S ALEXANDER HASLAM AND CRAIG MCGARTY

RESEARCH METHODS AND STATISTICS IN PSYCHOLOGY

3RD EDITION
# Contents

Preface to the Second Edition xiii
Preface to the Third Edition xv
The Authors xvii
Endorsements for the book xix
Key Features of this Book xx

1 Introduction 1
   ‘Why do I have to do this?’ 1
   The structure of this book and an overview of the chapters 4
   How to use this book 8
   Discussion/essay questions 11

2 Research in Psychology: Objectives and Ideals 13
   What is psychological research and why do it? 13
   How does psychological research progress? 16
   Principles of good research 22
   Some notes of caution 29
   Further reading 32
   Objective and ideals: A checklist for research evaluation and improvement 33
   Discussion/essay questions 38
   Exercises 38

3 Research Methods 39
   Psychological measurement 39
   The experimental method 48
   The quasi-experimental method 56
   The survey method 58
   The case study method 61
   Overview 64
   Further reading 66
   Research methods: A checklist for research evaluation and improvement 66
   Discussion/essay questions 70
   Exercises 70

4 Experimental Design 71
   Choosing an independent variable 72
   Choosing a dependent variable 76
CONTENTS

Statistical inference: A checklist for research evaluation and improvement 205
Discussion/essay questions 208
Exercises 208

8 Examining Differences between Means: The $t$-test 211
Student’s $t$-distribution 212
Comparing the results for a single sample with a specific value 216
Within-subjects $t$-tests 222
Between-subjects $t$-tests 225
The controversy about what to do with $t$-values 232
Handling the results of $t$-tests: The hypothesis-testing approach 234
Other ways of handling the results of $t$-tests: Probability-level, confidence-interval and effect-size approaches 244
Some notes of caution 251
Overview 257
Further reading 258
$t$-tests: A checklist for research evaluation and improvement 258
Discussion/essay questions 262
Exercises 263

9 Examining Relationships between Variables: Correlation 265
Some basic principles of correlation 267
The measurement of correlation 270
Interpreting and making inferences about correlations 275
Some notes of caution 279
Conclusion 287
Further reading 288
Correlations: A checklist for research evaluation and improvement 289
Discussion/essay questions 291
Exercises 292

10 Comparing Two or More Means by Analysing Variances: ANOVA 295
Analysing variances 296
Comparing multiple means using one-way analysis of variance 299
Another way to explain analysis of variance: Sums of squares and mean squares 307
How big does a difference need to be? Significance testing and effect sizes 312
What does analysis of variance buy us? Some notes on comparing individual means 315
Using $F$-ratios with and without comparisons planned in advance 317
An introduction to analysis of variance with two independent variables 320
A final word 336
Further reading 337
ANOVA: A checklist for research evaluation and improvement 337
Discussion/essay questions 340
Exercises 341

11 Analysing Other Forms of Data: Chi-square and Distribution-free Tests 343
Dealing with a single categorical variable that has two levels:
  The binomial test 347
Dealing with a single categorical variable that has more than
two levels: The chi-square test of goodness of fit 351
Examining the relationship between two categorical variables:
  The chi-square test of independence for 2 × 2 tables 354
Distribution-free tests 361
Examining differences between two groups with relaxed assumptions:
  The Mann–Whitney test 363
Some options for other cases 370
Overview and checklist 372
Further reading 375
Other forms of data: A checklist for research evaluation and improvement 376
Discussion/essay questions 379
Exercises 379

12 Classical Qualitative Methods 381
A classical critique of quantitative approaches 383
Classical methods of collecting qualitative data 387
Some general principles for collecting and analysing qualitative data 389
Examples of classical qualitative research methods 391
Critiques of classical qualitative methods 413
Further reading 415
Classical qualitative methods: A checklist for research evaluation
  and improvement 416
Discussion/essay questions 418
Exercises 419

13 Contextual Qualitative Methods 421
A contextual critique of quantitative approaches 422
Examples of contextual qualitative research methods 427
Critiques of contextual qualitative methods 445
Conclusion: The challenge of contextual qualitative methods 447
Further reading 449
Contents

Contextual qualitative methods: A checklist for research evaluation and improvement 449
Discussion/essay questions 452
Exercises 452

14 Research Ethics 455
Science and society 455
Ethical values 459
Ethical risks associated with participation in research 461
Research with animals 470
Final comment 472
Further reading 474
Research ethics: A checklist for research evaluation and improvement 474
Discussion/essay questions 478

15 Conclusion: Managing Uncertainty in Psychological Research 479
‘Where has all this got us?’ 479
Managing uncertainty in psychological research 480
Final comment 488
Further reading 489
Discussion/essay questions 490

Appendix A: Writing Research Reports in Psychology 491
Overview 491
Basic structure 491
The sections in detail 493
General stylistic issues 500
Presentational issues 502
Numbers in the text 505
Referencing 505
Further reading 507
Sample psychology research report 507

Appendix B: Areas under the Standard Normal Curve 513

Appendix C: Answers to Exercises 517

References 539
Author Index 549
Subject Index 553

Xi
Supplementary Online Appendices
A: Annotated Psychology Research Report
B: Hand Calculation of Statistical Tests
C: Statistical Tables
D: Data Files
Many people are familiar with opinion polls or surveys, and just about everyone who reads this book will have participated in such a survey at some time or other. In an opinion poll, the opinions of members of a sample are measured in order to estimate the amount of support for particular opinions in the general community. For example, pollsters might randomly select a sample of 2000 voters from around a country and try to establish which political parties and policies those people support. In this case the researchers conduct the poll not because they are more interested in the views of this sample of people than in the views of the rest of the population, but precisely because they want to use these people’s responses to estimate what the opinions of the rest of the population are. In a political opinion poll, the aim is often to predict which candidate or party will win a forthcoming election.

The term ‘survey’ is therefore one that most people will have come across before. However, in this chapter we want to work with a precise definition that distinguishes surveys from experiments. The definition rests on the idea that in any piece of research you always have to measure variables but you do not necessarily have to manipulate them. If you measure dependent variables and manipulate independent variables you are doing an experiment (or perhaps
In this chapter we start by considering the differences between surveys and experiments in more detail. We then look at the sorts of questions that surveys are used to answer. After that we examine the process of finding a sample or selecting participants to help answer research questions. Finally, we discuss some of the different types of survey technique available.

The differences between surveys and experiments

If they were asked to make a general observation about the difference between surveys and experiments, a lot of people would say that surveys involve doing research in the community and that experiments involve doing research in the laboratory. This is often true, but it is not always the case. We have already noted that surveys can be done in places that have the words ‘Psychology Laboratory’ on the door, and that experiments can be done in the field.

The main differences between experiments and surveys relate to the sorts of question that surveys and experiments can answer. As we discussed in Chapters 3 and 4, experimental research tends to be concerned with establishing causal relationships between variables; a quasi-experiment if you have not used full experimental control); if you measure both then you are doing a survey.

In a fascinating exploration of the rise of the survey as a research tool in the twentieth century, Jean Converse (2009) notes that, over time, this became an indispensable tool of both power and democracy. On the one hand, various elites (e.g., political and scientific) found surveys essential for understanding the sentiments of citizens that they wanted to influence, predict and control. On the other hand, the capacity for surveys to gather information from large representative samples of citizens also allowed people to participate more fully in the democratic process. For this reason, the survey can be thought of as a tool that proved to be every bit as important for industrialization and social progress as other forms of technology and communication.

Experimenters typically do this by comparing different groups that have been equated by randomization. If the differences between groups are big enough, they conclude that the difference arises from the manipulation of the independent variable.

In contrast, surveys tend to be concerned with measuring naturally occurring relationships between variables. The results from the sample surveyed are generalized to the population; that is, they are used to estimate the characteristics of the population of interest. It is important to note that any statement about the population is always an estimate and therefore statements about observed relationships are always made with some uncertainty.

Let us compare two studies of the same question to see how the experimental and the survey methods would treat them differently. We will look at the effect of televised violence on aggressiveness in children – a question of interest to social, developmental and personality psychologists. Some researchers would address this question using the experimental method. They might take a sample of 100 children, randomly divide them into two groups (experimental and control) and then show the experimental group a television programme containing some violence. The control group could be shown an episode from the same series that did not include violence. After the children had watched the programme the experimenters could measure the amount of violence that the children displayed in a specified period (e.g., the next hour).

Taking a rather different tack, survey researchers might measure (rather than manipulate) the amount of violent television that children watch and then measure the aggressiveness of the children. This research procedure might involve selecting a random sample of 1000 children and then giving their parents a questionnaire that asks them to list the television programmes that their children normally watch. The questionnaire would also ask the parents to describe their children’s behaviour. The researchers would code the level of violence in the television programmes that each child watches and the parents’ descriptions of the aggressiveness of their children’s behaviour. There could be lots of variations on this research that would improve the methodology of the study, but this is a fair example of how this question could be addressed by survey techniques.

This example illustrates two key differences between experiments and surveys. First, the experiment allows researchers to manipulate variables and to use experimental control. Second, surveys tend to involve many more participants (often more than 10 times as many). Among other things, this means that survey research can be much more expensive to conduct.

The fact that surveys require more participants is hard to get around. A major reason for this is that survey and experimental research often have different objectives. One key difference is in what they aim to generalize. As we noted in Chapter 3, survey researchers often try to generalize from the sample directly to the population in order to make statements about the relationships observed in everyday life between psychological, behavioural and environmental characteristics. In order to do this they need to ensure that their sample is a representative random sample of the population.

Experimenters, on the other hand, try to establish controlled conditions where they show the immediate effect of some independent variable on a dependent variable. As we discussed in Chapter 4, they seek to establish controls so that they can generalize observed processes to
a population where the same conditions are found. In contrast, much survey research tries to generalize a particular result to the population.

Experiments are therefore well suited to investigate things that are believed to be changeable. In particular, they are used to examine the dynamic mental processes that we talked about in Chapter 3. Surveys are more suitable for studying things that are believed to be constant and enduring – for example, such things as intelligence or personality.

We can illustrate these points by referring back to the example of research into televised violence. Let us imagine that the experimenters found that children in the experimental group were more likely to go out and hit another child than were children in the control group. The researchers might seek to generalize this finding by saying that their research shows that watching televised violence makes children more likely to commit *aggressive acts*. Survey researchers, however, might find that children who watched violent television programmes were more likely than those children who did not watch violent programmes to be rated as aggressive by their parents. They might then seek to generalize this finding by saying that *aggressiveness* is associated with watching violence on TV.

Survey techniques thus lend themselves to examination of stable long-term states or conditions, while experiments are more suited to studying the immediate effects of psychological processes on behaviour. So, in the above example the survey and the experiment actually address related but subtly different questions. The experiment addresses the immediate effects of one television programme, and the survey addresses the long-term effect of televised violence on aggressiveness as a stable characteristic.

As we also noted in Chapter 3, the main feature that distinguishes surveys from experiments is that the experiment allows us to infer causal relationships. In our example, the most that the survey researchers can hope to demonstrate is that watching violent television programmes is or is not associated with aggressive behaviour. Survey research never allows us to say that violent television *causes* aggressive behaviour because there are always other uncontrolled variables that could have produced any observed effect. Perhaps aggressive children like to watch violent programmes, or perhaps boredom leads people both to be aggressive and to watch violent programmes. That is, if children who watch more violent programmes are more aggressive it need not be the case that watching violent programmes *makes them* aggressive. Again, the point here is that correlation is not causation. The fact that two things are associated with each other does not mean that one causes the other.

In (well-designed) experiments we can be much more confident about causal inferences. If experimenters find a difference in aggressiveness between their experimental and control conditions then this difference can only be due to one of two things: random error or the manipulation of the independent variable. This is because we have *controlled* everything else. In Chapters 7, 8 and 10 we will discuss how we make the decision as to whether chance or the manipulation is the most plausible cause.

Given this enormous advantage that experiments offer the psychological scientist, one might reasonably ask why researchers bother doing anything other than experiments. The answer is that sometimes scientists are only interested in observing relationships. This is especially true in the initial stages of research where it is often useful to observe what is going on before
launching into experimental manipulations – not least because we may have no detailed idea as to which variables to manipulate. Indeed, even if people do not conduct formal surveys before they do experimental research on an issue, they will often at least attempt to make some observations relevant to the process in which they are interested. Also, there are circumstances under which scientists are not allowed or not able to conduct experiments.

These issues apply to all sciences, not just psychology. Astronomers and geologists rarely do experiments, simply because it would be difficult or impossible to manipulate the independent variables of interest (such as the position of certain stars or the depth of the earth’s crust). Instead they rely largely on the same logic of controlled observation that underpins psychological surveys. Clearly, this does not mean that astronomy and geology are unscientific. Similarly, an epidemiologist (who studies the spread of a disease) would not be allowed to release a disease into a human community. All the epidemiologist can do is study diseases that have already occurred. Psychologists often confront the same problem – one of the main reasons why they are unable to manipulate some important variables is that such manipulation would be unethical (an issue we focus on in Chapter 14).

There are two complications in this overview of the differences between surveys and experiments. Some surveys actually involve manipulations (e.g., those that use split-ballot techniques in which different people complete different versions of the survey), and if these involve full experimental control then they are really experiments. Some surveys also incorporate manipulations without randomization and, as we discussed in Chapter 4, these can be thought of as quasi-experiments.

Test Yourself 5.1*

Surveys are preferred in some areas of psychology for which of the following reasons?

a. They are usually cheaper.
b. They allow researchers to infer causal relationships.
c. They are more scientific because similar methods are used in astronomy and geology.
d. It is often impossible to manipulate the independent variables the researchers are interested in.
e. They use randomization to achieve random sampling.

The correct answer is (d). Answer (a) is wrong because surveys are often more expensive than other forms of research; (b) is wrong because surveys do not involve experimental or causal methods; (c) is wrong because similar methods are used in both surveys and experiments; (e) is wrong because randomization and random sampling are quite different ideas. Random sampling relates to the procedure for selecting the sample, randomization is the process for assigning participants to different conditions.
Another question you might reasonably ask is how surveys can have independent variables when no variables are actually manipulated. The answer is that in surveys independent variables are variables the researcher believes to be causal. If researchers do not manipulate variables they cannot prove that they are causal variables. They can, however, prove they are not causal variables. Understandably, there is quite a lot of confusion on this point: correlation cannot prove causation, but a lack of a correlation can disprove causation. If our television survey found no relationship at all between the amount of violent television watched and aggressive behaviour (using valid and reliable measures), then aggressive behaviour cannot be caused by violent television. Thus, under certain conditions, survey methods can allow researchers to falsify causal hypotheses.

code A scheme for converting responses into numerical values. Researchers might write something such as ‘Sex was coded 1 = male, 2 = female’. All this means is that, when describing the results, numbers are used rather than words. This is a useful thing to do for the purpose of analysing data.

questionnaire A set of questions to be answered by a research participant. These questions could be printed and given to the participant, completed by participants online, or asked by the researcher in an interview.

representative random sample A random sample of the population that has the same characteristics as the population.

split-ballot technique A survey research procedure whereby experimental manipulations are included. The most obvious form involves randomly selected participants receiving different questionnaires. In effect, split-ballot techniques are not surveys but experiments.

Setting the question

Many surveys are designed to estimate the level of one or more variables in the population. This is true of political opinion polls, for example. Psychological surveys are more commonly concerned with estimating relationships between variables. Indeed, just about any issue in psychology can be addressed in a survey design if a researcher has enough time and resources. For example, they can be used to answer questions about the relationship between gender and verbal reasoning, or between personality disorders and homelessness, or between age and the expression of a particular emotion. In each case we want to know if one of the variables is related to the other – whether gender differences are linked to differences in verbal reasoning, whether personality disorders are associated with homelessness or whether children’s emotional sophistication changes as they grow older.

In all of these cases we cannot manipulate the relevant variables. We cannot make people belong to different sexes, we cannot give people personality disorders, we cannot make children become older. Nevertheless, we can often eliminate the possibility that some
relationships are causal if it is the case that one thing can only cause another if it occurs before the other. Children’s eye colour could be caused by their parents’ eye colour (in fact children do inherit eye colour from their biological parents), but parents’ eye colour cannot be caused by their children’s eye colour. Similarly, it is silly to think that improving children’s emotional expression will increase their physical age.

Researchers decide which relationships to investigate in their surveys by a complex process (along the lines outlined in Chapter 2). Generally, though, the questions they address are determined by issues that arise at a particular point in time within particular scientific communities and within society at large. In this way theoretical development and pressing social issues often come together to determine the precise questions that survey researchers investigate.

**Finding a sample**

Once we have decided what research questions we want to address, we need to find some participants to help us answer them. That is, as we noted in Chapter 4, we need to decide what population to sample from. Most survey researchers want to control sampling for two reasons. They want to reduce the amount of uncertainty about their sample so that it is as small as possible (see the discussion of sampling error at the end of Chapter 6), and they want to be able to generalize the results of their study to the relevant population.

A good example of this process is a political opinion poll, where a sample of people is asked which political candidate they are going to vote for. Here the goal of ethical researchers will not be to produce change in the way that it generally is in an experiment. So in this case, they do not want to change people’s opinions; rather they want to find out what those opinions really are. More specifically, they want to eliminate uncertainty and know both that the methods they use are reliable and that the opinions expressed are valid reflections of people’s real voting intentions (in fact, we can never be totally sure about this latter point). In this case, the poll will only be useful if it can be generalized to the entire population of voters (in other words, the poll needs to have external validity).

An important point to note here concerns the relationship between sample representativeness and sample size. Here one might imagine that any problems of representativeness can be compensated for simply by having a larger sample. Unfortunately, this is not the case. When researchers are interested in any large population, if the sample they study is not representative of that population it does not matter whether they have two dozen, two thousand or two million people in their survey. If the sample is not representative, the only thing a bigger sample does is lead researchers to draw the wrong conclusions more confidently.

One of the best demonstrations of this point was provided by the *Literary Digest* survey of voter preference for the 1936 US presidential election. This magazine conducted what was one of the largest ever political opinion polls. Its researchers contacted every US voter who was listed in the telephone directory or who was registered as owning a car. The researchers obtained over 2 million responses (a typical opinion poll of the US population today will involve between 1000 and 2000 people). Despite their prodigious effort, the pollsters got the
result completely wrong. The poll said that the Republican candidate, Landon, would win by a landslide when in fact President Roosevelt was re-elected. The apparent reason for this failure to predict the correct outcome was that the researchers had limited their sample to the wealthier segments of the population who owned cars and telephones. This population was not the same as the US voting population in 1936 (though it would be far more representative now). So although the researchers had a big sample, it was a very biased (i.e., unrepresentative) sample. Similar problems would have occurred in the most recent US presidential election if pollsters had limited their sample to male voters, white European American voters or voters who were members of a golf club.

It is therefore essential that a survey sample is representative of the population of interest. If it is not, then the sample is useless for any purpose. We noted in Chapter 4 that, for many psychological experiments, the population in which researchers are interested is all human beings who display the psychological process they are investigating. For research into visual perception, the population of interest is usually people with normal vision. So for research in this area just about anybody with normal vision who can communicate what he or she sees and understands the experimenter’s instructions can be part of the sample. No matter how narrowly people are chosen, they should be representative of people with normal vision. Similar sampling considerations are true for many other areas of psychology, and they mean that in these areas the sample can effectively be comprised of anyone who is available to participate. This sampling procedure is called convenience sampling.

Photos 5.1 and 5.2 In the 1936 US presidential election, pollsters predicted that the Republican candidate, Alfred Landon, would win. However, these polls were flawed because the sampling frame was biased, and Landon lost the election to the Democrat Franklin Roosevelt (pictured in the left picture here) by 8 electoral votes to 523. In contrast, in 2012, pollsters using far superior sampling techniques were able to predict the support for Democratic Party candidate Barack Obama. In 2016 pollsters were less successful in predicting the electoral outcome because they were less successful in predicting exactly where the support for the Republican Party candidate, Donald Trump, would be strongest.
The alternatives to convenience sampling are systematic sampling methods, which can be divided into probability sampling and non-probability sampling techniques. Probability sampling involves drawing people from the population so that any member of the population has a specifiable probability of being sampled. This sounds complex, but all it means is that when we select a probability sample we have to know what each population member’s chance of being included in the sample is. In simple random sampling (a special case of probability sampling) every individual has exactly the same probability of being sampled.

Simple random sampling involves obtaining a complete listing of the population of interest, whether that is a class list, the voter roll, a telephone directory or something else. This listing is called a sampling frame. For the Literary Digest survey, the telephone directory and the list of car owners constituted the sampling frame (the problem was that the sampling frame did not list the entire population of interest). To draw a simple random sample from the population, the researchers would then attach a number to each person in the listing and if that number were chosen by a random process (e.g., pulling numbers out of a bucket, or using a table of random numbers) that person would be included in the sample. Simple random sampling is one method that makes it possible to obtain a representative random sample of the population.

Non-probability sampling includes all techniques where there is not an identifiable probability of each member of the population being included in the population. Convenience sampling is one form of non-probability sampling; another is purposive sampling. Purposive sampling involves obtaining a sample of people who all have a particular characteristic. For example, researchers who are interested in the effects of televised violence on children might take a random sample of the whole population and then exclude households where there were no children. But it might be more sensible just to go out and try to find a group of children. These children would suit the researchers’ purpose but would not have been drawn randomly from the population. The same point applies if researchers are interested in studying mental illnesses and personality disorders. If they want to find out what the level of the mental illness schizophrenia is in the community, it makes sense to sample the entire population. However, if they want to find the relationship between schizophrenia and other mental illnesses (such as depression), it makes sense to restrict their sample to schizophrenics.

The general rule is that non-probability samples are not useful for estimating the level of a phenomenon in a population. If we want to know whether young people who watch more television eat more junk food than those who watch less television, then we cannot find out how much junk food people eat or how much television they watch simply by going to the local junk food outlet and interviewing young people (a purposive sample). Their answers may be honest but they may still mislead us in the same way as the Literary Digest survey results misled people in 1936. Perhaps the people who watch a lot of TV are having their food home delivered. Despite this limitation, non-probability samples can be useful in mapping out the range of possibilities. A survey in the junk food outlet may give us clues about the relationship between variables. For example, if the hypothesis that people who watch more television eat more junk food is true, then it would be remarkable if the people in our local junk food outlet reported never watching any television at all unless there is something odd
about that outlet or this is a freak observation. A single non-probability sample on its own cannot resolve this matter but it is a piece of evidence that can add to one’s understanding of a given phenomenon in a useful and meaningful way.

Once researchers have decided on a sampling procedure they need to make a decision about the sample size. If (but, as we have seen, only if) the researchers eliminate systematic bias from the sample, then the bigger the sample, the better it will reflect the population and so the better it will be for the research. This is because, for reasons we will discuss in Chapter 7, providing it is unbiased, a larger sample size reduces uncertainty about what conclusions it is appropriate to draw on the basis of data obtained from the sample.

The size of the chosen sample will be reduced by non-response. This occurs, for example, where people do not feel like participating in or forget to participate in the research, or their responses are lost in the mail. This is a problem because a smaller sample size reduces the ability of researchers to draw conclusions from the research. More importantly, though, it can also introduce threats to both internal and external validity. For example, if people choose not to respond because they are offended by the survey instrument, then the sample will be unrepresentative of the population. For example, if a questionnaire sent to married women in North America is called ‘A Study of Housewives’ Attitudes’, then it might be the case that women who object to the term ‘housewife’ will refuse to fill in the questionnaire. This problem is one of reactivity – a threat to both internal and external validity that we discussed in the

**RESEARCH BITE 5.2**

*When do people respond to surveys?*

When people fail to respond to a survey this can introduce sampling bias because non-responders may be unrepresentative of the population as a whole. Taking steps to increase survey response rates is therefore very important. Julia Cook, Heather Dickinson, and Martin Eccles (2009) examined 350 studies to identify factors that increase the likelihood of medical staff responding to postal surveys. Among other things, they found that people were more likely to respond to relatively small surveys than to big ones (those with more than 1000 people) and if they were sent a reminder. These patterns fit with other evidence that people are more likely to respond to surveys when they feel that researchers are interested in them personally and that their responses really matter.

previous chapter. Similarly, it is conceivable that only highly motivated people will respond to the survey and not those who are too busy or too lazy (this is a problem of mortality). In both cases, the survey might therefore systematically exclude a relevant part of the population (e.g., feminists or the unmotivated) and, as in the case of the 1936 Literary Digest survey, this would lead to biased results that could easily lead researchers to draw misleading conclusions.

Test Yourself 5.2***

A television news programme shows a murder case including video footage of the grieving parents of the victim. The television station then conducts an opinion poll in which it asks viewers to phone in and vote for or against the death penalty for murder. The results of the survey show that 83% of the 20,000 viewers who ring in are in favour of the death penalty. Which of the following statements is true?

a. The results of this study can only be generalized to people who watch news programmes.
b. The results of this study can only be generalized to people who own a television.
c. The results of this study can only be generalized to people who care about the death penalty.
d. The results of this study can only be generalized when we know much more about the sampling method.
e. The results of this study can only be generalized if the bias created by showing the grieving parents is eliminated.

The correct answer is (d). Unless we know more about the way the sample is selected we cannot draw any conclusions about the true level of support for the death penalty. In this case the sample consists of the people who watch the programme, only those who are watching the programme and are able to access the phone-in system. We can reasonably assume that those who watch the programme are likely to be the people who are interested in the issue, but we do not know how many people watch the programme. We cannot conclude that the results are representative of the community as a whole. In short, we cannot draw any firm conclusions about the population based on the sample.

convenience sampling A sampling procedure where a sample is chosen from the people who are available to participate in research. Often just about any person who can understand the instructions and complete the task is acceptable. This is the technique used in many psychological experiments (in particular, those involving undergraduate students).
non-probability sampling A sampling technique where there is not a specifiable probability of a member of the population being sampled.

non-response The failure to obtain responses in circumstances where a person is selected from the population but does not actually participate in the survey.

probability sampling A sampling technique where there is a specifiable probability of each member of the population being sampled.

purposive sampling Selecting those members of a population who have a definable characteristic. For example, a study of depression might sample only those members of a population who are clinically depressed.

sample size The number of participants in a study. This is represented by the symbol \( N \).

sampling frame A listing of all members of the population of interest.

simple random sample A random sample in which every member of the population has the same probability of being included.

survey instrument A means of collecting data from the sample. This could be a questionnaire form, a web page or an interview.

systematic sampling Sampling where participants are selected according to a specific plan or method (either probability or non-probability sampling).

Types of survey

So far we have talked about survey design in very general terms. In fact there are many different survey methods, and the wide choice reflects the many different interests, approaches and objectives of psychologists. We cannot, for example, expect that the same techniques would be applicable to studying adults and children, or for answering questions in cognitive and clinical psychology. It is difficult to say that one method is always better than another, although some techniques are certainly better than others for particular purposes.

In the past, questionnaires tended to take the form of printed booklets with instructions and questions. These could be administered by researchers (i.e., handed out and then collected), mailed out to potential participants or distributed by other means such as email or in a magazine. Today, it is probably more common for questionnaires to be completed online, either by asking participants to access a link to the survey or by asking them to complete the survey on a computer or a tablet. Important advantages of this method are: (a) that it saves time and money (because questionnaires do not need to be printed and data do not need to be entered by the researcher), (b) that it reduces error (because data entry is done automatically rather than by hand), (c) that it increases the likelihood of obtaining a complete set of responses (because the survey can alert participants to questions they have missed), and (d) that it makes data easier to store. At the same time, though, data that are electronically stored require
Aside from questionnaires, survey data can also be obtained using a range of other procedures. Of these, interviews and observational studies are the most common, and we can consider each in turn.

**Interviews** are really just questionnaires where relevant questions are asked by researchers either in person or remotely (e.g., over the phone or via email). The interviewer writes down or records the response as he or she asks each question. **Computer-aided interviewing** is one particular form of interview that is used for telephone surveys. Here the computer assists interviewers by telling them which question to ask next. Computer-aided interviewing also allows researchers to conduct split-ballot surveys. That is, they can help to turn surveys into experiments by introducing manipulations. The computer achieves this because it is programmed either to produce different questions depending on previous responses or to ask different people different questions (randomly assigning them to different conditions).

**Observational studies** are widely used in both human and animal research and involve the observation of behaviour. **Naturalistic observational studies** do this without any attempt to interfere with that behaviour. As an example, comparative psychologists might study the behaviour of animals by observing them in the wild, or environmental psychologists might care.
study the movement of people through a shopping centre or the flow of traffic through an intersection. In all of these cases there is no attempt to change behaviour through intervention (e.g., in the form of an experimental manipulation).

Observational studies can take many forms. One possibility is direct public observation, where the participants are fully aware that they are being observed. This is normally true in case studies (as discussed in Chapter 3). For example, organizational psychologists might want to know how corporate boardrooms really work. Suppose they want to examine the hypothesis that male board members have more power than female members. The psychologists might obtain permission to attend a company’s boardroom meetings for a few months in order to get a picture of what is going on. While attending meetings they would attempt to measure the amount of control over decisions that was exerted by males and females. To do this they might record the amount of time male and female board members spent talking during meetings, the number of times their suggestions were agreed to by other members, or the number of times males and females were interrupted. However, if the board members know they are participating in research, their behaviour might clearly change. For example, men may want to appear more reasonable and so restrain from engaging in dominant behaviour. Here, then, observation raises the possibility of reactivity that we discussed in the previous chapter. This is one reason why researchers will often try to conduct non-obtrusive observational studies, in which people are not aware their behaviour is being observed.

Obviously, there are other possible research strategies that could be used in this case. For example, the researchers could give a questionnaire to the board members. The problem is that what people write down on their questionnaires would not necessarily correspond to their behaviour in their everyday lives. If the researchers ask male board members whether they give fair consideration to the proposals of female board members, they may well say that they do. They may say this because it is true, or they may say it because it is socially desirable (i.e., they think it is the right thing to say), or because they do not remember how they actually behave (can you remember which shoe you put on first every day?). The point here is that people do not have perfect memories and they are not always totally honest.

This problem is illustrated by attempts to establish the popularity of television programmes using questionnaires (ratings booklets). These usually obtain different results from those obtained by electronic devices attached to television sets. One likely reason for this discrepancy is that when people fill in the questionnaires at the end of the week they forget, or lie about, what they have actually watched. As a result, questionnaires have come to be seen as a rather unreliable measure of viewer preference. Questionnaires may also be invalid measures if responses are biased. This could occur if people reported that they had watched particularly desirable or popular programmes. For example, it is easy to imagine that teenagers might not wish to report everything they had seen if they thought their parents would see the ratings book.

So, as in much psychological research, we are caught between two problems: that if we ask people to tell us what they think or how they behave they may not be accurate (because they cannot tell us the truth or they do not want to tell us); and that if we simply observe behaviour we may also make mistakes (possibly due to experimenter bias or reactivity).
These problems to do with the social desirability of responses apply to many types of research, including experiments, questionnaires and observational studies. Again, the problem here reflects the more general issue of reactivity.

As we have suggested, one way around this problem is to use non-obtrusive research procedures that ensure that participants are not aware that they are being studied (these methods are also called non-reactive techniques). The logic of such research is that people cannot change their behaviour in response to being observed if they do not actually know they are being observed. In observational studies this can be done by making audio recordings, by using secret video cameras or one-way mirrors, or by infiltrating organizations with researchers who pretend not to be researchers. Again, there are serious ethical issues associated with conducting this sort of research because the participants have not given their consent to be observed. Such hidden observation is always considered to be acceptable by scientists where the participants are animals (i.e., it is always acceptable to observe animals in their natural habitat, providing that that habitat is not disturbed by the research). Hidden observation is also acceptable where human participants are unable to give consent but someone else is legally able to (such as the parent or guardian of young children). However, a hidden camera in a boardroom or living room is extremely unlikely to be considered acceptable (however interesting the results might be).

Another alternative open to researchers who want to conduct non-reactive research is to use archival records or behavioural trace measures (see Chapter 3). Archival studies involve checking through records of behaviour such as books, videos, Internet sites, and so on. These are non-reactive because the behaviour has occurred in the past so the participants cannot change it as a reaction to being observed. Rather than watching what goes on in a corporate boardroom, the researchers may obtain evidence of what has happened in boardrooms in the past (e.g., by examining minutes and other records of meetings, the autobiographies of chief executives, and so on). For example, the researchers might operationalize dominance as the number of times proposals brought by male and female board members were accepted. This technique is exactly the same as the methods that historians often use in their research.

Research examining physical traces of behaviour attempts to use evidence of a previous behaviour to define some dependent variable. These techniques are identical to the techniques used by archaeologists. So, in their study of boardrooms, researchers might look at which seats the female board members were assigned, perhaps assuming that more dominant board members would occupy seats closer to the chairperson of the meeting (naturally, they would need to establish that this was a valid and reliable measure of dominance). If females were allocated less dominant seats, then this evidence would be consistent with the researchers’ hypothesis (although it would not establish a causal relationship between gender and dominance).

As we noted earlier, a lot of survey research is concerned with variables that have long-term effects. Often it is not possible or appropriate for experimenters to control people’s lives for extended periods of time, so it can be difficult to investigate variables that have long-term effects using experimental techniques. Thinking back to the example of televised violence, it is obviously the case that very few parents would allow experimenters to control the television
watching of their children over a period of weeks or months. Surveys are therefore a more effective way of exploring these issues.

Research that examines the long-term effects of ageing and maturation uses developmental surveys. A common type of developmental survey is the cross-sectional study. In this type of study, a cross-section or slice of the population is taken at a particular point in time. If psychologists want to know whether the number of years of education has any effect on IQ scores, then they need to study people who differ in the years of education they have received. Obviously they cannot manipulate the length of education. One possibility would be to take a sample of people with (say) 10 years of education and compare them with a sample with 12 years of education. A problem would arise, however, in the internal validity of this study, for if a difference in IQ were found, we could not be sure that the difference was due to the number of years of education rather than to some other difference between the samples. Any such differences are confounds, and this example can again be used to make the point that correlation does not establish causation. Perhaps people with higher IQ scores tend to stay at school longer (i.e., IQ could cause years of education and not the other way around).

The main alternative to the cross-sectional design is the longitudinal study. A longitudinal study involves taking a sample of participants and following them over a period of time. As an example, a longitudinal study into the effects of education on IQ might measure the IQs of a sample of students with 10 years of education and then measure them again two years later. Using the terminology that was introduced in Chapter 4, the difference between cross-sectional and longitudinal designs is that cross-sectional studies are conducted between subjects and longitudinal studies are conducted within subjects. The advantage of the longitudinal study is that it allows the researchers to use each participant as his or her own control, and this serves to reduce the uncertainty involved in deciding whether key variables are really having an effect (see Chapter 7).

Longitudinal studies of this type, however, can fall victim to a number of additional threats to internal validity in the form of testing, mortality and maturation effects (see Chapter 4). In our example, the effects of all three problems can be seen quite clearly. First, it is apparent that repeated testing may produce changes in the participants. People may get better at doing IQ tests through practice. Alternatively, they may start to do worse on the tests because they become bored or lose interest. Mortality effects might be a problem because, due to the passage of time between measurements, it is quite likely that a number of people would drop out of the study (perhaps those who were less intelligent or less motivated). Finally, in this example, education is actually confounded with maturation. Accordingly, IQ could increase with age rather than with education.

Most of these problems can be dealt with through appropriate controls of the form discussed in the previous chapter. However, in relation to the final point, it is worth noting that in developmental research maturation is often the independent variable of interest. So rather than wanting to control for maturation, developmentalist psychologists actually want to examine it in its own right. This points to something that happens all the time in psychological research: what one researcher sees as a crucial variable another sees as an undesirable confound.

120
Yet another alternative to these cross-sectional and longitudinal strategies is to conduct a successive cross-sectional study. That is, rather than following the same sample (called a cohort) over time, the researchers would draw a different sample from the population at different times. This is the procedure normally used in opinion polling. It overcomes the problem of testing effects, but because it has a between-subjects design, more participants are needed to draw conclusions with the same certainty. The differences between longitudinal, cross-sectional and successive cross-sectional studies are summarized in Table 5.1. The longitudinal design involves measuring the same sample twice, the cross-sectional study involves measuring two different samples at the same time, and the successive cross-sectional study involves selecting two different samples at two different times. Of course, these principles can be applied to studies that involve more than two observations.

Table 5.1  Types of developmental survey

<table>
<thead>
<tr>
<th></th>
<th>Time 1</th>
<th>Time 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Longitudinal</td>
<td>Select and measure sample</td>
<td>Measure sample again</td>
</tr>
<tr>
<td>Cross-sectional</td>
<td>Select and measure Sample A</td>
<td>Select and measure Sample B</td>
</tr>
<tr>
<td></td>
<td>Select and measure Sample A</td>
<td></td>
</tr>
<tr>
<td>Successive cross-sectional</td>
<td>Select and measure Sample A</td>
<td>Select and measure Sample B</td>
</tr>
</tbody>
</table>

Test Yourself 5.3**

Which of the following can be a threat to the internal validity of longitudinal studies?

a. Testing effects.
b. The IQ of the participants.
c. Sample size.
d. Maturation effects.
e. Both (a) and (d).

Wrong because sample size is unrelated to internal validity.
Wrong because students can separate the general concept from the details of the specific example.
Wrong (c) is also wrong because we used an example to explain longitudinal designs—people writing exam questions often use this ploy to see if students can distinguish between different designs. This is also true for longitudinal designs where we included the example to help students distinguish between different designs. The correct answer is (e), as (a) suggests testing effects are a problem in longitudinal studies because the issues are always present, and (d) suggests maturation effects are a problem in longitudinal studies because the issues are always present.
Although this chapter (like those before it) focuses on procedures that all involve attempts to quantify mental states and behaviour, before we move on, we should note that there are also observational studies that do not attempt to measure variables in the way that is implied throughout most of this book. This research uses techniques called qualitative research methods, and we will discuss these in a lot more detail in Chapters 12 and 13. These methods are often similar in appearance to the observational techniques used in certain types of case study, and they are becoming increasingly popular in some areas of psychology. The techniques (especially those that we will discuss in Chapter 13) typically involve reporting behaviour and commenting on it, without attempting to quantify the behaviour – that is, without attempting to express it in numbers. These techniques are often used by critical psychologists who challenge conventional approaches to psychological research. In turn, many mainstream psychologists are critical of qualitative methods. Again, we will explore the basis and nature of this debate in later chapters.

archival records Written or other records that are stored and are available for study. Sometimes these are stored in institutions called archives, but any record that is not in current use is also called an archive.

cohort A group of participants who take part in a particular piece of research at the same time.

computer-aided interviewing A research procedure where the researcher is prompted by a computer to ask questions over the telephone in a particular order. The order of questions is set by the chief researchers. Many surveys are now administered entirely by computer using dedicated web pages.

critical psychologists Psychologists who are critical of the conventional methods and approaches of mainstream psychology. For this reason, they tend to use alternative methods.

cross-sectional studies Studies that sample a cross-section of the population at a particular time. In developmental psychology the sample normally comprises a number of different age groups (cohorts).

developmental surveys Surveys conducted into the effects of ageing and maturation. Such surveys may be conducted by developmental psychologists or others interested in changes in psychological processes and behaviour over time.

interview A series of questions asked by a researcher and answered by a research participant.

longitudinal studies Studies where the same sample of participants is measured on more than one occasion.

naturalistic observational studies Non-experimental studies in which the researcher observes behaviour and makes no attempt to interfere with the participants.
5. SURVEY DESIGN

**non-obtrusive observational studies.** *Observational studies* in which data are obtained without the participants being aware that they are participating in research.

**non-reactive techniques** Research procedures that use *non-obtrusive measures* to reduce the reactivity of participants’ behaviour.

**observational studies** Studies in which data are collected by directly observing behaviour.

**qualitative research methods** Procedures for studying psychological and behavioural phenomena that do not involve their quantification.

**social desirability** The extent to which people’s behaviour appears acceptable to other people. If behaviour is affected by people trying to behave in ways that they perceive to be desirable to the researcher, then this threatens both the internal and external validity of research.

**successive cross-sectional studies** Studies that take repeated samples from a cross-section of the population at different times. These combine the features of *cross-sectional* and *longitudinal studies*.

### Constructing a questionnaire

We noted earlier that questionnaires are one of the most common forms of survey instrument. It is worth noting, too, that the use of questionnaires is not restricted to survey research. For example, in many of the experimental designs we have discussed, the effect of an independent variable would be assessed by using a questionnaire to collect data on relevant dependent variables. In light of the questionnaire’s importance as a tool in psychological research, it is therefore worthwhile outlining some general principles of questionnaire design.

Perhaps the most important feature of a questionnaire is that it needs to be user-friendly and welcoming. Remember that the respondents are doing you a favour by responding, so you want to do everything you can to make their task as easy and pleasant as possible. You should start by providing a clear introduction to the questionnaire that explains what it is about and why it is worthwhile, and provides key information that the participants need to know (e.g., who is going to use the information and what it will be used for) and which may reassure them (e.g., that their responses are anonymous and will be confidential).

The questionnaire as a whole then needs to be set out clearly and to ask questions in as straightforward
a way as possible. In particular, avoid using jargon and avoid asking two things in the same question (using what are known as double-barrelled questions). It should be easy to read, pleasing to the eye and create a positive impression (e.g., it should not contain typographic errors or bad grammar). It should not be any longer than it needs to be, because the longer it is, the more off-putting it is likely to be. Moreover, even if someone goes to the trouble of completing a long questionnaire, it is likely that by the time they get to the end, their concentration will have lapsed and hence the responses they give will be less valuable. This is especially true if you are asking for their opinion rather than for factual information. This is one reason why it is a good idea to ask for demographic information (e.g., about a person’s age, sex and level of education) at the end of a survey.

Nonetheless, sometimes a questionnaire needs to be long, and in such cases it is worthwhile paying careful attention to the way it is structured. In particular, the questionnaire as a whole should be organized so that the process of responding makes as much sense as possible. Sub-headings can help in this process, and, as far as possible, there should be a logical sequence to the questions.

The fact that people’s attention can wander while completing a questionnaire is also a factor to bear in mind when thinking about the order in which questions are asked. For this reason, it is often best to put the most important questions towards the start of the questionnaire, and to place open-ended questions towards the end (as these take longer to complete and are often more fatiguing). As noted above, this is also true in the case of questions that ask for demographic information. Another reason for placing these at the end of a questionnaire is that they can be quite reactive. For example, if you ask people whether they are heterosexual or homosexual, this may lead them to believe that the research is ‘about’ sexuality and hence affect their responses. Generally, then, reactive and sensitive items are placed at the end of a questionnaire.

It is also very common for researchers to ask respondents to provide answers to questions using response scales, where the nature of a response is indicated by selecting one of a number of options that are sequentially ordered on a scale. An example of this was provided in Chapter 4, where we discussed a manipulation check in which participants were asked to respond to the question ‘How much are you personally responsible for your failure?’ on a seven-point scale with end-points labelled ‘not at all’ (1) and ‘completely’ (7). The first thing to be sure of is that if you are using existing measures (that have been developed by previous researchers and have been reported in peer-reviewed publications) you should keep the scales and labels that the original researchers used. This allows you to compare your results to the previous research.

If you are creating new scales, it is important to attach meaningful labels to the numbers. This can be done either by providing anchors in which only the end-points are labelled (as in this example), or by attaching labels to all scale points, as in the following example:

```
Was this year’s exam more difficult than last year’s?

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>much less difficult</td>
<td>slightly less difficult</td>
<td>the same</td>
<td>slightly more difficult</td>
<td>much more difficult</td>
</tr>
</tbody>
</table>
```

Indeed, here the numbers need not be included at all, and can simply be added by the researchers when they set about performing quantitative analysis at a later stage.
Labelling all scale points reduces some of the ambiguity in a scale and makes it easier for participants to understand what a particular response means. However, this procedure takes up more space and can create problems if the labels do not correspond to the particular response that the participant has in mind. A general principle of scale labels is that they need to be non-overlapping and discrete, and this can sometimes be hard to achieve if all points are given separate labels.

It can also be a mistake to use a response scale with too many points. There is little to be gained by using more than nine, and five- or seven-point scales are the norm. Note that scales typically have an odd number of points because this means that they have a clear midpoint. On the other hand, under certain circumstances, using a scale with an even number of points can be useful, as the absence of a clear mid-point makes it impossible for respondents to ‘sit on the fence’. At their simplest, such scales take the form of two-option forced-choice responses (e.g., ‘yes–no’).

If a questionnaire contains a large number of such scales, this may also increase the likelihood of respondents simply making an identical response to each scale. To check for and to discourage response bias of this form, it can be useful to measure the same psychological state using multiple items where some items require reverse scoring. This involves asking questions so that a person with a given psychological state would answer some by marking a higher score and others by marking a lower score. For example, a questionnaire to measure depression might include the normally scored item ‘In the past week did you feel down-hearted and blue?’ as well as the reverse-scored item ‘In the past week did you feel you had a lot to look forward to?’. On the second item, a higher score would indicate lower depression, and so responses on it would need to be reversed when creating a single depression score. However, in constructing reverse-scored items, it is important to avoid the need for double-negative responses (e.g., the response ‘not at all’ to the statement ‘Are you never happy?’), as these are very confusing.

**double-barrelled questions** Questionnaire items that ask two (or more) questions at the same time (e.g., ‘Do you believe in climate change and does this upset you?’).

**forced-choice responses** Questionnaire items where a respondent has to select one response from two or more options.

**response scale** An ordered sequence of responses to a particular questionnaire item. These responses can be numbered in the questionnaire or numbers can be attached to them prior to quantitative analysis.

**reverse scoring** The practice of having some of the items that measure a particular construct worded so that a higher score is associated with a lower level of the construct. For example, on a scale designed to measure happiness, if a question asked ‘Are you sad?’, higher agreement would indicate less happiness. So, before calculating an overall score for the construct, scores on these particular items are transposed so that on all measures a higher score is associated with a higher level of the construct. This is done by (a) subtracting the participants’ responses from the scale mid-point and (b) adding the resulting score, including the + or – sign, to the scale mid-point to provide a new score.
Test Yourself 5.4**

Which of the following statements is true?

a. Questionnaires should not ask people to provide personal information.
b. Questionnaires should aim to obtain as much information from people as possible.
c. The order of items in a questionnaire is not particularly important.
d. Questionnaires should contain a mixture of open-ended and forced-choice items.
e. None of the above.

The correct answer is (e). All of these statements are false. Answer (a) is incorrect because, in some sense, all questionnaires require people to provide personal information. Note, however, that steps need to be taken to deal with sensitive personal information appropriately (e.g., by assuring respondents that their responses are anonymous and confidential). If this is true, the broad ethical implications of a particular questionnaire need to be addressed before it is administered (see Chapter 14). Answer (b) is incorrect because the order of questionnaire items is very important. For example, more reactive items need to be asked last and (d) is incorrect because the order of questionnaire items is very important. For example, more forced-choice items are required before the order of questionnaire items is very important (e.g., in reducing response fatigue). Answer (c) is incorrect because the order of questionnaire items is very important. For example, more forced-choice items are required before the order of questionnaire items is very important (e.g., in reducing response fatigue).

Conducting research online

If you do decide to conduct a survey early in your research career, then unless you are part of a large research team or you have access to relatively large amounts of research funding, there are two main methodological possibilities that you are likely to choose. Both strategies involve convenience sampling. If your target population is people in school classrooms, or you have access to people who belong to a particular organization or who work or live at a particular location, then you will probably be using a pen-and-paper survey. This will involve asking people whether they wish to participate in the survey and, if they do, handing them a questionnaire and waiting for them to return it to you.

In many other circumstances the best available option will be to conduct a non-probability survey via the Internet. The steps to follow here involve creating a questionnaire in very much the same way that you would for a pen-and-paper survey. When you have the list of questions you need a platform on which to administer the survey and a way of recruiting people to complete it.

The platform to administer the survey might belong to a university or private company (such as Qualtrics), but in either case it will have procedures (computer programs) for turning
127

5. SURVEY DESIGN

After the survey has been created, you also need to bring your survey to the attention of the people you wish to recruit. This may involve using a commercial service (e.g., a crowdsourcing platform such as Amazon’s Mechanical Turk, Crowdflower or Prolific Academic) to advertise the survey. Alternatively, it may involve making people aware of the survey via email, social media or through an organization. One critical issue here is that you need to avoid spamming. It is highly likely that you are already being sent lots of junk email that is of no interest to you. Some people will regard your request for them to participate in the same way. If you do use an email list or organization to send your request, then you need to make sure that you have permission to use it in this way. This relates to your ethical duty as a researcher (see Chapter 14) and it is another important professional responsibility.

The other issue that you will face is that, unless you use a reliable commercial or university service, you may be unable to confirm that your participants are who they say they are. For example, if you are looking for a sample of women who have experienced domestic violence and you receive 100 completed surveys, how can you be sure that the participants are women, or that they have experienced domestic violence, or that the same person has not completed the survey 100 times? The same issue applies with other methods, but it is more likely to be perceived to be a problem in the case of Internet-based research.

Photo 5.4 Online surveys are increasingly common in psychology. They have at least four important advantages over pen-and-paper surveys. First, they are easy and cheap to distribute. Second, the data are entered directly into a computer program ready for analysis, thereby saving researchers the trouble of having to do this and eliminating errors in this process. Third, they can be constructed so that respondents have to answer every question. And fourth, they save paper and storage space.

Your questionnaire into web pages and for recording and storing the responses. Many computer programs for doing this are relatively easy to use.
However, Birnbaum (2004) argues that this problem is likely to be more apparent than real – not least because people generally have no interest in lying or in wasting their time. There are also a number of ways in which researchers can reduce the risk of this problem. These include having procedures to ensure that no more than one set of responses can be submitted from the same computer, and being able to corroborate some findings with data drawn from other sources (e.g., another survey; see Research Bite 5.4).

Overview: Designing a survey

One thing that is clear from this chapter is that there are many options open to researchers who wish to use survey methods. Despite this, we should add that our coverage of survey strategies has been far from exhaustive and we have only hinted at the richness of possible survey designs. However, most of the strategies that we have not dealt with build upon the core principles and practices outlined above. In order to summarize these, it is instructive to highlight the main steps involved in designing a survey:

1. *Formulate the research question.* Before researchers can begin to collect data they must be clear about the question they want answered. In survey research the question is often about the relationship between two (or more) variables. Survey methods cannot be used to tell us whether one variable causes another. However, providing
they contain valid measures that yield valid data, they can be used to falsify causal relationships by showing that certain variables are not related. This is because one variable cannot cause another if it is not related to it.

2. **Decide on the measurement of dependent and independent variables.** Measuring variables often involves complex steps of operationalization (see Chapter 4). The variables should be both relevant and sensitive and every effort should be made to eliminate extraneous variables (see Chapter 3).

3. **Decide on the population of interest.** This involves deciding exactly to which population the results will be generalized.

4. **Decide on the sample size.** Other things being equal, the larger the sample size, the less uncertainty there will be about what is going on in the population from which the sample is drawn. This is because large surveys have a greater probability of revealing relationships between variables that exist in the population. However, increasing the sample size does not increase the validity of a survey if there is any bias in the methods or the sampling. In this case, a large sample will only increase the chance of drawing the wrong conclusions.

5. **Decide on the sampling method.** The key choice of sampling method is between a probability and a non-probability sample. In order to obtain a representative random sample of the population, appropriate probability sampling methods have to be used.

6. **Decide on the technique.** There are many ways of administering a survey. These include observational, questionnaire, interview, archival and other techniques. Choices about which to use will be dictated by a number of factors, including the ease and likelihood of obtaining appropriate responses and established practice in the research area.

7. **Consider possible experimental manipulations.** Researchers should always consider the possibility of using some experimental manipulations in surveys (the split-ballot technique). Experimental manipulations allow researchers to make causal inferences, and help them to be more certain about the reasons for any effects they obtain. These are massive benefits that are sometimes overlooked by survey researchers. One possible reason for this is that when some researchers conduct surveys, they believe they are collecting pure measures of psychological variables and thus do not feel there is a need to complicate the research with experimental manipulations. Some researchers also believe that such manipulations are artificial. This cuts to the heart of the distinction between surveys and experiments in psychology. Of course, many interesting variables cannot be manipulated experimentally (for ethical or other practical reasons), but that is not a reason to rule out experimental manipulations if they are possible. Even if it is not possible to manipulate the main independent variable of interest, it may still be possible to control experimentally for some confounds (e.g., order effects).

In conclusion, we should note that none of the above decisions can be made in isolation, and that it is not always possible (or necessary) to make decisions in the order they are presented here. Researchers may decide that they will measure the dependent variable using a particular psychological test and then find that the test is too long to include in a telephone interview and
that they have to change the measure. For this reason, we often liken designing any piece of research (whether it be an experiment or a survey) to a juggling act. As you get further into the act you pick up new balls. Each step in the design of the research is like a new ball that has to be worked seamlessly into the act, lest everything comes crashing down around your ears. Because of this, when dealing with a new problem, even the most experienced researchers often need to conduct preliminary **pilot studies** before they get their research procedures to work well.

**pilot study** A preliminary piece of research designed to ‘road-test’ various design elements (e.g., independent variables, dependent variables, details of procedure), in order to establish their viability and utility prior to the investment of time and money in a full study.

---

**CHECKLIST**

Revisiting the key goals for this chapter

- I understand the differences between surveys and experiments, and the implications of these for the research process.
- I know what the key components of a survey are, and what decisions I need to make when putting them together to design a good survey.
- I understand the different types of survey and the issues that surround their use.

---

**Further reading**

There are a number of very good volumes dedicated to issues of survey design. However, the book chapter by Reis and Gosling (2014) provides an accessible, comprehensive and relatively succinct treatment of many of the methods covered here. Fraser and Lawley’s (2000) text is also very user-friendly and provides multiple worked examples of appropriate survey practice. Birnbaum (2004) provides a useful treatment of web-based research – although we do not endorse his arguments about institutional review boards.


The original publication of the *Literary Digest* article that we discussed above is also available at: http://historymatters.gmu.edu/d/5168/
### Survey design: A checklist for research evaluation and improvement

Table 5.2

<table>
<thead>
<tr>
<th>Potential problem</th>
<th>Question to ask</th>
<th>Potential improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unrepresentativeness; sampling bias</td>
<td>Have the researchers collected data from a sample that is representative of the population of interest?</td>
<td>Think about the characteristics of people included in the survey and of those not included. If they differ, this suggests the sampling process was not representative (and was not random). Reflect on the impact that this may be having on the research findings. If this impact is significant, take steps to improve the sampling process (e.g., by making it truly random).</td>
</tr>
<tr>
<td>Limited consideration of options</td>
<td>Could the research have been improved by conducting a split-ballot survey?</td>
<td>Think of ways in which it may have been possible to investigate the impact of a relevant variable by systematically varying the form and content of the survey instrument.</td>
</tr>
<tr>
<td>Inappropriate sampling</td>
<td>Is the researchers’ sampling strategy appropriate?</td>
<td>Think about the range of sampling options available to the researchers in addition to the one that they have chosen (e.g., convenience sampling, purposive sampling). If any of these options would have had appreciable benefits (e.g., in saving time without compromising external validity) consider implementing it.</td>
</tr>
<tr>
<td>Inappropriate sample size</td>
<td>Is the sample size large enough?</td>
<td>If a sample is too small to reduce uncertainty about a particular finding to a tolerable level, then (providing that the sampling strategy is appropriate) increase the sample size.</td>
</tr>
<tr>
<td>Reactivity</td>
<td>Has the nature of the survey instrument affected who responds or the way they respond?</td>
<td>Think about ways in which the nature or content of the survey may be encouraging participants either to respond in particular ways or to avoid responding. Reflect on the impact that this may be having on the research findings. Consider trying to assess the extent of this impact by conducting a split-ballot survey. If the impact is significant, modify the survey instrument appropriately (e.g., by rewording the rubric or particular questions).</td>
</tr>
<tr>
<td>Inappropriate survey choice</td>
<td>Have the researchers used the appropriate type of survey?</td>
<td>Think about the survey techniques that the researchers might have used but did not (e.g., archival, observational, non-reactive). If any of these have obvious advantages, design a study that incorporates them.</td>
</tr>
</tbody>
</table>

(Continued)
Question to ask

Is the questionnaire polished, clear, and in a form suited to obtaining information that will allow researchers to address the research question(s) in which they are interested?

Potential improvement

Think about how the questionnaire will appear to potential respondents. Why might they not want to complete it? What might they misunderstand? Also, think about the nature of the information the questionnaire will yield. Will it allow you to answer your research question? Make changes in light of your reflections on these questions.

Test Yourself 5.5**

In a survey where the results are obtained from a representative random sample of a population, which of the following is true?

a. The results can be generalized to that population.
b. The sample can only be obtained by simple random sampling.
c. The sampling procedure is similar to that used in most experimental research.
d. Both (a) and (b).
e. None of the above.

Test Yourself 5.6**

Which of the following statements about convenience sampling is true?

a. It should always be avoided.
b. It is appropriate providing the sample size is extremely large.
c. It can be used under some circumstances.
d. It is a non-probability sampling technique.
e. Both (c) and (d).

Test Yourself 5.7***

Which of the following is likely to be a problem for non-reactive studies that use non-obtrusive measures?
5. SURVEY DESIGN

The correct answer to 5.5 is (a). Answer (b) is not correct because there are other probability sampling techniques that will produce representative random samples, and (c) is wrong because representative random samples are not always easy to obtain. The correct answer to 5.6 is (e). If (a) were correct only a very small proportion of published research in psychology (and almost no experiments) would ever have been conducted. Answer (b) is not true, as the 1936 Literary Digest survey shows, and (c) must be true if (a) is not true. Because (a) and only (a) is correct, both (d) and (e) must be wrong. The correct answer to 5.7 is (c). Appropriate operationalization of variables is always a major problem in any research. In non-reactive, non-obtrusive studies, social desirability and other reactive effects are eliminated because the participants do not know that they are being observed — so (a) is incorrect, and (b) is meaningless. Behavioural traces are not a problem, as such. Instead, they are potentially a source of the dependent variable in non-reactive, non-obtrusive studies. (b) is therefore wrong, and (e) must also be wrong. The correct answer to 5.8 is (d). Because of this between-subjects design, neither (a) nor (b) is true because this study includes a different sample of participants at the two stages of testing. Because of this between-subjects design, (c) and (e) are all true, too — nothing else produces problems of testing. Because (a) is correct, neither (b) nor (c) can be correct. Because (a) and only (a) is correct, both (d) and (e) must be wrong. The correct answer to 5.8 is (e). Developing a suitable design and survey can involve any form of measurement variables. Because (a) is correct, neither (b) nor (c) can be correct. Because (a) is correct, neither (b) nor (c) can be correct. Because (a) is correct, neither (b) nor (c) can be correct.

b. Defining the probability of maturation for each member of the population.

c. Appropriate operationalization of variables.

d. Behavioural traces.

e. Both (c) and (d).

Test Yourself 5.8**

A team of researchers conducts a study in which they ask boys and girls from a local high school to complete a battery of psychological tests that investigate their social skills and levels of sociability. Ten years later they ask boys and girls from the same school to perform the same tests. Which of the following statements is false?

a. The study has a successive cross-sectional design.

b. The research is invalid because developmental studies should not include psychological tests.

c. Testing effects are eliminated by using a between-subjects design.

d. Mortality effects are eliminated by using a between-subjects design.

e. Maturation effects are eliminated by using a between-subjects design.
Discussion/essay questions

a. What are the main differences between the surveys typically used in psychological research and opinion polls?
b. Under what conditions should researchers consider using non-obtrusive measures?
c. Discuss the strengths and limitations of the various sampling strategies available to researchers using surveys.
d. Is the inability to draw causal inferences an insurmountable limitation of the survey method?
e. Discuss the strengths and limitations of telephone-based survey techniques.

Exercise

a. Researchers want to find out whether men and women have the same level of colour vision. They select 100 men and 100 women by giving volunteers a test of colour vision in a shopping centre.
   i. What sort of sampling method have the researchers used?
   ii. Is the sample likely to be representative of the population in general?
   iii. Is the sampling method appropriate given the research question?

b. An economic psychologist, Maisie, conducts a survey into people’s spending habits and financial plans. The questionnaire she designs contains the following items. What are the problems with these, and how might they be improved?
   i. How much do you earn?
   ii. Do you ever worry that you haven’t got enough money? (tick one) yes no
   iii. Are you in a pension plan and a superannuation scheme? (tick one) yes no
   iv. Do you have an ISA? (tick one) yes no
   v. On the scale below indicate your level of agreement with the following statement (circle one number)
      I do not always organize my finances as well as I might agree 1 2 3 4 5 6 7 disagree
   vi. How much do you agree with the following statement? (circle one)
      I organize my finances quiet well disagree completely 1 2 3 4 5 6 7 agree completely