise out of interviewing time, travel time and transport costs of the interviewers themselves. There are additional demands on the survey budget. Training and supervising the panel of interviewers can often be as expensive as the costs incurred during the time that they actually spend in the field. Questionnaire construction, piloting, printing, posting, coding, together with computer programming and processing all eat into financial resources.

#### Mode of data collection

There are two main issues to be addressed here:

- 1 Will the researcher be completing the survey by entering data, or will the participants be selfadministering the survey?
- 2 How will the survey be administered, for example, a postal survey, a telephone survey, an Internet survey, by face-to-face interviews, group-administered surveys, self-administered surveys, drop-off surveys, email? A full account of the interview as a research technique is given in Chapter 25.

Dillman *et al.* (2014) advise researchers to use multiple and mixed modes of delivery/administration, as this helps response rates.

#### Self-reporting

There can be a large difference in the responses gained from self-reporting and those obtained from face-toface survey interviews or telephone interviews (Dale, 2006, p. 145; Dillman et al., 2014). Many surveys ask respondents not only to administer the questionnaires themselves but also to report on themselves. This may introduce bias, as respondents may under-report (e.g. to avoid socially undesirable responses) or over-report (to give socially desirable answers). Self-reporting also requires the researcher to ensure that: respondents all understand the question, understand it in the same way and understand it in the way intended by the researcher (Kenett, 2006, p. 406). The difficulty here is that words are inherently ambiguous (see Chapter 24 on questionnaire design), so the researcher should be as specific as possible. The researcher should also indicate how much contextual information the respondent should provide, what kind of answer is being sought (so that the respondent knows how to respond appropriately), how much factual detail is required and what constitutes relevant and irrelevant data (e.g. the level of detail or focus on priority issues required) (pp. 407-8). Further, surveys that rely on respondents' memory may be prone to the bias of forgetting or selective recall.

#### **Ethics**

Ethical issues are discussed in Chapters 7, 8 and 18, and we refer readers to these; here we note the importance of gaining the informed consent of respondents. Whilst completion of the survey might be taken as giving consent, this may not always be the case, and the completion of a consent form may be needed (though some participants may be suspicious of this), and indeed Dillman *et al.* (2014) note that asking for consent requires the researcher to make it clear what the consent is being given for, as, for example, to ask for consent before the questions have been asked is asking participants to take a leap of faith. We also note in Chapters 7 and 28 that informed consent is complex, as it is unclear what is being consented to, and for how long, and for what purposes and uses, and that these problems are exacerbated when data are archived for future use as secondary data sets. Informed consent should also include the right not to participate or to withdraw at any time.

Ethical issues here also concern attention to confidentiality, anonymity, privacy and non-traceability. In paper-based surveys this may be easy to guarantee, but, as we indicate in Chapter 18, for electronic and Internetbased (e.g. website and email surveys), no such absolute guarantees are available. Such computer-related problems raise the matter of data security and identity protection. In electronic and paper surveys, telephone interviewing and face-to-face surveys, the researcher might not ask for, or require, identifying features, or might remove these when storing and archiving data.

However, in group interviews these may not be so easy to protect (e.g. members of the group may talk to others), and in electronic/Internet-based surveys, the service provider can log and track participants, and data miners and hackers can break into data, particularly email, even when security steps have been taken. We discuss this in Chapter 18.

The researcher can, and should, take all reasonable steps to protect confidentiality, anonymity, privacy and non-traceability and indicate to respondents what those steps are, recognizing that where there are limits (e.g. in electronic surveys), this may lead to some respondents not taking part.

Underpinning ethical issues in surveys is the requirement of *primum non nocere*: primarily, do no harm. The researcher must take every step necessary to address this. This concerns access to, collection, storage, use, dissemination and reporting of data, and subsequent archiving of data or locating the data in the public domain, with immense care being taken with regard to identification and sensitive information. This raises issues not only of removing identifying features, removing certain data, aggregating or anonymizing data, but who owns the data and what rights the owner has, once the data have been given to the researcher. The researcher has a duty of care and of trust here.

### 17.5 Planning and designing a survey

Whether the survey is large scale and undertaken by some governmental bureau, or small scale and carried out by the lone researcher, the collection of information typically involves one or more of the following datagathering techniques: structured or semi-structured interviews, self-completion (e.g. postal and Internet questionnaires), telephone interviews, Internet surveys, standardized tests of attainment or performance, and attitude scales.

Planning a survey involves knowing: (a) what exactly you wish to find out, and why; (b) what data you need to be able to answer (a); (c) what questions you will ask to acquire the data. Researchers must also consider: sample selection and access to the sample; distribution/data collection and return of surveys; measurement design and data types; ethical issues; piloting; analysis and reporting.

Sapsford (1999, pp. 34–40) suggests that there are four main considerations in planning a survey:

problem definition (e.g. deciding what kinds and contents of answers are required; what hypotheses

there are to be tested; what variables there are to explore);

- sample selection (e.g. what is the target population; how can access and representativeness be assured; what other samples will need to be drawn for the purpose of comparison);
- design of measurements (e.g. what will be measured, and how (i.e. what metrics will be used see Chapter 24 on questionnaires); what variables will be required; how reliability and validity will be assured);
- concern for participants (e.g. protection of confidentiality and anonymity; avoidance of pain to the respondents; avoiding harm to those who might be affected by the results; avoiding over-intrusive questions; avoiding coercion; informed consent; see Chapters 7 and 8).

Typically surveys proceed through well-defined stages, outlined in Figure 17.1. Though these are set in a



sequence, the sequence may alter and the process is iterative and recursive. The process moves from the general to the specific. A general research topic is operationalized into component issues and questions, and, for each component, questions are set. As with questionnaires (Chapter 24), it is important, in the interests of reliability and validity, to have several items or questions for each component issue, as this does justice to the all-round nature of the topic.

Rosier (1997, pp. 154–62) suggests that the planning of a survey must include clarification of:

- the research questions to which answers need to be provided;
- the conceptual framework of the survey, specifying in precise terms the concepts that will be used and explored;
- operationalizing the research questions (e.g. into hypotheses);
- the instruments to be used for data collection, for example, to chart or measure background characteristics of the sample (often nominal data), academic achievements (e.g. examination results, degrees awarded), attitudes and opinions (often using ordinal data from rating scales) and behaviour (using observational techniques);
- sampling strategies and sub-groups within the sample (unless the whole population is being surveyed, e.g. through census returns or nationally aggregated test scores etc.);
- pre-piloting the survey (to generate items for the survey);
- piloting the survey;
- data-collection practicalities and conduct (e.g. permissions, funding, ethical considerations, response rates);
- data preparation (e.g. coding, data entry for computer analysis, checking and verification);
- data analysis (e.g. statistical processes, construction of variables and factor analysis, inferential statistics);
- reporting the findings (answering the research questions).

Ruel *et al.* (2015) comment that researchers need to consider:

- the kind of survey to be used;
- ethical issues;
- questionnaire and instrument design and appearance;
- question construction (measures, responses and measurement error);
- validity and reliability;

- sampling;
- response rates, non-responses and attrition;
- the medium of delivery, completion and return of the survey;
- data entry and data cleaning;
- data analysis and reporting;
- missing data;
- data archiving.

It is important to pilot and pre-pilot a survey. The difference between the pre-pilot and the pilot is this: the pre-pilot is usually a series of open-ended questions that are used to generate items and categories for closed, typically multiple-choice questions, whilst the pilot is used to test the draft of the actual survey instrument itself (see Chapter 24).

A rigorous survey formulates clear, specific objectives and research questions; ensures that the instrumentation, sampling and data types are appropriate to yield answers to the research questions; and ensures that as high a level of sophistication of data analysis required can be done (i.e. as the data will sustain).

Attention must be given to: the mode of data collection; respondent effort (too much and this can lead to non-response); question wording, sequence and format.

#### Some challenges in planning surveys

A survey is no stronger than its weakest point, and we consider a range of issues here in order to strengthen each aspect of a survey (e.g. OECD, 2012). Surveys must minimize errors caused by:

- poor sampling (e.g. failure to represent or include sufficiently the target population);
- poor question design and wording (e.g. failure to catch accurately the views of, or meanings from, the respondents or to measure the factors of interest);
- incorrect or biased responses;
- low response or non-response.

The first of these – a sampling matter – may be caused by a failure correctly to identify the population and its characteristics, or a failure to use the correct sampling strategy, or systematically to bias the sample (e.g. using a telephone survey based on telephone directory entries, when key people in the population – the poor – may not have a telephone, or may have a cellphone rather than a fixed line (the young, the middle aged but not the elderly), or using an Internet- or email-based survey when many respondents do not have access). We address sampling issues in Chapter 12 and below.

The second of these is a failure to operationalize the variables fairly (i.e. a validity issue) or a failure in the

wording or meanings used or inferred, such that incorrect responses are collected (a reliability issue) (e.g. people may not understand a question, or may misinterpret it, or interpret it differently). We address this in Chapter 14 and below.

The third problem is that some participants may deliberately over-report or under-report the real situation in – often sensitive – matters. For example, teenage alcohol, smoking or drug use, underage sexual relations, bullying, domestic violence, petty criminality may be *systematically* under-reported (i.e. be biased), whereas the popularity of a teacher or students might be over-reported (i.e. biased). Bias obtains where there is a *systematic* skewing or distortion in the responses.

Further, some questions may rely on memory, and memory can be selective and deceptive (e.g. people may not remember accurately). Also, some responses will depend on a person's state of mind at the time of completing the survey – asking a teacher about teacher stress and tiredness late on a Friday afternoon in school with a difficult class could well elicit a completely different response from asking her directly after a week's holiday. Some questions may be so general as to be unhelpful (e.g. 'how stressed do you feel?'), whereas others might be so specific as to prevent accurate recall (e.g. 'how many times have you shouted at a class of children in the past week?') (one solution to the latter might be to ask participants to keep a diary of instances).

Fowler (2009, p. 15) suggests that a respondent's answer is a combination of the true response plus an error in the answer given, with errors coming from many sources.

The fourth of these - low response or non-response - is a problem that besets researchers, and is so significant that we devote a separate section to it below.

Dillman *et al.* (2014) identify four key errors to be avoided in surveys which seek to represent a wider population:

- coverage error (poor and incomplete representation of the population in the sample). For example, a coverage error might be made if telephone or Internet surveys are used, as not everyone has a telephone (particularly a landline) or access to, and familiarity with, the Internet;
- sampling error (including inaccurate estimates of the population);
- non-response error (the difference between a representative result and that obtained from non-response of different individual or groups, i.e. a skewed response); and

measurement error: inaccurate and unreliable response because of (a) the metrics, scales and units of measurement used; (b) socially desirable responses and respondent acquiescence (the tendency to agree with an interviewer rather than disagree) in face-to-face survey interviews; (c) questionnaire features, for example, length, difficulty, questions asked, complexity, order effects, interviewer effects, survey mode (post, telephone, email, interview, Internet etc.).

#### 17.6 Survey questions

Though we go into detail about questions and questionnaires in Chapter 24, here we give advice on some important issues in writing and asking questions in surveys (Creswell, 2012; OECD, 2012; Abascal and Diaz de Rada, 2014; Champagne, 2014; Dillman *et al.*, 2014; Colorado State University, 2016):

- Ensure that the questions cover the topics and research questions comprehensively and with the appropriate scales of measurement and scales (e.g. 1-5, -4 to +4, 'strongly disagree' to 'agree').
- Keep the survey simple and short, and use whole, short sentences.
- Consider respondent effort: avoid overloading the respondent with thinking, recalling, reading and responding.
- Ensure that the questions apply to all the respondents.
- Consider the order of the questions (questions are not independent of each other, and the answer to one question may affect the answer to another in the respondent's mind, e.g. the primacy effect, 'carry over' and 'anchoring effect' (Dillman *et al.*, 2014, p. 235), i.e. what comes first affects what comes later and respondents use the early questions as a standard against which they compare the later questions).
- Arrange the order and organization of the survey in a way that is easy for the respondent to understand (subheadings in a written survey are important here).
- Group together questions that cover similar topics, with subheadings in written surveys, to parallel what would naturally happen in a conversation (NB if respondents see two questions as similar then, for consistency, they will give answers which are similar).
- Start the survey with questions that respondents will find meaningful and interesting, and will be able to answer.

- If you are using branching questions, ask all the branching questions before you ask the follow-up questions.
- Ensure that the wording is comprehensible to the respondent (use easy words) and judge how the respondent will regard and feel about the question asked.
- Keep sensitive questions until later in the survey.
- Avoid putting the important questions right at the end of the survey.
- Consider the willingness of the respondent to answer the questions correctly and honestly, and whether the respondent will actually know the answer (e.g. to factual questions or to questions which require long-term memory), i.e. whether the question really applies to the respondent.
- Consider what the question is asking for for example, factual answers; attitudes, perceptions and opinions; behaviours; events – and how to make these clear to the respondent. Some factual information is easy (e.g. gender, age) but other data (e.g. attitudes, behaviours, those which rely on memory) may be less accurate.
- Use concrete, specific and precise terms (define terms concretely) so that the respondent understands exactly what is being asked for in the survey.
- Consider the suitability of question types and formats: (a) for nominal variables: dichotomous, multiple choice (single choice, restricted number of choices, free number of choices); (b) for ordinal variables: rating scales, ranking scales; (c) for interval, ratio and continuous variables: constant sum, percentages/marks out of ten, open number (e.g. number of hours of study in a week); (d) for nonnumerical answers: open questions. Decide whether to have a mid-point in scale items; use large-range scales if subsequent factor analysis is intended; and ensure that response categories are exhaustive, to fit the choices that participants will really want, i.e. that they enable respondents to say what they want to say (and this underlines the importance of running a pilot).
- Avoid: double-barrelled questions (asking more than one thing in a single question); long and complex questions and vocabulary; technical language; negatively worded items; ambiguous questions; leading questions (those which influence the response and indicate a desired response); questions which may cause embarrassment.
- Consider the medium of the administration/ conduct/'delivery' of the survey, for example, postal service, email, face-to-face interview, website, telephone, i.e. visual, oral and aural administration of

the survey, and who enters the responses (the respondent or the interviewer).

Consider whether it is advisable to have an interviewer present or absent, as the interviewer's presence may bias the respondent, raising issues of the respondent's concern for (a) social desirability and (b) acquiescence (defined above); acquiescence is a particular problem in questions which include 'agree', as there is a tendency to agree.

Magee et al. (2013) advise researchers to consider:

- how others have addressed the constructs in question; developing and writing relevant survey items clearly;
- the mode of the item, for example, a statement or a question (a question is preferable);
- the response (number and type, with no smaller than a five-point scale; odd numbers or even numbers in scaling; inclusion of positive and negative options or only positive options: avoid agreement- or positive-only responses; label each point in an ordinal scale);
- reliability and validity of items;
- ensuring that the question is interpreted by respondents in the way intended.

Given these points, it is essential that a survey be piloted, and we give guidelines to piloting in Chapter 24, for example, for content, coverage, ease of understanding, timing, redundancy, sensitivity, question types, question order, mode of delivery, ease of completion, answerability.

## 17.7 Low response, non-response and missing data

Response and non-response are related to contact, cooperation and ease of conduct, completion and return of the survey (Dillman *et al.*, 2014). Non-response to a whole questionnaire ('unit non-response'; Durrant, 2009, p. 293) or to a specific item ('item non-response'; p. 293) is a serious problem for much survey research, though Denscombe (2009b, p. 282) notes that online surveys tend to have lower item non-response than paper-based surveys, though there may be more dropouts before reaching the end of an online survey than in a paper-based survey.

Dale (2006, p. 148) suggests that 'non-respondents almost invariably differ from respondents', and that this affects the validity and reliability of the responses obtained, and their analysis. If non-response is received from a very homogeneous sample then this might be less of a problem than if the sample is very varied. Further, if non-response is received randomly across a sample then this might be less of a problem than if the non-response was from a particular sub-sector of the respondents (e.g. a very low or a very high socioeconomic group), as this would bias the results (cf. Dale, 2006, p. 148). A subset of non-response to a whole questionnaire is item non-response, and here missing data should not be ignored (Dale, 2006, p. 15).

Rubin (1987), Little and Rubin (1989), Allison (2001), Dale (2006, pp. 149-50) and Durrant (2006, 2009) review a range of different 'imputation methods' for handling and weighting non-response, i.e. methods for filling in missing data with 'plausible values' in order to render a set of data complete and yet to reduce bias in the non-responses, i.e. that bias which might be caused by the non-responses having different values from the non-missing responses (Durrant, 2009, p. 295). These depend on whether the non-response is largely confined to a single variable or many variables. The researcher has to determine whether there are patterns of non-response, as these affect the method for handling non-response. For example, if the nonresponse is randomly distributed across several variables, with no clear patterns of non-response, then this may be less problematic than if there is a systematic non-response to one or more variables in a survey (Durrant, 2009, p. 295; Dillman et al., 2014). Durrant (2009) sets out several ways of calculating missing values, including:

- calculating missing values from regression techniques using auxiliary variables (p. 296);
- 'hot deck' methods, in which sub-groups of participants (based on their scores on auxiliary variables) are constructed and the researcher compares their results to the non-missing results of the respondent who had omitted a particular response (p. 297);
- 'nearest neighbour' techniques, in which the results from a person whose data diverge as little as possible from those of the missing person are used to replace the missing values.

Durrant (2006, 2009) and Dillman *et al.* (2014) identify further, statistical methods of calculating missing scores, such as multiple and fractional imputation and propensity score weighting. Durrant makes the point that how one calculates the values of missing data depends on a range of factors such as the purpose of the analysis, the variable(s) in question, the kinds of data, any patterns of missing data, and the characteristics and fittingness of the assumptions on which the particular intended imputation method is based. The National Centre for Research Methods (2016) also suggests that using means of groups and sub-groups responding to a particular item can be used for imputation. Here one looks for patterns of missing data (any groups of units/ cases or items) and calculates an average value (e.g. on a scaled item) for groups/sub-groups of cases (individuals), and reporting standard error.

Ary *et al.* (2002) note that non-respondents may be similar to late responders, so it might be possible to use data from late responders to indicate the possible responses from non-respondents. This requires the researcher to identify late responders.

Missing data within a survey can have many causes. For example, people may not be present on the day of its administration, or they may not understand the question, or they may take exception to the question or overlook it by mistake. Pampaka et al. (2016) give the example of the administration of a school survey on bullying, where students may be absent without predictable reasons, or they are representing their school in a competition (e.g. high-performing and highly motivated students), or they may be more likely to be bullied (p. 19). All these, the authors note, lead to biased data. They note that missing data are a particular problem in longitudinal surveys and surveys across phase transitions. They note that statistical analysis (e.g. stepwise regression, which ignores missing data) is dangerous if there are missing data, and they argue for multiple imputation methods. However, they also note that multiple imputation methods are essentially speculative, based on simulations (p. 21).

Pampaka *et al.* (2016) distinguish between missing data from units (individuals) and items, but both can lead to a biased response. There are many ways to address this, for example, by simply analysing incomplete data, or by weighting, and by imputation. Weighting is designed to ensure a better representation of the population, and it can be used to adjust data for non-response, to bring the data into the correct matching of the population. If the incomplete data are random, i.e. all cases have equal probability of being missing (as in their example of those students who are absent for unpredictable reasons), then the analysis may be unbiased (the claim of randomness for equality of distributions, see Chapter 20 on experiments).

For further guidance on weighting, standard error and imputation, we refer the reader to the sources indicated above and to the guidance from the National Centre for Research Methods (www.restore.ac.uk).

In some cases (e.g. when all the students in a class complete a questionnaire during a lesson) the response rate may be very high, but in other circumstances the response rate may be very low or zero, either for the whole survey or for individual items within it, for several reasons, for example:

- the survey never reaches the intended people;
- people refuse to answer;
- people may not be available (e.g. for a survey administered by interview), for example, they may be out at work when a telephone survey administrator calls;
- people may not be able to answer the questions (e.g. language, reading, speaking or writing difficulties);
- people may not actually have the information requested;
- people may overlook some items in error;
- the survey was completed and posted but failed to return;
- the pressure of competing activities on the time of the respondent;
- potential embarrassment at their own ignorance if respondents feel unable to answer a question;
- ignorance of the topic/no background in the topic;
- dislike of the contents or subject matter of the interview;
- fear of possible consequences of the survey to himself/herself or others;
- lack of clarity in the instructions;
- fear or dislike of being interviewed (or of the interviewer);
- sensitivity of the topic, or potentially insulting or threatening topic;
- betrayal of confidences;
- losing the return envelope or return address;
- the wrong person may open the mail, and fail to pass it on to the most appropriate person.

Non-response can lead to responses that are systematically different (i.e. biased) than those from the whole sample or population, as the responses from those who did not respond might be distinctively different from those who actually responded.

Later in this chapter we discuss ways of improving response rates. However, here we wish to insert a note of caution: some researchers suggest that, for nonresponders to an item, an average score for that item can be inserted. This might be acceptable if it can be shown that the sample or the population is fairly homogeneous, but, for heterogeneous populations or samples, or those where the variation in the sample or population is not known, it may be dangerous to assume homogeneity and hence to infer what the missing replies might have been, as this could distort the results.

Let us suppose that, out of a sample of 200 participants, 90 per cent reply (180 participants) to a 'yes/no' type of question, for example, for the question 'Do you agree with public examinations at age 11?', and let us say that 50 per cent (90 people) indicate 'yes' and 50 per cent indicate 'no'. If the 10 per cent who did not reply (20 people) had said 'yes' then this would clearly swing the results as 110 people say 'yes' (55 per cent) and 90 people say 'no' (45 per cent). However, if the response rates vary, then the maximum variation could be very different, as in Table 17.1 (cf. Fowler, 2009, p. 55). Table 17.1 assumes that, if 100 per cent had replied, 50 per cent said 'yes' and 50 per cent said 'no'; the rest of the table indicates the possible variation depending on response rate.

Table 17.1 indicates the possible variation in a simple 'yes/no' type of question. If a rating scale is chosen, for example a five-point rating scale, the number of options increases from two to five, and, correspondingly, the possibility for variation increases even further.

#### Improving response rates in a survey

A major difficult in survey research is securing a sufficiently high response rate to give credibility and reliability to the data. In some surveys, response rates can be as low as 20–30 per cent, and this compromises the reliability of the data very considerably. There is a difference between the *intended* and the *achieved* sample (Fogelman, 2002, p. 105). Punch (2003, p. 43) suggests that it is important to plan for poor response rates (e.g. by increasing the sample size) rather than trying to adjust sampling *post hoc*. He also suggests that access to the sample needs to be researched before the survey

# TABLE 17.1MAXIMUM VARIATION FOR<br/>LOW RESPONSE RATES IN A<br/>YES/NO QUESTION FOR A 50/50<br/>DISTRIBUTION

Response rate (%)	Variation in the true value of 'yes' and 'no' votes (lowest % to highest % in each category)
100	50–50
90	45–55
80	40–60
70	35–65
60	30–70
50	25–75
40	20–80
30	15–85
20	10–90
10	5–95

commences, maybe pre-notifying potential participants if deemed desirable. He argues that a poor response level may also be due to the careless omission of details of how and when the questionnaire will be returned or collected. This is a matter that needs to be made clear in the questionnaire itself. In the case of a postal survey a stamped addressed envelope should always be included.

Kenett (2006) and Fowler (2009, p. 52) report that responses rates increase when people are interested in the subject matter of the survey, or if the subject is very relevant to them, or if completing the survey brings them a sense of satisfaction. Denscombe (2009b, p. 288) reports that response rates increase if the 'respondent burden' (the effort required by the respondent to answer a question) is low.

Further, the design, layout and presentation of the survey may also exert an influence on response rate. It is important to include a brief covering letter that explains the research clearly and introduces the researcher. The timing of the survey is important, for example, schools will not welcome researchers or surveys in examination periods or at special periods, for example, Christmas or inspection times (Fogelman, 2002, p. 106).

Finally, it is important to plan the follow-up to surveys, to ensure that non-respondents are called again and reminded of the request to complete the survey. Fowler (2009, p. 57) indicates that between a quarter and a third of people may agree to completing a survey if a follow-up is undertaken.

There are several possible ways of increasing response rates to mailed surveys (Aldridge and Levine, 2001; Diaz de Rada, 2005; Fowler, 2009, p. 56; Denscombe, 2014; Dillman *et al.*, 2014; Williams *et al.*, 2016), including:

- use follow-ups and polite reminders (e.g. by mail, email, telephone call) in which the reminder is short, polite, indicating the value of the respondent's participation and, if the reminder is postal, another clean copy of the questionnaire;
- use multiple and mixed modes of responding (i.e. avoid relying on a single mode, such as post, email, website, cellphone app, interview);
- give advance notification of the survey (e.g. by telephone, post or email);
- indicate how the survey is important and the benefits from it, and how (and what) the respondents can help in answering the survey;
- indicate the institutional affiliation (with a logo) that is sponsoring or supporting the survey and support for the survey from high-status or influential persons;

- provide information about the research through a covering letter and/or advance notification;
- avoid making the survey look like junk mail;
- thank the participants in advance;
- indicate that others have already answered the survey (do not be dishonest);
- give pre-paid stamped addressed envelopes for return of the survey;
- offer incentives for return (though increasing the financial incentive to a high figure does not bring commensurate returns in response rates);
- for a follow-up reminder, include a cover page, as this increases response rates;
- make it easy to answer the survey, keeping the respondent effort and burden to a minimum;
- make the questionnaire topic interesting, the design attractive and the questions interesting, clear and easy to answer, with easy-to-follow instructions and spacing of the text. Make instructions about responses and return very clear and easy;
- keep the survey short, easy to read and complete, and very clear;
- make response modes easy: giving too many kinds can lower response rates;
- avoid open-ended questions unless these are really important (as the quality of responses is usually poor to open-ended questions: people tend not to write anything or to write very little). Avoid placing open-ended questions at the start of a questionnaire;
- consider asking the respondents for an interview to complete the survey questionnaire;
- deliver the questionnaire personally rather than through mail;
- ensure that the questions or items are nonjudgemental (e.g.in sensitive matters);
- avoid asking for sensitive or personal information unless it is absolutely necessary, particularly if asking for identifying features of children;
- indicate you own contact details, relevant and authentic professional information about yourself and how you can be reached;
- assure confidentiality, anonymity, privacy and security of information;
- send an email reminder to participants very shortly after the distribution of the survey.

Cooper and Schindler (2001, pp. 314–15) and Fowler (2009, p. 58) report that the following factors make little or no appreciable difference to response rates:

- personalizing the introductory letter;
- writing an introductory letter;
- promises of anonymity;

- questionnaire length (it is not always the case that a short questionnaire produces more returns than a long questionnaire, but researchers will need to consider the effect of a long survey questionnaire on the respondents – they may feel positive or negative about it, or set it aside temporarily and forget to return it later);
- size, reproduction and colour of the questionnaire;
- deadline dates for return (it was found that these did not increase response rate but did accelerate the return of questionnaires).

Potential respondents may be persuaded to participate depending on, for example:

- the status and prestige of the institution or researcher carrying out the research;
- the perceived benefit of the research;
- the perceived importance of the topic;
- personal interest in the research;
- interest in being interviewed, i.e. the interview experience;
- personal liking for, or empathy with, the researcher;
- feelings of duty to the public and sense of civic responsibility;
- loneliness or boredom (nothing else to do);
- sense of self-importance.

Dillman (2007) suggests that response rates can be increased if, in sequence: (a) non-respondents are sent a friendly reminder after ten days, stressing the importance of the research; (b) non-respondents are sent a further friendly reminder ten days after the initial reminder, stressing the importance of the research; (c) a telephone call is made to the respondents (if the number is known) shortly after the second reminder, indicating the importance of the research.

Fowler (2009, p. 60) suggests that the initial questionnaire might also include a statement to say that completion and return of the questionnaire will ensure that no follow-up reminders will be sent (though this may be regarded by some respondents as presumptuous).

#### 17.8 Survey sampling

Sampling is a key feature of a survey approach, and we advise readers to look closely at Chapter 12 (sampling). Researchers must take sampling decisions early in the overall planning of a survey (see Figure 17.1) in light of the population from which they want to sample, and this involves, for example:

 identifying the target population (who, how large and what are their characteristics of interest?);

- deciding whether a sample or the whole population is necessary (e.g. it may be possible to have a whole population if access and size render it feasible, such as all the staff of a school);
- the sampling frame (all those to be included in the sample);
- the sampling strategy (probability and nonprobability) and type of sample;
- sampling error;
- weighted samples for small groups (e.g. before the survey is conducted and post-stratification: after the survey has been conducted).

Often the researcher will not know the population size or heterogeneity of the characteristics of the population, and, in this event, it is advisable to have as large a sample as possible (see Chapter 12 for determining sample size).

We have already seen that due to factors of expense, time and accessibility, it is not always possible or practical to obtain measures from a population. Indeed Wilson et al. (2006, p. 352) draw attention to the tension between the need for large samples in order to conduct 'robust statistical analysis', and issues of resources such as cost and practicability (p. 353). Researchers endeavour, therefore, to collect information from a smaller group or subset of the population in such a way that the knowledge gained is representative of the total population under study, i.e. a sample. Unless researchers identify the total population in advance, it is virtually impossible for them to assess how representative the sample is which they have drawn. Chapter 12 addresses probability and nonprobability samples, and we refer readers to the detailed discussion of these in that chapter. The researcher will need to decide the sampling strategy to be used on the basis of fitness for purpose, for example:

- a probability and non-probability sample;
- the desire to generalize, and to whom;
- the sampling frame (those who are eligible to be included);
- the sample size;
- the representativeness of the sample;
- access to the sample;
- the anticipated response rate.

Even if the researcher has taken extraordinary care with the sampling strategy, there may still be problems (e.g. response rate, respondent characteristics or availability) that can interfere with the best-laid strategies.

In addition to the sampling strategy to be used, there are the issues of sample size and selection. We discussed

this in Chapter 12, but here we wish to address the issue of practicability. For example, let us say that, in the interests of precision, the researcher wishes to have a sample in which there are four strata (e.g. age groups in a primary school), and that each stratum comprised 50 students, i.e. 200 students in total. If that researcher wished to increase the sample size of one stratum by, say, 20 students, this would necessitate an overall increase of 80 students ( $20 \times 4$ ) in the sample. Do the benefits outweigh the costs here?

An alternative to increasing the *total* size of the sample would be to increase the size of one stratum only, under certain conditions. For example, let us say that the researcher is studying attitudes of males and females to learning science, in a secondary school which had only recently moved from being a single-sex boys' school to a mixed sex school, so the ratio of male to female students is 4:1. The researcher wishes to include a minimum of 200 female students. This could require a total of 1,000 students in the sample (200 females +  $\{200 \times 4\}$  male students in the sample); this could be unmanageable. Rather, the researcher could identify two female students for each male student (i.e. 400 females) and then, when analysing the data, could give one quarter of the weight to the response of the female students, in order to gain a truer representation of the target population of the school. This would bring the total sample to 600 students, rather than 1,000, involved in the survey. Oversampling a smaller group (in this case the females) and then weighting the analysis is frequently undertaken in surveys (cf. Fowler, 2009, p. 27).

In sampling, the probability might also exist of excluding some legitimate members of population in the target sample; however, the researcher will need to weigh the cost of excluding these members (e.g. the very hard to reach) against the cost of ensuring that they are included – the benefit gained from including them may not justify the time, cost and effort (cf. Fowler, 2009, p. 179). Similarly, the precision gained from stratified sampling (see Chapter 12) may not be worth the price to be paid in necessarily increasing the sample size in order to represent each stratum.

In many cases a sampling strategy may be in more than one stage. For example, let us consider the instance of a survey of 1,000 biology students from a population of 10,000 biology students in a city. In the first stage, a cluster group of, say, ten schools is identified (A), then, within that, a cluster by age group of students (B), and then, within that, the cluster of individuals in that group who are studying biology (C), and, finally, the sample (D) is taken from that group. The intention is to arrive at (D), but in order to reach this point a series of other steps has to be taken.

This raises the matter of deciding the steps to be taken. For example, the researcher could decide the sampling for the survey of the biology students by taking the random sample of 1,000 students from ten schools. The researcher lists all the 1,000 relevant students from the list of 10,000 students, and decides to select 100 students from each of the ten schools (a biology student, therefore, in one of these ten schools has a one in ten chance of being selected). Alternatively, the researcher could decide to sample from five schools only, with 200 students from each of the five schools, so students in each of these five schools have a one in five chance of being selected. Alternatively, the researcher could decide to sample from two schools, with 500 students, so students in each of these two schools have a one in two chance of being selected. There are other permutations. The point here is that, as the number of schools decreases, so does the possible cost of conducting the survey, but so does the overall reliability, as so few schools are included. It is a trade-off.

In order to reduce sampling error (the variation of the mean scores of the sample from the mean score of the population), a general rule is to increase the sample size, and this is good advice. However, it has to be tempered by the fact that the effect of increasing the sample size in a small sample reduces sampling error more than in a large sample, for example, increasing the sample size from 50 to 80 (30 persons) will have greater impact on reducing sampling error than increasing the sample size from 500 to 530 (30 persons). Hence it may be of little benefit simply to increase sample sizes in already-large samples.

The researcher has to exercise his or her judgement in attending to sampling. For example, if it is already known that a population is homogeneous, then the researcher may feel it a needless exercise in having too large and unmanageable a sample if the results are not likely to be much different from those of a small sample of the same homogeneous group (though theoretical sampling (see Chapter 37) may suggest where a researcher needs to include participants from other small samples). As Fowler (2009, p. 44) remarks, the results of a sample of 150 people will describe a population of 15,000 or 25 million with more or less the same degree of accuracy. He remarks that samples of more than 150 or 200 may yield only modest gains to the precision of the data (p. 45), though this, of course, has to be addressed in relation to the population characteristics, the number, size and kind of strata to be included, and the type of sample being used. Sampling errors, he notes (p. 45) are more a function of sample size than of the proportions of the sample to the population. Further, he advocates probability rather than non-probability samples, unless there are persuasive reasons for non-probability samples to be used.

Whilst sample sizes can be calculated on the basis of statistics alone (e.g. confidence levels, confidence intervals, population size, statistical power and so on, see Chapter 12), this is often not the sole criterion, as it accords a degree of precision to the sample which takes insufficient account of other sampling issues, for example, access, variation or homogeneity in the population, levels of literacy in the population (e.g. in the case of a self-administered questionnaire survey), number and type of variables and costs.

Sampling is one of several sources of error in surveys, as indicated earlier in this chapter.

#### 17.9 Longitudinal and cross-sectional surveys

The term 'longitudinal' describes a variety of studies that are conducted over a period of time. A clear distinction is drawn between longitudinal and crosssectional studies. The longitudinal study gathers data over an extended period of time: a short-term investigation may take several weeks or months; a long-term study can extend over many years. Where successive measures are taken at different points in time from the same respondents, the term 'follow-up study' or 'cohort study' is used in the British literature, the equivalent term in the US being the 'panel study'. The term 'cohort' is a group of people with some common characteristic.

Where different respondents are studied at one or more different points in time, the study is called 'crosssectional', i.e. a cross-section of the population is taken to investigate the topic(s) of interest. Where a few selected factors are studied continuously over time, the term 'trend study' is employed. One example of regular or repeated cross-sectional social surveys is the General Household Survey, in which the same questions are asked every year, though they are put to a different sample of the population each time. The British Social Attitudes Survey is an example of a repeated crosssectional survey, using some 3,600 respondents.

A famous example of a longitudinal (cohort) study is the UK's National Child Development Study, which started in 1958. The British General Household Panel Survey interviewed individuals from a representative sample each year in the 1990s. Another example is the British Family Expenditure Survey. These latter two are cross-sectional in that they tell us about the population at a given point in time, and hence provide aggregated data.

By contrast, longitudinal studies can also provide individual-level data, by focusing on the same individuals over time (e.g. the Household Panel Studies which follow individuals and families over time (Ruspini, 2002, p. 4). Lazarsfeld introduced the concept of a panel in the 1940s, attempting to identify causal patterns and the difficulties in tracing these (Ruspini, 2002, p. 13)).

#### Longitudinal studies

Longitudinal studies can be of the survey type or of other types (e.g. case study). Here we confine ourselves to the survey type. Longitudinal studies can include trend studies, cohort studies and panel studies (Creswell, 2012), and we discuss these below. A useful centre for longitudinal studies in education is at the University of London: www.cls.ioe.ac.uk/default.aspx.

Longitudinal studies can use repeated crosssectional studies, which are conducted regularly, each time with a largely different sample or, indeed, an entirely new sample (Ruspini, 2002, p. 3), or use the same sample over time. They enable researchers to: 'analyse the duration of social phenomena' (p. 24); highlight similarities, differences and changes over time in respect of one or more variables or participants (within and between participants); identify long-term ('sleeper') effects; and explain changes in terms of stable characteristics, for example sex, or variable characteristics, such as income. The appeal of longitudinal research is its ability to establish causality and to make inferences. Ruspini adds to these the ability of longitudinal research to 'construct more complicated behavioural models than purely cross-sectional or time-series data' (p. 26); they can catch the complexity of human behaviour. Further, longitudinal studies can combine numerical and qualitative data.

Retrospective analysis is not confined to longitudinal studies alone. For example, Rose and Sullivan (1993, p. 185) and Ruane (2005, p. 87) suggest that cross-sectional studies can use retrospective factual questions, for example, previous occupations, dates of birth within the family, dates of marriage and/or divorce, though Rose and Sullivan (1993, p. 185) advise against collecting other types of retrospective data in cross-sectional studies, as the quality (e.g. reliability) of the data diminishes the further back one asks respondents to recall previous states or even facts.

It is important in longitudinal studies to decide when and how frequently to collect data over time, and this is informed by issues of fitness for purpose as well as practicability. Further, in order to allow for attrition (dropout) of the sample, it is wise to have as large a sample as practicable and possible at the start of the study (Wilson *et al.*, 2006, p. 354).

#### Cohort studies

A cohort study focuses on a specific population in which all its members have the specific defining characteristic that is of interest to the researcher (e.g. the National Child Development Study in the UK; the Millennium Cohort Study). In a cohort study the specific population is tracked over a specific period of time but selective sampling within that sample occurs. This means that different members of a cohort are included each time. For example, the population might be eighteen-year-olds in the UK; at one time point (say, when they are twentyfive years old) the population might be sampled, and then at another time point (say, when they are thirty-five) the same population might be sampled but different members of the population will be in the sample.

Cohort studies and trend studies can be *prospective* longitudinal methods, in that they are ongoing in their collection of information about individuals or their monitoring of specific events. *Retrospective* longitudinal studies, on the other hand, focus upon individuals who have reached some defined end-point or state. For example, a group of young people may be the researcher's particular interest (intending social workers, convicted drug offenders or university dropouts, for example), with questions such as: 'Is there anything about your previous experience that can account for your present situation?' Retrospective longitudinal studies will specify the period over which to be retrospective, for example, one year, five years.

#### Panel studies

In contrast to a cohort study, in a panel study exactly the same individuals are tracked over time. An example of this is the Panel Study of Income Dynamics in the US. Another example from the UK is the '7 Up' study which started in 1964 and tracks a small group of individuals every seven years, yielding insights into social and cultural stratification, reproduction and the selffulfilling prophecy.

Whilst this type of study has the attraction of tracking the same people over time, this same requirement also has its disadvantages in terms of keeping contact with those individuals and addressing attrition. Panel studies are useful for investigating causality and change over time.

#### Trend studies

Trend studies focus on factors (e.g. mathematics performance) rather than people, and these factors are studied over time. New samples – different people – are drawn at each stage of the data collection, but focus on the same factors, and if random samples are used, they can be representative of the wider population. By taking different samples the problem of reactivity is reduced (see below: 'pre-test sensitisation'), i.e. earlier surveys affecting the behaviour of participants in the later surveys. This is particularly useful if the research is being conducted on sensitive issues, as raising a sensitive issue early on in the research may change an individual's behaviour, which could affect the responses in a later round of data collection. By drawing a different sample each time, this problem is overcome.

Trend or prediction studies have an obvious importance to educational administrators or planners. Like cohort studies, they can be of relatively short or long duration. Essentially, the trend study examines recorded data to establish patterns of change that have already occurred in order to predict what will be likely to occur in the future. In trend studies, two or more crosssectional studies are undertaken with identical age groups at more than one point in time in order to make comparisons over time (e.g. the Scholastic Aptitude and Achievement tests in the US and the National Assessment of Educational Progress results). A major difficulty that researchers face in conducting trend analyses is the intrusion of unpredictable factors that invalidate forecasts formulated on past data. For this reason, shortterm trend studies tend to be more accurate than longterm analyses. Trend studies do not include the same respondents over time, so the possibility exists for variation in data due to the different respondents rather than the change in trends. Gorard (2001b, p. 87) suggests that this problem can be attenuated by a 'rolling sample' in which a proportion of the original sample is retained in the second wave of data collection, and a proportion of this sample is retained in the third wave, and so on.

#### **Cross-sectional studies**

A cross-sectional study is one that produces a 'snapshot' of a population at one particular point in time. The epitome of the cross-sectional study is a national survey in which a representative sample of the population consisting of individuals of different ages, different occupations, different educational and income levels, and residing in different parts of the country, is interviewed on the same day. In education, crosssectional studies can involve indirect measures of the nature and rate of changes in the physical and intellectual development of samples of children drawn from representative age levels. The single 'snapshot' of the cross-sectional study provides researchers with data for either a retrospective or a prospective enquiry. A cross-sectional study can also bear several hallmarks of a longitudinal study of parallel groups (e.g. age groups) which are drawn simultaneously from the population. For example, drawing students aged five, seven, nine and eleven at a single point in time would bear some characteristics of a longitudinal study in that developments over age groups could be seen, though, of course, it would not have the same weight as a longitudinal study conducted on the same age group over time. This is the case for international studies of educational achievement, requiring samples to be drawn from the same population (Lietz and Keeves, 1997, p. 122) and for factors that might influence changes in the dependent variables to remain constant across the age groups.

Cross-sectional studies, catching a frozen moment in time, may be ineffective for studying change or causality. If changes are to be addressed through crosssectional surveys, then this suggests the need for repeated applications of the survey, or the use of trend analysis.

The main types of longitudinal study are illustrated in Figure 17.2.

## 17.10 Strengths and weaknesses of longitudinal, cohort and cross-sectional studies

Longitudinal studies of the cohort analysis type have an important place in the armoury of educational researchers. Longitudinal studies have considerable potential for yielding rich data that can trace changes over time, and with great accuracy (Gorard, 2001b, p. 86). On the other hand, they suffer from problems of attrition (participants leaving the research over time, a particular problem in panel studies which research the same individuals over time), and they can be expensive to conduct in terms of time and money (Ruspini, 2002, p. 71). Gorard (2001b) reports a study of careers and identities that had an initial response rate of between 60 and 70 per cent in the first round, and then risked dropping to 25 per cent by the third round, becoming increasingly more middle class in each wave of the study; the same publication discusses a Youth Cohort Study in which only 45 per cent of the respondents took part in all three waves of the data collection. Ruspini (2002, p. 72) identifies an attrition rate of 78 per cent in the three waves of the European Community Household Panel survey of the UK in 1997.

Ruspini also indicates how a small measurement error in a longitudinal study may be compounded over time. She gives the example of an error in income occurring at a point in time (p. 72) that could lead to 'false transitions' appearing over time in regard to poverty and unemployment.

Further, long-term studies, Gorard (2001b, p. 86) avers, face 'a threat to internal validity' that stems from the need 'to test and re-test the same individuals'. Dooley (2001, p. 120) terms this 'pre-test sensitisation'; it is also termed 'panel conditioning' or 'time-in sample bias' (Ruspini, 2002, p. 73). Here the first interview in an interview survey can cause changes in the second interview, i.e. the first interview might set up a self-fulfilling prophecy that is recorded in the second interview. He gives the example of a health survey in the first round of data collection, which may raise



participants' awareness of the dangers of smoking, such that they reduce or give up smoking by the time the second round takes place. Trend studies overcome this problem by drawing different populations at each round of data collection.

Dooley (2001) also identifies difficulties caused by changes in the research staff over time in longitudinal surveys. Changes in interviewee response, he suggests, may be due to having different researchers rather than to the respondents themselves. Even using the same instruments, different researchers may use them differently (e.g. in interviewing behaviour).

To add to these matters, Ruspini (2002, p. 73) suggests that longitudinal data are affected by:

- history (events occurring may change the observations of a group under study);
- maturation (participants mature at different speeds and in different ways);
- testing (test sensitization may occur participants learn from exposure to repeated testing/interviews);
- the timing of cause and effect (some causes may produce virtually instantaneous effects and others may take a long time for the effects to show);
- the direction of causality not always being clear or singular.

A major concern in longitudinal studies concerns the comparability of data over time. For example, though public examinations may remain constant over time (e.g. GCSE, A levels), the contents and format of those examinations do not. (This rehearses the argument that public examinations are becoming easier over time.) This issue concerns the need to ensure consistency in the data-collection instruments over time. Further, if comparability of data in a longitudinal study is to be addressed then this means that the initial rounds of data collection will need to anticipate and include all the variables that will be addressed over time.

Longitudinal studies are more prone to attrition than cross-sectional studies, and are more expensive to conduct in terms of time and cost. On the other hand, whereas trend studies change their populations, thereby disabling micro-level – individual-level – analysis from being conducted, longitudinal analysis enables such individual-level analysis to be performed. Indeed whereas cross-sectional designs (even if they are repeated cross-sectional designs) may be unsuitable for studying developmental patterns and causality within cohorts, in longitudinal analysis this is a strength. Longitudinal data can supply 'satisfactory answers to questions concerning the dynamics and the determinants of individual behaviour' (Ruspini, 2002, p. 71), issues which are not easily addressed in cross-sectional designs.

Retrospective longitudinal studies rely on participants' memories which may be faulty; the further back one's memory reaches, the greater is the danger of distortion or inability to recall. Memory is affected by, for example (Ruspini, 2002, p. 97):

- the time that has elapsed since the event took place;
- the significance of the event for the participant;
- the amount of information required for the study the greater the amount, the harder it is to provide;
- the contamination/interference effect of other memories of a similar event (i.e. the inability to separate similar events);
- the emotional content or the social desirability of the content;
- the psychological condition of the participant at interview.

Further, participants will look at past events through the lens of hindsight and subsequent events rather than what those events meant at the time. Moreover, it is not always easy for these participants to recall their emotional state at the time in question. Factually speaking, it may not be possible to gather data from some time past, as they simply do not exist (e.g. medical records, data on income) or they cannot be found, recovered or accessed.

Cohort studies of human development conducted on representative samples of populations are uniquely able to identify typical patterns of development and to reveal factors operating on those samples which elude other research designs. They permit researchers to examine individual variations in characteristics or traits, and to produce individual development curves. Cohort studies, too, are particularly appropriate when investigators attempt to establish causal relationships, as this involves identifying changes in certain characteristics which result in changes in others.

Cross-sectional designs are inappropriate in causal research as they cannot sustain causal analysis unless they are repeated over time, as causality has a necessary time dimension. Cohort analysis is especially useful in sociological research because it can show how changing properties of individuals fit together into changing properties of social systems as a whole. For example, the study of staff morale and its association with the emerging organizational climate of a newly opened school would lend itself to this type of developmental research. A further strength of cohort studies in schools is that they provide longitudinal records whose value takes account of the known fallibility of any single test or assessment. Finally, time, often a limiting factor in experimental and interview settings, is generally more readily available in cohort studies, allowing the researcher greater opportunity to observe trends and to distinguish 'real' changes from chance occurrences (see Bailey, 1994).

In longitudinal, cohort and trend studies the characteristics of respondents are likely to affect results (Robson, 1993, p. 128). For example, their memory, knowledge, motivation and personality may affect their responses, and indeed they may withhold information, particularly if it is sensitive.

Longitudinal research indicates the influence of biological factors over time (e.g. human development), environmental influences and intervention influences (Keeves, 1997a, p. 139) and their interactions. Addressing these, the appeal of longitudinal analysis is that it enables researches to conduct causal analysis. Time series studies in longitudinal research also enable emergent patterns to be observed over time, by examining a given range of variables over time, in addition to other factors. This enables individual and group profiles to be examined over time and development, indicating similarities and differences within and between individuals and groups in respect of given variables.

Longitudinal studies suffer several disadvantages (though the gravity of these weaknesses is challenged by supporters of cohort analysis). The disadvantages are, first, that they are time-consuming and expensive, because the researcher is obliged to wait for growth data to accumulate. Second, there is the difficulty of sample mortality. Inevitably during the course of a long-term cohort study, subjects drop out, are lost or refuse further cooperation. Such attrition makes it unlikely that those who remain in the study are as representative of the population as the original sample. Sometimes attempts are made to lessen the effects of sample mortality by introducing aspects of cross-sectional study design, that is, 'topping up' the original cohort sample size at each time of retesting with the same number of respondents drawn from the same population. The problem here is that differences arising in the data from one survey to the next may then be accounted for by differences in the persons surveyed rather than by genuine changes or trends.

A third difficulty has been termed the 'control effect' (sometimes referred to as 'measurement effect'). Often, repeated interviewing results in an undesired and confusing effect on the actions or attitudes under study, influencing the behaviour of subjects, sensitizing them to matters that have hitherto passed unnoticed, or stim-

ulating them to communicate with others on unwanted topics (see Riley, 1963). Fourth, cohort studies can suffer from the interaction of biological, environmental and intervention influences (Keeves, 1997a, p. 139). Finally, cohort studies in education pose considerable problems of organization due to the continuous changes that occur in pupils, staff, teaching methods and the like. Such changes make it highly unlikely that a study will be completed in the way that it was originally planned.

Cohort studies, as we have seen, are particularly appropriate in research on human growth and development. Why then are so many studies cross-sectional rather than cohort studies? The reason is that they have a number of advantages over cohort studies: they are less expensive; they produce findings more quickly; they are less likely to suffer from control effects; and they are more likely to secure the cooperation of respondents on a 'one-off' basis. Generally, crosssectional designs are able to include more subjects than are cohort designs.

The strengths of cohort analysis are the weaknesses of the cross-sectional design. The cross-sectional study is a less effective method for the researcher who is concerned to identify individual variations in growth or to establish causal relationships between variables. Sampling in a cross-sectional study is complicated because different subjects are involved at each age level and may not be comparable. Further problems arising out of selection effects and obscuring irregularities in growth weakens the cross-sectional study so much that one observer dismisses the method as a highly unsatisfactory way of obtaining developmental data except for the crudest purposes.

Douglas (1976a), who pioneered the first national cohort study in any country, makes a spirited defence of the method against the common criticisms that are levelled against it – that it is expensive and time-consuming. His account of the advantages of cohort analysis over cross-sectional designs is summarized in Box 17.1.

Cross-sectional studies require attention to sampling in order to ensure that the information on which the sample is based is comprehensive (Lietz and Keeves, 1997, p. 124). Further, there is a risk that some potential participants may decline to take part, thereby weakening the sample, or some may not answer specific questions or, wittingly or unwittingly, give incorrect answers. Measurement error may also occur if the instrument is faulty, for example, using inappropriate metrics or scales.

The comparative strengths and weaknesses of longitudinal studies (including retrospective studies),

#### BOX 17.1 ADVANTAGES OF COHORT OVER CROSS-SECTIONAL DESIGNS

- 1 Some types of information, for example, on attitudes or assessment of potential ability, are only meaningful if collected contemporaneously. Other types are more complete or more accurate if collected during the course of a longitudinal survey, though they are likely to have some value even if collected retrospectively, for example, length of schooling, job history, geographical movement.
- 2 In cohort studies, no duplication of information occurs, whereas in cross-sectional studies the same type of background information has to be collected on each occasion. This increases the interviewing costs.
- 3 The omission of even a single variable, later found to be important, from a cross-sectional study is a disaster, whereas it is usually possible in a cohort study to fill the gap, even if only partially, in a subsequent interview.
- 4 A cohort study allows the accumulation of a much larger number of variables, extending over a much wider area of knowledge than would be possible in a cross-sectional study. This is of course because the collection can be spread over many interviews. Moreover, information may be obtained at the most appropriate time, for example, information on job entry may be obtained when it occurs even if this varies from one member of the sample to another.
- 5 Starting with a birth cohort removes later problems of sampling and allows the extensive use of subsamples. It also eases problems of estimating bias and reliability.
- 6 Longitudinal studies are free of one of the major obstacles to causal analysis, namely, the reinterpretation of remembered information so that it conforms to conventional views on causation. It also provides the means to assess the direction of effect.

Source: Adapted from Douglas (1976b)

cross-sectional analysis and trend studies are summarized in Table 17.2 (see also Rose and Sullivan, 1993, pp. 184–8).

Several of the strengths and weaknesses of retrospective longitudinal studies share the same characteristics as those of *ex post facto* research, discussed in Chapter 20.

### 17.11 Postal, interview and telephone surveys

#### Postal surveys

There are strengths and difficulties with postal and interview surveys. Postal surveys can reach a large number of people, gather data at comparatively low cost and quite quickly, and can give assurances of confidentiality (Robson, 1993; Bailey, 1994, p. 148; Dillman *et al.*, 2014). Similarly they can be completed at the respondents' own convenience and in their preferred surroundings and own time; this can enable them to check information, if necessary (e.g. personal documents), and think about responses. As standardized wording is used, there is a useful degree of comparability across the responses, and, as no interviewer is present, there is no risk of interviewer bias. Further, postal questionnaires enable widely scattered populations to be reached.

Postal surveys can also be used to gather detailed sensitive qualitative data (Beckett and Clegg, 2007),

not least because the non-presence of another person (e.g. an interviewer) can increase the honesty and richness of the data, whereas the presence of an interviewer might inhibit the respondent. Further, in a postal survey, the relations of power between the researcher and the respondent are often more equal than in an interview situation (in which the former often controls the situation more than the latter) (p. 308).

On the other hand, postal surveys typically suffer from a poor response rate, even though Dillman *et al.* (2014) comment they have moved from having the lowest response rate to having response rates higher than telephone surveys. Mailed surveys are reported to have an approximately 20 per cent response rate, which is far lower than telephone and face-to-face surveys (Colorado State University, 2016). Diaz de Rada and Dominguez (2015) note that postal surveys feature greater acquiescence than other kinds of survey, with more unanswered questions.

Because researchers may not have any information about non-respondents, they may not know whether the sample is representative of the wider population. Further, respondents may not take the care required to complete the survey carefully, and, indeed, may misunderstand the questions. There is no way of checking this. Bailey (1994, p. 149) suggests that the very issues that make postal surveys attractive might also render them less appealing, for example:

	CROSS-SECTIONAL, TREND ANALYSIS AND RETROSPECTIVE LONGITUDINAL, STUDIES		
Study type	Features	Strengths	Weaknesses
Longitudinal studies (cohort/panel studies)	<ol> <li>Single sample over extended period of time.</li> <li>Enables the same individuals to be compared over time (diachronic analysis).</li> <li>Micro-level analysis.</li> </ol>	<ol> <li>Useful for establishing causal relationships and for making reliable inferences.</li> <li>Shows how changing properties of individuals fit into systemic change.</li> <li>Operates within the known limits of instrumentation employed.</li> <li>Separates real trends from chance occurrence.</li> <li>Brings the benefits of extended time frames.</li> <li>Useful for charting growth and development.</li> <li>Gathers data contemporaneously rather than retrospectively, thereby avoiding the problems of selective or false memory.</li> <li>Economical in that a picture of the sample is built up over time.</li> <li>In-depth and comprehensive coverage of a wide range of variables, both initial and emergent – individual specific effects and population heterogeneity.</li> <li>Enables the dynamics of change to be caught, the flows into and out of particular states and the transitions between states.</li> <li>Individual level data are more accurate than macro-level, cross- sectional data.</li> <li>Sampling error reduced as the study remains with the same sample over time.</li> <li>Enables clear recommendations for intervention to be made.</li> </ol>	<ol> <li>Time-consuming – it takes a long time for the studies to be conducted and the results to emerge.</li> <li>Problems of sample mortality heighten over time and diminish initial representativeness.</li> <li>Control effects – repeated interviewing of the same sample influences their behaviour.</li> <li>Intervening effects attenuate the initial research plan.</li> <li>Problem of securing participation as it involves repeated contact.</li> <li>Data, being rich at an individual level, are typically complex to analyse.</li> </ol>

## 

continued

<b>TABLE 17.2</b>	CONTINUED		
Study type	Features	Strengths	Weaknesses
Cross- sectional studies	<ol> <li>Snapshot of different samples at one or more points in time (synchronic analysis).</li> <li>Large-scale and representative sampling.</li> <li>Macro-level analysis.</li> <li>Enables different groups to be compared.</li> <li>Can be retrospective and/ or prospective.</li> </ol>	<ol> <li>Comparatively quick to conduct.</li> <li>Comparatively cheap to administer.</li> <li>Limited control effects as subjects only participate once.</li> <li>Stronger likelihood of participation as it is for a single time.</li> <li>Charts aggregated patterns.</li> <li>Useful for charting population-wide features at one or more single points in time.</li> <li>Enable researchers to identify the proportions of people in particular groups or states.</li> <li>Large samples enable inferential statistics to be used, e.g. to compare sub-groups within the sample.</li> </ol>	<ol> <li>Do not permit analysis of causal relationships.</li> <li>Unable to chart individual variations in development or changes, and their significance.</li> <li>Sampling not entirely comparable at each round of data collection as different samples are used.</li> <li>Can be time-consuming as background details of each sample have to be collected each time.</li> <li>Omission of a single variable can undermine the results significantly.</li> <li>Unable to chart changing social processes over time.</li> <li>They only permit analysis of overall, <i>net</i> change at the macro-level through aggregated data.</li> </ol>
Trend analysis	<ol> <li>Selected factors studied continuously over time.</li> <li>Uses recorded data to predict future trends.</li> </ol>	<ol> <li>Maintains clarity of focus throughout the duration of the study.</li> <li>Enables prediction and projection on the basis of identified and monitored variables and assumptions.</li> </ol>	<ol> <li>Neglects influence of unpredicted factors.</li> <li>Past trends are not always a good predictor of future trends.</li> <li>Formula-driven, i.e. could be too conservative or initial assumptions might be erroneous.</li> <li>Neglects the implications of chaos and complexity theory, e.g. that long- range forecasting is dangerous.</li> <li>The criteria for prediction may be imprecise.</li> </ol>
Retrospective longitudinal studies	<ol> <li>Retrospective analysis of history of a sample.</li> <li>Individual- and micro-level data.</li> </ol>	<ol> <li>Useful for establishing causal relationships.</li> <li>Clear focus (e.g. how did this particular end state or set of circumstances come to be?).</li> <li>Enables data to be assembled that are not susceptible to experimental analysis.</li> </ol>	<ol> <li>Remembered information might be faulty, selective and inaccurate.</li> <li>People might forget, suppress or fail to remember certain factors.</li> <li>Individuals might interpret their own past behaviour in light of their subsequent events, i.e. the interpretations are not contemporaneous with the actual events.</li> <li>The roots and causes of the end state may be multiple, diverse, complex, unidentified and unstraightforward to unravel.</li> <li>Simple causality is unlikely.</li> <li>A cause may be an effect and vice versa.</li> <li>It is difficult to separate real from perceived or putative causes.</li> <li>It is seldom easily falsifiable or confirmable.</li> </ol>

Copyright © 2017. Routledge. All rights reserved.

- the standardization of wording;
- the inability to catch anything other than a verbal response;
- the lack of control over the environment in which the survey questionnaire is completed;
- the lack of control over the order in which the questions are read and answered;
- the risk that some questions will not be answered;
- the inability to record spontaneous answers;
- the difficulty in separating non-response from bad response, the former being where the intended respondent receives the survey but does not reply to it, and the latter being where the intended recipient does not receive the survey, for example, because she/he has moved house;
- the need for simplicity in format as there is no interviewer present to guide the respondent through a more complex format.

Postal surveys are an example of self-administered surveys. The anonymity and absence of face-to-face interaction between the interviewer and the respondent can render these useful for asking sensitive questions (Strange *et al.*, 2003, p. 337), though Fowler (2009, p. 74) also counsels that sensitive questions can sometimes be handled better in private face-to-face interviews. In self-administered surveys, Fowler (2009, p. 72) remarks that it is advisable to keep to closed questions and make the response categories simple and explicit (e.g. ticking a box). If open questions are to be asked then, he indicates, it is better to gather the survey data in a face-to-face interview.

Further, Diaz de Rada (2005) reports that the design, size and colour of the paper used in postal surveys affects response rates. Small-sized questionnaires were mostly returned by males and those under sixty-four years of age (p. 69), whilst larger-sized questionnaires were mostly returned by females and those over the age of sixty-five (p. 70). He recommends using paper size  $14.85 \times 21$  cm (i.e. a sheet of A4-sized paper folded in half), white paper, and including a cover page (p. 73) (though this inevitably increases the number of pages in a questionnaire, and this can be off-putting for respondents). He reports that paper size has no effect on the quality of the responses.

#### Interview surveys

Whereas postal surveys are self-administered, interview surveys are supervised and hence potentially prone to fewer difficulties. Interview methods of gathering survey data are useful in that the presence of the interviewer can help clarify queries from the respondents and can stimulate the respondent to give full answers to an on-the-spot researcher rather than an anonymous researcher known only through an introductory letter (Robson, 1993). Indeed face-to-face encounters can improve response rates. Further, as interviews can be flexible, questioners are able both to probe and to explain more fully (Bailey, 1994, p. 174). Interviews are also useful when respondents have problems with reading and writing. Using non-verbal behaviour to encourage respondents to participate is also possible. Moreover, with interviews there are greater opportunities to control the environment in which the survey is conducted, particularly in respect of privacy, noise and external distractions.

The effective interviewer, Fowler (2009, p. 128) claims, is business-like and assertive whilst being engaging, friendly and kind. Fowler argues for great care with choosing interviewers and training them, as much can hang on their behaviour.

The potential for trust, rapport and cooperation between the interviewer and the respondent is strong in face-to-face encounters (Dooley, 2001, p. 122; Gwartney, 2007, p. 16). Further, interviewers can either ensure that the sequence of the survey protocol is strictly adhered to or they can tailor the order of responses to individual participants, making certain that all questions are answered. Interview surveys, moreover, can guarantee that it is the respondent alone who answers the questions, whereas in postal surveys the researcher never knows what help or comments are solicited from, or given by, other parties. Bailey (1994) adds that the opportunity for spontaneous behaviour and responses is also possible in interview surveys, and interviews can use more complex structures than postal questionnaires, the researcher being on hand to take participants through the items.

On the other hand, the very features which make interview methods attractive may also make them problematic. For example, interview survey methods may be affected by the characteristics of the interviewer (e.g. sex, race, age, ethnicity, personality, skills, perceived social status, clothing and appearance). They may also be affected by the conduct of the interview itself (e.g. rapport between the interviewer and the interviewee), and interviewees may be reluctant to disclose some information if they feel that the interview will not be anonymous or if sensitive information is being requested. The flexibility which the interview gives also contributes to the potential lack of standardization of the interview survey, and this may render consistency, and thereby reliability, a problem.

Interview surveys are costly in time for the researcher and the interviewee, and, as they are conducted at a fixed time, they may prevent the interviewee from consulting records that may be important to answer the questions. Further, they may require the interviewer to travel long distances to reach interviewees, which can be expensive both in time and travel costs (Bailey, 1994, p. 175). If interviews are intended to be conducted in the participants' own homes, then participants may be unwilling to admit strangers. Moreover, neighbourhoods may be dangerous for some researchers to visit (e.g. a white researcher with a clipboard going into a non-white area of great deprivation, or a black researcher going into a conservative white area).

#### **Telephone surveys**

Telephone surveys are located between mailed questionnaires and personal interviews (Arnon and Reichel, 2009). Dillman *et al.* (2014) note the rapid decline in telephone interviewing (p. 11) with the reduction in landlines, the rise in cellphones, the lack of listing of call numbers and the rise in screening callers. However, telephone interviews have the attraction of overcoming bias in the researcher or the interviewee that may be caused by social characteristics or matters of age, dress, race, ethnicity, appearance etc. (e.g. Gwartney, 2007, p. 16). Indeed Denscombe (2014) suggests that people are 'more honest and open' on the phone than in a postal questionnaire (p. 12).

Telephone surveys require the interviewer to be an articulate, clear speaker and a good listener, and able to key in interviewee responses onto a computer whilst listening and speaking (Denscombe, 2014, pp. 42-3). They have the advantage of reducing costs in time and travel, for when a potential respondent is not at home, a call-back is cheap and the time to redial is short (Dooley, 2001, p. 122; Arnon and Reichel, 2009, p. 179), and, using Internet services such as Skype, telephone surveys can be almost free of charge and include face-to-face viewing. Revisits to often distant locations, on the other hand, can incur considerable expense in time and travel. Furthermore, if the intended participant is unable or unwilling to respond, then it is a relatively easy matter to maintain the required sample size by calling a replacement. Again, where respondents are unable or unwilling to answer all the questions required, then their partial replies may be discarded and further substitutes sought from the sample listing. It is easy to see why telephone interviews must always have a much longer list of potential respondents in order to attain the required sample size.

Not everyone has a telephone (e.g. the poor, the young) and this may lead to a skewed sample (Arnon and Reichel, 2009, p. 179). Nor, for that matter, is everyone available for interview, particularly if they work. Further, many people are 'ex-directory', i.e. their numbers are withheld from public scrutiny. In addition, Dooley (2001, p. 123) reports that younger, single and higher occupational status groups use electronic facilities that screen out and delete researchers' calls and these could lead to a skewed sample. Indeed Fowler (2009, p. 75) indicates that telephone surveys tend to elicit more socially desirable answers than face-to-face interviews.

Even when the telephone is answered, the person responding may not be the most suitable one to take the call; she/he may not know the answer to the questions or have access to the kind of information required. For example, in an inquiry about household budgets, the respondent may simply be ignorant about a family's income or expenditure on particular items. A child may answer the call, or an elderly person who may not be the householder. Interviewers will need to prepare a set of preliminary screening questions or arrange a callback time when a more appropriate person can be interviewed.

Telephone interviewing has its own strengths and weaknesses. For example, more often than not a respondent's sex will be clear from their voice, so some questions may be unnecessary or inappropriate. On the other hand, it is unwise to have several multiple choices in a telephone interview, as respondents will simply forget the categories available, there being no written prompts to which the respondent can refer.

Similarly, order effects can be high: items appearing early in the interview exert an influence on responses to later ones, whilst items appearing early in a list of responses may be given greater consideration than those occurring later, a matter not confined to telephone surveys but to questionnaires in general. Dooley (2001, p. 136) indicates a 17 per cent difference in agreement recorded to a general statement question when it appeared before rather than after a specific statement, and other research demonstrates that responses to particular questions are affected by questions surrounding them. His advice is to ask general questions before spe*cific* ones, otherwise the general questions are influenced by earlier responses to specific questions. Once again, this is a matter not confined to telephone surveys but to questionnaires in general.

Further, if the questioning becomes too sensitive, respondents may simply hang up in the middle of the survey interview, tell lies or withhold information. Dooley (2001, p. 123) reports that, in comparison to face-to-face interviews, telephone respondents tend to produce more missing data, to be more evasive, more acquiescent (i.e. they tend to agree more with statements) and more extreme in their responses (e.g. opting for the extreme ends of rating scales).

Fowler (2009, pp. 73-4) also indicates that, in a telephone survey, it is unwise to have too many response scale points, that it is better to avoid long lists of items and that it is advisable to read the statement before indicating the response categories, unless a long list of items is to be given (i.e. is unavoidable), in which case he suggests that it is better to read and re-read the response categories to the respondent before the list of statements. All of these points take account of the limits of the short-term memories on which respondents often rely in a telephone interview. He also suggests (p. 73) that complex questions can be approached in a staged manner. For example, if a researcher wishes to ask about a ten-category item (e.g. income level of the teacher), then the researcher could start with a general question (e.g. above or below a particular figure), and then, once that category has been identified, proceed to a sub-category, for example, between such-and-such a figure; this avoids overload of asking a respondent to remember ten categories.

Because telephone interviews lack the sensory stimulation of visual or face-to-face interviews or written instructions and presentation, it is unwise to plan a long telephone survey call. Ten to fifteen minutes is often the maximum time tolerable to most respondents, and indeed fifteen minutes for many people is too long. This means that careful piloting must take place in order to include those items, and only those items, that are necessary for the research. The risk to reliability and validity is considerable, as the number of items may be fewer than in other forms of data collection.

Procedures for telephone interviews also need to be decided (Gwartney, 2007), for example:

- how many times to let the telephone ring before conceding that there is nobody to answer the call (Gwartney (2007, p. 99) suggests eight rings);
- how to introduce the caller and the project;
- what to say and how to introduce items and conduct the interview;
- how to determine who is receiving the call and whether he/she is the appropriate person to answer the call;
- whether to leave a message on an answerphone/ voicemail/call-back facility and, if so, what that message will be;
- how to handle language problems (e.g. which language is being used, meanings/explanations/ vocabulary);
- how to handle the situation if the receiver asks to call back later;
- what to say and how to control the caller's voice/ tone/pitch/speed/pace of questions/repetitions/language/intonation/register;

- the caller's pronunciation, enunciation and reading out loud;
- the caller's ability to clarify, summarize, reiterate, probe (and when to stop probing), prompt (if the receiver does not understand), confirm, affirm, respond, give feedback, encourage respondents, keep respondents focused and to the point;
- how to conduct closed and open questions, sensitive, factual and opinion-based questions;
- how to indicate the nature and format of the responses sought;
- the caller's ability to handle the called person's initial hostility, refusal, reluctance to take part, feelings of invasion of privacy, lack of interest, reluctance to disclose information, feelings of being harassed or singled out, anger, antagonism, lack of interest, incomplete answers, hurriedness to complete, slowness or hesitancy, mistrust, rudeness, abusive responses, or simply saying that they are too busy;
- the caller's ability to remain neutral, impartial and non-judgemental;
- how to record responses;
- how to end the interview.

It is also advisable, in order to avoid the frequent responses to 'cold-calling' (where the called person simply slams down the telephone), for the interviewer to contact the person in advance of the call, perhaps by mail, to indicate that the call will come, when, what it is about, and to ask for the recipient's cooperation in the project.

Many of the features of telephone interviewing are similar to those of effective interviewing per se, and we advise the reader to consult the comments on interviewing earlier and also in Chapter 25.

### 17.12 Comparing methods of data collection in surveys

Aldridge and Levine (2001, pp. 51–4) and Fowler (2009, pp. 80–3) offer useful summary guides to the advantages and disadvantages of several methods of data collection in surveys: personal face-to-face interviewing; telephone interviewing; self-administered/ self-completion versus interviewer-administered; group administered; mailed surveys; delivered (distributed) surveys (e.g. personally delivered or delivered to an institution); Internet surveys. We refer the reader to these useful sources.

Additionally, Fowler (2009) and Dillman *et al.* (2014) discuss the benefits of combining methods of data collection (e.g. face-to-face interviews with telephone interviews, Internet surveys with postal surveys, advance

### TABLE 17.3 ADVANTAGES AND DISADVANTAGES OF DATA-COLLECTION METHODS IN SURVEYS

	Advantages	Disadvantages	Either advantages or disadvantages
Postal	Time to think Costs may not be too expensive Opportunity for attractive survey design and graphics Complete at respondent's convenience, with opportunity for respondent to check Can reach many people No risk of interviewer intrusion or bias Can reach scattered populations Can gather sensitive data (nobody else is present) Can offer secure confidentiality, anonymity and non-traceability Standardized wording	Cost: printing, postage Time: response time and data entry Low response rates Need for contact details Risk of superficial coverage of topics No checking on understanding or seriousness of response Missing data Respondents may misunderstand instructions or items	Self-completion Impersonal Need for simple format Completion of sensitive information
Interviews face-to-face (individual)	Opportunity for gathering in-depth data Reduction of false responses Benefits of human-to-human contact and interpersonal behaviour High response rate Useful for exploring complex issues Opportunity to explain and clarify items and take questions from respondents Flexibility in item sequence Can build trust and rapport Ensure that only the respondent answers	Potential for perceived threat and bias in face-to-face meeting Costly: time for conducting interview, data entry and travel Not possible for large-scale survey Need to train interviewers Long data-collection period Access to sample Little time to think or reflect Flexibility can reduce standardization	Location of interviews Participation Personal Interviewer and interviewee characteristics Conduct of interview affects responses Small samples Standardization
Interviews (group)	Time-saving (compared to individual) Opportunity for gathering in-depth data Reduction of false responses Benefits of human-to-human contact and interpersonal behaviour High response rate Useful for exploring complex issues	Risk of 'group think' Potential for perceived threat and bias in face-to-face meeting Threat to confidentiality Not possible for large-scale survey Scheduling time and location for whole group to be present Costly: time for conducting interview, data entry Little time to think	Participation Personal Interviewer and interviewee characteristics Conduct of interview affects responses Small samples Standardization

Cohen, 1359 non, L., & Morrison, K. (2017). Research methods in education. Retrieved from http://ebookcentral.proquest.com Created from gasouthern on 2020-03-05 19:17:31.

#### SURVEYS, LONGITUDINAL, CROSS-SECTIONAL AND TREND STUDIES

	Advantages	Disadvantages	Either advantages or disadvantages
Telephone	Opportunity to explain and clarify items and take questions from respondents Flexibility in item sequence Can build trust and rapport Ensure that only the respondent answers Honesty Anonymity (absence of the human face) Reduction in costs: time, money and travel Rapid contact Random dialling Access to dispersed sample and distant locations Response rate higher than postal survey Short data-collection period Opportunity to explain and clarify items Opportunity to probe participants Reduced interviewer and interviewee bias	Lack of visuals and non-verbal cues: oral and aural medium only Finding telephone numbers (particularly with cellphones) Easy for respondents to refuse or quit through the survey (i.e. to hang up) Limited time (no more than ten minutes) Cold calling has a bad name Time of day for calling may be inappropriate Biased sampling (no telephone) Respondents are ex-directory Immediacy: no time to think of responses Cost (phone charges) Personal answering the call may not be suitable Multiple-choice, rating scale and ranking questions are difficult Order effects can be strong Risk of socially desirable	Sensitive questions: absence of an interviewer may encourage or discourage honesty of response Personal and yet impersonal Well-prepared and trained interviewer
Internet-based	Cost saving: time, money, data entry by researcher Speed: rapid distribution, completion and return Wide distribution: no problem of time and distance Access to minority and marginalized groups Opportunity for large samples and data volume Rapid data entry	acquiescence Security of data and confidentiality Biased sampling (no Internet, or respondents' limited Internet expertise, volunteer samples) No checking on understanding or seriousness of response Need for email addresses or posting opportunities Multiple submissions Risk of superficial coverage of topics	Honesty of responses Impersonal Anonymity, confidentiality and privacy

Copyright © 2017. Routledge. All rights reserved.

<b>TABLE 17.3</b>	CONTINUED		
	Advantages	Disadvantages	Either advantages or disadvantages
	Easy access to people and dispersed populations	Computer software compatibility and technical problems	
	Time to think	Limited number of items per screen	
	Opportunity to complete it in stages, i.e. with time breaks	Respondents give a minimal response	
	Complete at respondent's	Order effects	
	convenience	People quit if it is too long or	
	Opportunity for attractive survey	complex	
design and graphics Higher response rates than posta	Higher response rates than postal	Missing data (or resentment if forced responses are required)	
	surveys	Respondents regard it as spam	
	Environmentally friendly (no paper)	Design expertise of the researcher	
	Easy skip and branching arrangements	Respondents may misunderstand instructions or items	
	Honest responses to sensitive issues	Overall low response rates	
		Satisficing and acquiescence (see Chapter 18)	
	Standardized wording		
<b>D</b>			
Dropping off questionnaires	Opportunity to explain the survey face-to-face	Costly: distribution staff and time	Impersonal
	No training required for distribution staff (i.e. no interviews)		
	Respondents have time to think and reflect		
	Complete at respondent's convenience		
	Higher response rates than postal survey		

emails with interviews). Single mode surveys, write Dillman *et al.* (2014), are less effective than mixed mode surveys (e.g. telephone calls and emails, emails and websites, etc.) in terms of response rates.

Table 17.3 sets out advantages and disadvantages of these different types of survey administration.

We include Internet methods in Table 17.3, for ease of comparison with other methods, and our discussion turns to Internet surveys, devoting the next chapter entirely to this topic.

### Companion Website

The companion website to the book includes PowerPoint slides for this chapter, which list the structure of the chapter and then provide a summary of the key points in each of its sections. These resources can be found online at www.routledge.com/cw/cohen.