SAMPLE CHAPTER:
PSYCHOLOGICAL RESEARCH AND SCIENTIFIC METHODS

Crown House Publishing Limited
www.crownhouse.co.uk

A2 LEVEL Psychology

the student's textbook

Nigel Holt and Rob Lewis

Psychological Research and Scientific Methods

NIGEL HOLT

WHAT YOU NEED TO KNOW

This section builds on the knowledge and skills of research methods developed at AS level.

You are expected to be able to understand the application of scientific method in psychology, design investigations, understand how to analyse and interpret data arising from such investigations, and report on practical investigations.

METHODS AND TECHNIQUES

- The major features of science
- The scientific process, including theory construction, hypothesis testing, use of empirical methods, generation of laws/principles
- Validating new knowledge and the role of peer review

INVESTIGATION DESIGN

- Selection and application of appropriate research methods
- Implications of sampling strategies
- Issues of reliability, including types of reliability, assessment of reliability, improving reliability
- Assessing and improving validity (internal and external)
- Ethical considerations in design and conduct of psychological research

DATA ANALYSIS AND PRESENTATION

- Appropriate selection of graphical representations
- Probability and significance, including the interpretation of significance and Type 1/Type 2 errors
- Factors affecting choice of statistical test, including levels of measurement
- The use of inferential analysis, including Spearman's Rho, Mann-Whitney, Wilcoxon, chi-square
- Analysis and interpretation of qualitative data
- Conventions of reporting on psychological investigations

THE APPLICATION OF SCIENTIFIC METHOD IN PSYCHOLOGY

RESEARCH METHODS IS NOT JUST MATHS

Research methods is one of the areas in psychology that some students find extremely difficult, and that most students have trouble with from time to time. Many people doing psychology confuse 'statistics' with 'research methods'. It's true that you do have to use some mathematical skills, but really methods are the vital skills that allow you to carry out research in psychology, and to look carefully and critically at the research that has already been done. A decent knowledge of research methods will enable you to look at the work of psychologists and to criticise their methods and findings. Evaluating research in this way allows us to think clearly about whether the work is ***reliable*** and ***valid***, two terms that are extremely important in psychology, and which will feature later in this chapter.

Another problem people have is that they think that 'research methods' is just another separate area of psychology. That's partly the fault of textbook writers like us, and partly the fault of generations of psychologists. It just so happens that research methods tends to feature as a separate chapter and is therefore regarded as something you read about once, and forget about for the rest of the book. We would urge you to regard research methods as central to your study of psychology, with all the other bits (developmental, cognitive, social, etc.) as separate important extras. Methods will provide the backbone and the other areas will provide the fabulous and interesting covering. Without one you cannot have the other.

ASK AN EXAMINER

If you are using this book as part of an A Level course you will already be familiar with a good deal of the information you need to know about the scientific method, since it has been covered in the AS course. Much of it will be summarised as we go through this chapter. Methods can be tricky in places: there's no getting around the fact, but a sensible understanding and knowledge of where to find refreshment for an overworked memory can do wonders for confidence. There is nothing at all to say that you cannot flick back to your AS books, or even get a study guide to help you summarise what you need to know. Remember, research methods at A2 extends the knowledge and understanding of this area that you gained at AS.

THE MAJOR FEATURES OF SCIENCE

By now you should be quite accustomed to the idea that psychology is a science. Many people are not at all familiar or comfortable with our subject being described in this way. They feel that science is something that people in white coats do while locked up in laboratories housing dangerous chemicals or machines. They feel that because psychology often deals with opinions and emotions it is not a science at all. They are wrong. Psychology follows the principles of all good sciences. These are the application of the scientific method, with careful consideration of replicability and objectivity.

There is no great secret to identifying the scientific method. Your previous studies have already introduced you to the concept. The scientific method refers to ways of thinking about evidence and of carrying out research that allow us to develop what we already know. It is because of the scientific method that we now have the technology we do, and it is largely because of the application of the scientific method that we now know as much we do about psychology.

HYPOTHESIS TESTING

The scientific method progresses by hypothesis testing. You know by now that a hypothesis is a statement that can be tested in research. For instance, we might be interested in investigating whether people really do feel happier in the summer when there is more sunshine. Our hypothesis might be:

'People feel happier in the summer.'

We must now do our best to test this hypothesis. We must try everything we can to disprove it. Our efforts must be focused on finding out whether we can prove that people do not feel happier in the summer. If we can do that then we can reject our hypothesis. The process of trying to disprove a hypothesis is absolutely central to the progression of all sciences, and psychology is no different in that respect from physics, chemistry or biology.

Extract from A2 Psychology: The Student's Textbook © Nigel Holt and Rob Lewis ISBN: 9781845901004 www.crownhouse.co.uk

REPLICABILITY

If we can carry out our research again at another time and find the same or similar results then we can say that we have managed to replicate our findings. Being able to do this is extremely important in the scientific method. If we find something on Monday and carry out the same research on Wednesday but find a different result we might, understandably, be a little cautious about Monday's hypothesis. After all, if we have found something on Monday, why *would* the effect not be there still on Wednesday? A high level of consistency (we use the term 'reliability' in psychology) means that we and those reading about our research can be confident that our findings are really what we say they are. Other researchers in our field (our *peers*) can then extend our research without having to carry out our work again. This process further improves our knowledge of psychology. Our subject is pushed on and on, developing every time a piece of work is carried out and published for others to read.

OBJECTIVITY

A good scientist should always be an objective scientist. This means that the findings of a piece of research should not depend on the person that did the research in the first place. The results should not be influenced by the person who designed the study, carried it out, analysed the results or drew the conclusions. A high level of objectivity increases other people's confidence in the results, as they are able to say that it would have made no difference who did the work, the results would have been the same. If the researcher was not objective then the results might have been biased or influenced in some way by his or her ideas and feelings. In cases like these we say that the research was subjective – the researcher's opinion influenced the study and the interpretation of the results. This means that another researcher doing exactly the same thing might not find the same results. For instance, a researcher is investigating the statements of people who said they had seen UFOs, to determine whether there was any consistency in the statements that might lead us to believe that UFOs really do exist. If the research is objective then true findings will result from this study. If the researcher's approach is subjective, the findings may be distorted or biased – for example, a researcher who believes that he has at one time been abducted by aliens might, if investigating the existence of UFOs, allow his conclusions to be influenced by this belief.

ASK AN EXAMINER

In the exam you might be asked to comment on whether something is a science or not. You might be asked to explain why psychology is a science, or what makes something a science and something else not. To answer these types of questions you need to know what constitutes a science, so make sure you read this section carefully!

THE SCIENTIFIC PROCESS

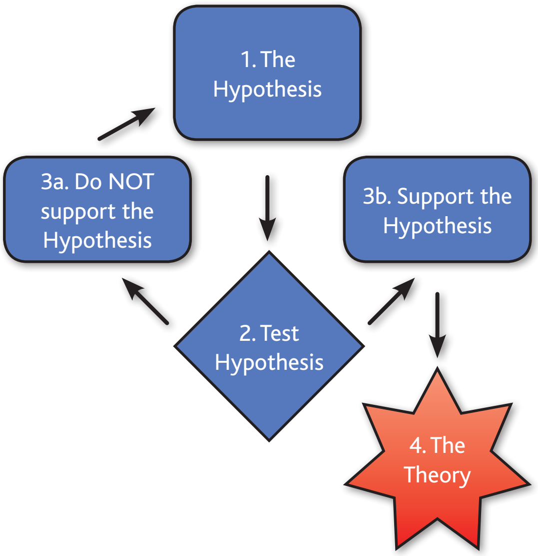
The process of science is always to strive for the truth. The end point of the scientific method is a theory that explains what it is that we are investigating. Building a theory is the end point of what can sometimes seem like a long process. Science does not stop there. Once a theory is built it is developed and modified by research that continues to be carried out. For instance, a well-known theory of memory, the Working Memory Model, has been developed, added to and altered on many occasions since its original introduction.

EMPIRICAL METHOD AND DEDUCTIVE REASONING

An empirical method is one that allows us to collect data that can lead us to a conclusion. In psychology, an empirical method is one which permits us to observe or measure some aspect of behaviour and use our findings either to develop existing theories or to introduce new ones. This is central to our understanding of psychology as a science.

By employing empirical methods that allow us to develop hypotheses which we can go out and test, we are engaging in a process called *deductive reasoning* – we have a hypothesis or theory we assume to be correct and we test it in various ways to see if it stands up. It follows from this that psychology is only a science as long as its theories are actually testable, using the scientific method. This is why the issue of whether psychology is a science is still debated by scientists and philosophers, since many psychological theories cannot be tested empirically.

This process, often called the *hypothetico-deductive method*, is shown in the following diagram.



1. The Hypothesis

The scientific process begins with the hypothesis: this is the statement, based on the aims of our research, that will be tested.

2. Test Hypothesis

Appropriate methods of investigation are chosen that allow us to test the hypothesis.

3a. Do NOT support the Hypothesis

If the results of the tests are inconsistent with the hypothesis and therefore do not allow us to support it, we go back to the drawing board, and rethink the hypothesis itself. The procedure begins again, flowing through points 1, 2 and 3.

3b. Support the Hypothesis

If the results of the tests chosen in point 2 are consistent with the hypothesis then we can accept it and move on to our final goal.

4. The Theory

The findings of carefully constructed research that support the hypothesis allow us to begin to form a theory or modify one that already exists.

THEORY CONSTRUCTION

Theories are constructed by using the route of hypothesis testing, development and retesting that we have identified here as the scientific process. In the majority of cases a theory does not depend on one piece of research. It is the end result of a range of work, usually from a number of different sources and researchers. Once the theory is developed it continues to be tested – and this brings us to a very important point. The most important aspect of a scientific theory is that it must be testable and ultimately falsifiable. This may sound odd, but a good theory must always be falsifiable. By this we mean that researchers must be able to develop ideas that can ultimately end in the rejection or development of the theory. Some of the most famous theories ever developed in science are just theories and are still being tested. One day they too may be shown to be incorrect in some way, making room for new theories. For instance, Einstein developed the theory of relativity in a series of scientific works between 1905 and 1915. While a huge amount of research in science has taken place on the assumption that this famous idea is entirely correct, the theory may yet be shown to be wrong. If this is the case then it may be developed or rejected altogether. This is the hallmark of a good, scientific theory.

The problem of unfalsifiability is one of the criticisms of the work of Sigmund Freud. His theory of the subconscious is that it is made up of three interacting components – the id, ego and superego. This idea is unfalsifiable. It cannot be tested and as such is not regarded as truly scientific. It seems that some parts of psychology are more scientific than others!

THE GENERATION OF LAWS/PRINCIPLES

A scientific law, or principle, is a statement that describes the behaviour of things in the real world. For instance, Newton described a series of laws that tell us how items behave in the universe. Newton's first law, for instance, says that every object in a uniform state of motion tends to remain in that state of motion until an external force is applied to it. This means that if something is moving along it will continue to move along in the same way unless it is influenced by something else. This is a law. It always happens. A scientific law or principle can be thought of as a development of a scientific theory. The theory is tested and retested and, once there is enough evidence in its favour, it can be developed into laws and principles.

INDUCTIVE REASONING

Let's return for a moment to Newton's laws. How can we really accept Newton's ideas as laws when it is impossible to have tested every single incidence of an object bumping into another object. We cannot possibly know whether Newton's ideas hold true all the time so how can they be regarded as laws? The answer to this lies in the principle of *inductive reasoning*. If something usually works and if every observed case of something provides the same result, we can *induce* from this that the same thing will happen for all cases. In other words, we can *generalise* our law and apply it in many situations to explain other things. Here's an example. What colour are crows? You have not seen every crow in the world, but no doubt you have seen quite a few. The ones you will have seen are all black. You therefore induce from your observations that all crows are black. This example also demonstrates the value of inductive reasoning in science. Conclusions based on inductive reasoning are only ever generalisations based on limited evidence (for example, the evidence of your observations that all the crows you have ever seen are black). We can make inductive arguments more probable by adding evidence in their favour, or we can question them with contradictory evidence (for example, we go out and look for a crow that is not black). Either way, we are increasing our understanding through science.

PARSIMONY

There is often more than one explanation for a hypothesis. One of them might be quite complicated, requiring complex arguments and further untested assumptions. On the other hand, another explanation might be much more straightforward, with its assumption based firmly on what we already know. Despite the fact that both accounts explain the hypothesis, the principle of parsimony states that we should select the least complex one. In this way, parsimony ensures that our explanation does not go beyond the available empirical evidence. We can see the principle of parsimony in the logical principle of Occam's razor. This is a principle identified by a monk called William of Ockham in the 14th century which basically states that, given two or more explanations for an

GREAT THINKERS: KARL POPPER

Karl Popper (1902–1994) provided a vital way of thinking about science that has led to today's understanding of hypothesis testing and theory construction. It was Popper who said that a theory must be falsifiable in order to be considered as scientific. Any amount of evidence can be presented to support a theory, but only a single piece of evidence that shows the theory to be incorrect is needed for it to be falsified. Popper went on to use this central idea of science to attack the study of psychoanalysis and the work of Sigmund Freud as unfalsifiable and therefore not scientific.

GREAT THINKERS: THOMAS KUHN

Thomas Kuhn (1922–1996) wrote a very important book called ***The Structure of Scientific Revolutions***. In it he says that science progresses quite happily most of the time but occasionally it undergoes a sort of shake up, which is described as a ***paradigm shift***. A paradigm is a way of thinking about something. If, perhaps because of the importance of a small group of researchers, people begin to think about their science differently, then the direction taken by science will change. Psychodynamic theory, behaviourism and biological ways of thinking about psychology, for example, can all be described as paradigms.

Kuhn says that each paradigm, or way of thinking, has its own framework of assumptions. Because of this, someone working in one paradigm cannot provide evidence in support of, or falsifying, a theory in another paradigm, because the principles upon which each paradigm is based are different. In psychology we can think of it as being illustrated by the example that research into the ideas of Freud under the paradigm of psychodynamicism cannot falsify research provided by medical psychologists in the neuropsychological paradigm. This is because the psychodynamic paradigm assumes the existence of the subconscious, and it is partly upon this assumption that the research is built. The neuropsychological paradigm is concerned with the physical activity of the brain and does not accept that the subconscious exists. It follows that information from the one paradigm cannot be used to falsify theories from the other.

The implications of this idea are huge for those interested in driving science forward and testing and building theories. Earlier in this chapter we discussed one of the principles of science, formalised by Popper, that a theory must be falsifiable. If the work of a researcher shows that a theory is incorrect, that theory will be rejected or changed to accommodate the new findings. If Kuhn's thinking is to be accepted, a paradigm shift will change this idea completely. If the discipline undergoes a change in direction, brought about by a paradigm shift, then the findings of a researcher, working under assumptions of an old paradigm, may be seen not as 'important, and falsifying the theory'. Instead they may be seen as a mistake made by the researcher, and the new paradigm theory continues, unchanged, simply because an old paradigm was used to derive the results that would otherwise have falsified the theory.

observed event, we should select the simplest one as the better. It's as straightforward as that.

VALIDATING NEW KNOWLEDGE

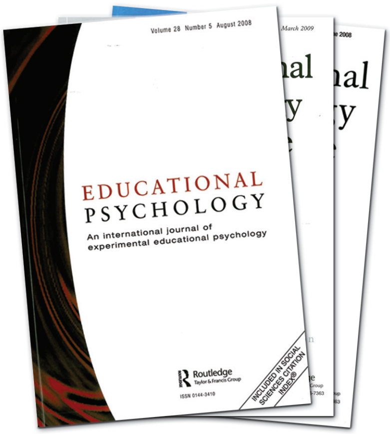
Just because a researcher carries out some work that he or she is convinced will change the world, this does not mean that it will be accepted by the rest of the scientific community. Having work published really gives it a stamp of approval. If something appears in print then other scientists can have access to it, and thus have the chance to read and challenge it. Traditionally, research is published in scientific journals. These are magazine-like publications which appear one or more times a year. Getting research published however is not as easy as just sending it to the publisher. In science the principle of peer review is applied.

PEER REVIEW

When a psychologist completes a piece of research it is written in a generally agreed format and then sent to a psychological journal for consideration. If the journal thinks the work is good enough they will publish it for all to see. The process of assessing whether it is good enough for publication is normally peer review. The work is sent to one or more established scientists with similar expertise (who usually remain anonymous) for them to read and perhaps criticise its method and thinking. It is then sent back to the researchers with recommendations for revision. If the work does not meet the standards of the peer reviewers it will not be published. This system ensures that standards of quality are maintained in research and that unsubstantiated claims are not made.

The peer review system is certainly very highly regarded but it is not without its critics. For instance, just because something is not thought by a reviewer to be appropriate for publishing, this does not mean that others would feel the same way. It could be that the reviewer is providing a fair assessment, in which case the work should be rejected and looked at by the researchers again. It could be, however, that the reviewer is biased in some way. It may be that the paper under consideration is expressing an opinion that the reviewer disagrees with. If this is the case the reviewer may be approaching the review from a position of bias. In situations like these, research may not be widely seen by the community, just because of the opinions of a single reviewer.

Some peer reviewed academic journals are more highly thought of than others, and researchers try their best to have their work published in them. For instance, the journals *Nature* and *Science* are regarded as extremely prestigious. Universities encourage their researchers to publish in such journals and not in less prestigious titles – and for this reason, some research, if rejected by highly respected journals, may never see the light of day because researchers feel "If it's not good enough for *Nature*, it's just not good enough."



Academic journals are extraordinarily expensive to buy. Many university libraries can only afford a relatively small number each year, and most provide the journals online through very expensive licensing agreements, with annual costs often running into hundreds of thousands of pounds. This means that the general public, or less wealthy colleges and universities, may not have access to the published research and so it remains relatively unread. Its use in advancing psychology is therefore limited. For reasons like these, exciting alternatives to traditional publishing have appeared – the latest of which is open access.

With open access, new knowledge can be viewed by the scientific community as well as the general public, allowing everyone to make up their own minds about the research. Open access can be combined with a form of peer review, as with the online journal *Philica.com*. On sites like these new ideas are rated by academics. Those rated as good contain ideas that might be trusted; those rated as bad are less likely to be trusted.

Despite numerous obstacles and sometimes long time scales involved in getting research published, the publication of psychological research is important for many reasons. Journals are international, which means that there will be widespread dissemination of the new research among peers. The work will therefore be more likely to be discussed and debated and the ideas possibly built upon

THE PUBLISHING PROCESS

Researcher submits an article to a journal. The choice of journal may be determined by:

- The journal's audience: is it the appropriate audience for the article?
- The journal's prestige: is it well known? Is it often cited?

The author is unpaid, but may pay 'page charges' (see terminology).

The journal selects two or more appropriate experts to peer review the article, without payment.

The peer reviewers assess:

- The quality of the research and the way it is reported
- The relevance of the article to the journal's readership
- Its novelty and interest
- Its content, structure and language.

Feedback from the reviewers determines whether or not the article is accepted. Acceptance rates vary from journal to journal.

The rejected article is returned to the author or the accepted article is passed to the editors employed by the publishers, either in-house or freelance.

The editors ensure that:

- The language of the article (particularly if it has been produced by a non-native speaker) is clear and unambiguous
- The standard style of nomenclature is adhered to
- Illustrative material is of a sufficiently high standard.

Editors also establish live links in the electronic version to the references cited and to other material such as data sets.

Where a journal exists in both paper and digital formats, the article is sent to the printer to produce a paper copy; and will be formatted for the online version of the journal.

Reproduced from: ***Select Committee on Science and Technology: Tenth Report***

with further research. The whole community will benefit from news of new developments, and because peer review ensures that only the best research gets published, the scientific community can be assured of the quality of the research.

Publication in peer reviewed journals can do a great deal to enhance the reputation of researchers. Not only are they more esteemed but, because of a good research and publication record, they also stand a better chance of obtaining funding for further research. Published research also provides enormous benefits for the institutions the researchers work for. Most research takes place in universities, which are funded by government grants. The amount of money a university gets is partly based on periodic audits of the university's research output, including the amount of published work they generate (the process is called the Research Assessment Exercise, or RAE). The publication of journal articles is therefore of crucial importance to both the researcher and their respective institutions.

Extract from A2 Psychology: The Student's Textbook © Nigel Holt and Rob Lewis ISBN: 9781845901004 www.crownhouse.co.uk

DESIGNING PSYCHOLOGICAL INVESTIGATIONS

In AS psychology you will have had experience of a number of investigative methods and techniques that are available to psychologists. Which they choose will depend on the type of research they are doing.

CARRYING OUT RESEARCH

Psychology is a science and psychologists rely on scientific methods of knowledge acquisition in just the same way as do other scientists, such as chemists and biologists. The goal of psychology is to collect information about human behaviour and this is done by careful observation and measurement, which provides *empirical* evidence.

Research must be carried out carefully, and psychologists must be aware of issues that could influence their results: this includes the requirement to be as objective as possible. One way to be sure of taking appropriate care in our research is to follow certain steps to ensure we really are investigating what we want to investigate and asking the right sort of questions in our work. These steps are as follows:

Step 1: AIMS Be sure of the aim of your research. For instance, you might be interested in investigating whether those that smoke cigarettes have a worse memory than those that do not smoke cigarettes. Your *aim* here is to investigate the effects of smoking on memory.

Step 2: HYPOTHESIS Based on your aims, your hypothesis makes a statement that you can set about testing. It can be directional, and predict a certain type of influence on behaviour (Smoking makes memory ability worse), or it can be non-directional, where a change in behaviour in either direction is not predicted, just that there will be a change (Smoking will *alter* people's memory ability). In the latter case memory ability may get better or it may get worse.

Step 3: IDENTIFY VARIABLES Variables are things that change in the research. The one you control and change is called the *independent variable*. The one that you measure is called the *dependent variable*. Anything that *might* change the dependent variable other than the independent variable is called an *extraneous variable*. If the extraneous variable *does* influence the dependent variable it has confounded your results and is referred to as a *confounding variable*.

Step 4: OPERATIONALISE VARIABLES This rather grand term simply means 'make them happen', i.e. make the variables measurable. So, if your independent variable is 'smoking', you will operationalise this by choosing some people who smoke and some who do not. Your dependent variable is 'memory ability' so you can operationalise this by designing a memory test, remembering a list of words perhaps, or a string of numbers – the more a person can remember, the better their memory.

Step 5: DECIDE ON A METHOD There may be a number of methods available for your study, and the choice depends in part on what it is specifically you are investigating.

Step 6: CONSIDER ETHICS Ethics are a vital part of research. Ethical principles include informed consent, deception, debriefing, the right to withdraw, confidentiality and protection.

Step 7: LOCATE YOUR POPULATION AND SAMPLE We choose a sample of a population. There are essentially three ways open to you: random, opportunity and volunteer sampling.

Step 8: PILOT IT and COLLECT YOUR DATA A pilot is a smaller scale version of the main study, used as a sort of 'dry run'. It helps you iron out any problems. When the pilot is complete and you are happy with the procedure, gather the data for real.

Step 9: ANALYSIS This means looking at what the data are telling us, i.e. what they mean. Later in this chapter we'll add more about this idea of analysis.

Step 10: PRESENT YOUR DATA There's more on this later on, but essentially you need to decide on the best way to present your data. You can do this with numbers, using tables to show your descriptive statistics (measures of central tendency and dispersion). You might also want to use a graph.

Step 11: FINDINGS AND CONCLUSIONS Look carefully at your results and decide what they show in terms of your hypothesis and psychological theory.

SELECTING APPROPRIATE RESEARCH METHODS

The good researcher is aware of a number of different methods and techniques. In reality, many professional researchers tend to concentrate on one or two methods because their research is very focused on particular areas, but for new psychologists a good overall knowledge is required.

It is useful to make sure that you know the difference between an experimental design and a non-experimental design. In a true laboratory experiment, it is the researcher who makes changes to variables and measures the result of those changes. For instance, in an experiment to measure the influence of chocolate eating on feelings of happiness it will be the experimenter who will give the participants one, two or three pieces of chocolate to eat, and who will measure, after each is consumed, how happy the person is. It is the experimenter who is in charge of the independent variable, in this case the chocolate. In a non-experimental design a variable is not manipulated by the researcher.

ASK AN EXAMINER

*Much of this material will be familiar to you from AS level. However, because it is vital that you understand **all** the research methods you cover on the A Level course, we will now provide a useful overview. Remember – you are expected to know this stuff! Be prepared to revisit your AS work to expand on what we present here. For example, make sure you are aware of the advantages and weaknesses of the various methods.*

EXPERIMENTAL DESIGNS

When we carry out an experiment, we change something and see what happens to something else. As we've already said, the something we change is called the independent variable, and the thing that changes as a result is the dependent variable. Once you've decided that you are running an experiment there are three designs open to you. Each has its advantages and weaknesses.

ASK AN EXAMINER

Think carefully about the three different design choices. If you are asked in the exam to evaluate the methodology used in a piece of experimental research, displaying a knowledge of the different types of designs and the advantages and weaknesses of each would be an ideal way to gain marks.

To make things simpler, we will discuss three different designs for the same experiment. Imagine you were interested in investigating whether it was possible to stand on one leg longer in silence than when listening to music. You never know, you might be interested in something like that. Your independent variable would be whether there was music, or whether there was silence: it has two levels, music and silence. Your dependent variable (the thing you measured) would be the time the person managed to spend standing on one leg without falling over. The three possible designs for your experiment are identified and described below.

1. REPEATED MEASURES DESIGN

A repeated measures design is one where each participant takes part in each level of the independent variable. In our imagined experiment, this means that each participant does the test in silence and then each of them does the test with music.

Advantages

The same person takes part in each level of the independent variable and so there are no problems of individual differences, such as fitness and age. This means you do not need to use a huge number of participants, so the research can be completed sooner and, if you are paying your participants, for less money.

Weaknesses

There are possible order effects. This means that the time spent standing on one leg in silence may have influenced the time spent on one leg when listening to music. The person may have become tired and fallen over more easily on the second occasion. This is called a fatigue effect. Similarly, the person may have become more used to standing on one leg, lasting longer on the second occasion. This is called a practice effect.

2. INDEPENDENT GROUPS

An independent groups design uses two groups of participants. One group stands on one leg in silence and one group stands on one leg while listening to music.

Advantages

The major advantage of an independent groups design is that it completely eliminates order effects such as practice and fatigue.

Weaknesses

The difficulty with independent groups designs is that they are open to problems associated with individual differences. For instance, imagine the UK 'standing on one leg' Olympic team was in one group, and the old ladies' lawn bowls team was in the other group. That would hardly make a fair comparison! The way to avoid this is to have two huge groups. If you do that, the range of individual differences in one group would be similar to the range in the other group.

3. MATCHED PAIRS

The matched pairs design is like an independent groups design, but each person in one group is matched with a person in the other group. For instance, if you had an 82-year-old pensioner in one group, you should match them with an 82-year-old pensioner in the other group. Each person in each group has a matched person in the other group.

Advantages

The great thing about matching is that it controls for individual differences, and since it is a variation of an independent groups design, there are no problems with order effects either.

Weaknesses

Matching is extremely difficult, or even impossible. Even if you do have an 82-year-old pensioner in each group you are not to know whether one has had a huge amount of 'standing on one leg' experience while the other may be an absolute novice. In fact, even if you had identical twins, and allocated one to each group, you could not be sure whether their experiences and skills were identical. These individual differences have not been matched, and may influence your result. It is impossible to match people in different groups perfectly.

THE LABORATORY EXPERIMENT

In general, the laboratory experiment is run in carefully controlled conditions, and uses a standardised procedure. Variables are manipulated (the independent variable), are controlled (extraneous or confounding variables) and measurements are carefully taken (the dependent variable).

A good laboratory experiment has the following characteristics. It is…

Replicable – it can be repeated by other researchers

Generalisable – the results can refer to people outside the sample that you tested

Reliable – a repeat of the experiment will yield the same results

Valid – It measures what it says it measures.

THE FIELD EXPERIMENT

This is like a laboratory experiment, but the variables are manipulated by the researcher out in a more real-world setting. The advantages of this are that the real-world setting offers ecological validity and the people acting as participants in the experiment often need not know they are part of the study, eliminating any demand characteristics. The problem however is that these studies are often extremely expensive and time consuming to carry out.

THE NATURAL EXPERIMENT

Here the researcher makes use of a naturally occurring independent variable. For instance, the researcher may have the idea that towns near the seaside are perceived as more fun than towns inland. The researcher cannot change the position of the sea and so the independent variable, in this case proximity to the sea, is naturally occurring. The advantages of the natural experiment are those of the field experiment: it is high in ecological validity and there are few or no demand characteristics. The disadvantages are that the researcher has no control at all over the independent variable and so the chance of unwanted things influencing the experimental findings (confounding variables) is quite high.

THE QUASI EXPERIMENT

This looks rather like a lab experiment but it's not. Imagine you wanted to see whether Americans or English people were better at mathematics. You would locate a group of Americans and a group of English people and give each group a maths test. The allocation of participants to a group is done on the basis of their nationality, your independent variable. However, nationality cannot be manipulated, and so it's not a true experiment.

THE CASE STUDY

Some research offers a careful and systematic investigation of the case of an individual. Examples

LABORATORY EXPERIMENT	
ADVANTAGES	**WEAKNESSES**
• Replicability	• Cannot always use a lab experiment
• Easy to control variables	• More control means less natural
• Controllability	• May be demand characteristics and experimenter effects, both of which may influence the results

Extract from A2 Psychology: The Student's Textbook © Nigel Holt and Rob Lewis ISBN: 9781845901004 www.crownhouse.co.uk

might include Freud's investigation of Anna O, or the case of Phineas Gage.

The advantages of a case study are that it can be extremely detailed and may provide wonderfully rich and informative data about the person. The disadvantage of course is that generalisability is extremely low, as the case of that particular individual may well be unusual and unrelated to the rest of the population. Also, as case studies gather data from the past, the information may not be very reliable. Memory has an annoying habit of fading over time, and the responses a person may give regarding their past may not be entirely accurate.

THE OBSERVATIONAL METHOD

This is extremely obvious at first glance, and deceptively simple. It clearly means a method where researchers observe the behaviour of others. You will not be surprised to hear that it is not quite that simple. There are two main divisions of observational method. *Participant observation* is where the person doing the observing is part of the group they are observing. For instance, if observing the behaviour of a football team, the observer may be part of that team. *Non-participant observation* is where the observer is not part of the group they are observing. For instance, an observer may be investigating incidences of aggressive behaviour in the playgrounds of secondary schools. They themselves are not part of the group of students being observed. There are three things to watch out for in the observational method:

1. Awareness of the observer: If those being observed are aware of the person doing the observing then their behaviour may change and will not be natural.

2. Behavioural categories: The observer must decide what exactly they are looking for in the observation. These categories can form a checklist for use during the observation.

3. Bias: If the observer is not objective the results will be biased. Using multiple observers and comparing their observations after each session can overcome this problem.

Finally, the observational method can be carried out in two different ways. Having chosen the behaviour to be observed – for example, aggression – the observer may count how many incidences of the behaviour occur during a period of time (*time sampling*), or may record all incidences of the behaviour (*event sampling*).

ASK AN EXAMINER

When you think about observational methods, consider possible ethical implications of these methods. Is it ethically acceptable to observe someone who doesn't know they are being observed? Is it ethically acceptable to observe groups such as children, who may be regarded as 'vulnerable'?

OBSERVATIONAL METHODS VS. OBSERVATIONAL TECHNIQUES

Observation can be a *method* and observation can be a *technique*. The observational method is a type of research in its own right. The observational technique, however, is something that you can use in any type of research. It's best to think of the observational technique as a tool that researchers use and the observational method as a style of research that they undertake. For instance, you might use the observational technique to count up how many people on the high street are carrying sales bags after an advertising campaign by a local shop. You are using the technique of observing to measure the impact of the campaign in a natural experiment.

CONTENT ANALYSIS

Content analysis is a form of observational method, but instead of watching people behave, the researcher analyses the content of a text or film. For instance, they may be investigating the use in newspapers of terms that might be regarded as racist, and how this might have changed over the years. They would decide on the racist terms

SAMPLING OF OBSERVATIONS

ADVANTAGES	WEAKNESSES
• Natural settings, so natural behaviours are observed.	• Observer bias if researchers don't remain objective.
• Allows otherwise impossible or hard-to-example behaviours (e.g. aggression in children) to be investigated.	• Confounding variables may interfere, making it hard to determine the cause for something happening.
• Few demand characteristics because the observed are not put into a false situation and may not know they are being observed at all.	• Small groups are usually observed, making the results hard to generalise.

to watch out for and then look through newspapers marking down how often the terms were used.

SELF-REPORT TECHNIQUES

These include questionnaires and interviews. Here a participant records or expresses their own opinions and feelings about something. Both techniques have individual strengths and weaknesses, but both share the weakness that they are self-report and so the truthfulness of the answers may be in question as participants may respond in such a way as to make themselves appear socially desirable.

A questionnaire is a list of prewritten questions. These may be 'closed', where alternative responses are offered; or 'open', where free responses are offered; or possibly rank-order, where you are required to place things in some kind of order. The advantages of questionnaires are that they are a cheap way of collecting lots of data while retaining participant anonymity. The disadvantages are that those who agree to complete the questionnaires may be doing so because they like to help out, or the topic covered is of interest to them. The sample may, in this way, be biased. Also, some people chosen to complete the questionnaire at 'random' may decide not to do so, which means your random sample is no longer random, as you have to choose someone else to take part.

An interview is like a questionnaire, in that it provides self-report data drawn directly from a participant. Interviews may be 'structured', where the questions are very organised and rigid; they may be 'semi-structured', where the control and organisation is less rigid and the interviewer is offered more flexibility; or they may be 'unstructured', where the interview is completely free and uncontrolled. The advantages of the interview are that it can provide very rich and insightful data and can be very simple and fast to carry out, with generalisable findings if the sample is big enough. The disadvantages are those of the questionnaire. In addition to this, a person may provide quite different responses to the same questions in an interview on different occasions, thus threatening the validity of the results.

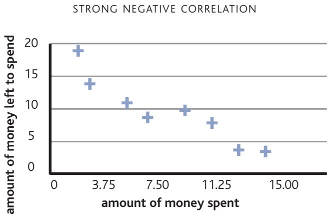
CORRELATION METHOD

Correlational analysis lets us see how two variables are related. For instance, height and shoe size; or ability at maths and ability at chess. Just as with observation, the correlational method is a research design option in its own right, but correlation analysis can be used with data from any other method where appropriate. The analysis provides a statistic called the correlation coefficient, which varies from –1 to +1.

A correlation coefficient of near to –1 indicates a strong negative correlation, which means as one variable increases in size the other decreases. For instance, the more money you spend, the less money you have to spend. This may not be a perfect negative correlation because we have not accounted for the amount of money you may be earning, so occasionally spending some money may not, in fact, result in a reduction in the amount that remains, if some wages have been paid to you.

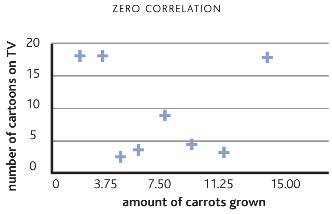
A correlation coefficient of 0 means that there is no relationship at all between the two variables. For instance, there is no relationship between the amount of carrots grown in Israel and the number of cartoons on the TV between 3 and 4p.m. on a Tuesday afternoon.

NEGATIVE (LESS THAN 0)
A 'downward slope' is seen.



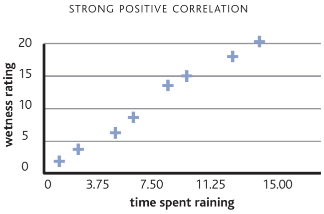
The more perfect the line, the closer to –1 the correlation is. A strong negative correlation is one close to a perfectly straight downward-sloping line.

ZERO CORRELATION
No slope is seen.



There seems to be no pattern here.

POSITIVE (MORE THAN 0) An 'upward slope' is seen.



The more perfect the line, the closer to +1 the correlation is. A strong positive correlation is one close to a perfectly straight upward-sloping line.

CORRELATIONS	
ADVANTAGES	**WEAKNESSES**
● Naturally occurring variables can be measured and a relationship between the two identified. This may lead to future research.	● Cannot imply cause or effect, and variables are almost impossible to control. Even if the analysis indicates that there is no relationship we cannot be certain of that. It could be that something else that we have not measured or considered is happening that the analysis is not sensitive to.

A correlation coefficient of near to +1 means that there is a strong positive correlation between the two variables. As one goes up, so does the other. For instance, the longer it rains, the wetter you will get if standing outside. This may not be a perfect positive correlation because it may be that at some point your clothing becomes completely saturated and, however much more rain falls on you, you cannot possibly get any wetter.

Correlations are represented on scattergraphs (also known as scattergrams). On page 13 we have three scatterplots identifying the relationships described above.

Do not think that correlation means cause. It does not. Just because two variables are positively correlated does not mean that one causes the other. For instance, it is extremely likely that were you to collect the data you would find that the number of cars on the roads of the United Kingdom has risen steadily over the last 10 years. You would also probably find that the number of people taking holidays abroad has gone up. The two variables (car sales and holidays abroad) are likely to be positively correlated. We cannot conclude, however, that increased car sales have caused people to take more holidays. There may be a third factor that we have not considered, such as 'amount of disposable income' that is related to both of these variables, but that we have not measured.

WATCH OUT! DEMAND CHARACTERISTICS AND INVESTIGATOR EFFECTS

There are a few things to keep in mind when designing research. Demand characteristics, investigator effects, reliability and validity can all cause problems for a researcher.

Demand characteristics – Something in a research design may change a participant's behaviour. If this happens then your data will not give you a true picture of what's going on in your study. For instance, participants who know something about the procedure may alter their behaviour to provide the researcher with the responses they *think* the researcher would like. Researchers often design 'single blind' procedures to eliminate, or at least attempt to eliminate, demand characteristics. These are procedures where participants are kept naive about the aims of the research. It is the job of the researcher to plan the study so as to minimise demand characteristics. Eliminating them is often very difficult indeed.

Investigator effects – These are a little like demand characteristics, but you should try not to mix them up! Research suffers with investigator effects when the researchers themselves influence the behaviour of the participant in some way. They may not do so on purpose, but they may give away their opinion in some way which may encourage

Investigator Effects

Demand Characteristics

the participant to respond differently for the rest of the session. Researchers may introduce a *double-blind* procedure to eliminate these effects. Here the researcher *and* the participant do not know of the aims of the research, and so neither can influence the results by inadvertently responding in certain ways.

ASK AN EXAMINER

Demand characteristics, investigator effects, reliability and validity are extremely useful things to consider when you comment on the research you read about. Have these in mind when thinking critically about the methodology used. You can often pick up valuable marks by making one or two effective comments based on these issues when evaluating research.

SAMPLING STRATEGIES

Finding participants is something that all researchers have to do. A group of participants is referred to as a *sample*. A researcher cannot test everyone in the world, and would not want to. The idea is to select a sample of people who can be described as representative of the wider population. If the sample really is representative then we can say that we are able to *generalise* our results to the larger group. There are three sampling strategies, each with strengths and weaknesses that have implications for your research.

RANDOM SAMPLING

This is where all of the people in a population have the same chance of being selected.

Generalisability and bias

The vulnerability of random samples to drop-outs means that the sample is easily biased, as the researcher must either go with a reduced sample or choose another random participant, who by the very fact that they were not originally 'randomly chosen' is not very random at all.

Similarly, it's very difficult to carry out a truly random sample if you have a very large population as it is unlikely that you will have a complete list of names from which to choose. This means that not everyone is equally likely to be chosen and so the sample is not truly random, but rather, it is biased towards 'people on your list'.

You must ask yourself how the names got onto the list in the first place. For instance, researchers may have used a telephone directory. If this was the case then the sample would be limited to those that own a phone and have their details listed. Some researchers may use the electoral register. If they do this then the sample is limited to those people who are 18 or over who have indicated that they wish to be able to vote in elections; and further limited because people whose names are on the full register are allowed to opt out of inclusion on the shorter version, to which researchers and others have access. In both these cases the sample is not generalisable to those in the population who do not meet these criteria.

OPPORTUNITY SAMPLING

This is where anyone handy is asked to participate. The most typical opportunity sample in psychology is one where researchers simply ask people in their department or college to take part. For instance, a researcher may put up a notice asking for people to help out in their research, or may simply approach people in the cafe and ask them for help.

Generalisability and bias

Typically, researchers choose people to take part from their own social group. They may, if not deliberately, choose people they would like to take part, whom they find attractive in some way. Or they may choose people who have taken part in previous research, and whom they trust. It may even be that the researchers only ask those who they think are least likely to reject their requests. Since choices are made, and the samples are drawn, from such a narrow population, the extent to which the findings can be generalised to the wider population is very limited.

Similarly, those who take part may have their own motives for doing so. At university, students may take part in their lecturer's research because they may feel that helping out will gain them more marks. This may leave the research open to bias in that the participant may attempt to provide data that will please the researcher in some way.

The other problem with an opportunity sample is that only those who want to take part do so. This means that you may find that the sample has a large proportion of 'helpful' or 'interested' participants, who may provide biased data. Obviously, this reduces the generalisability of the findings, as the sample is not reflective of the wider, general population. This means it has low ecological validity – we'll cover this more in a moment.

VOLUNTEER SAMPLING

This is also known as 'self-selected' sampling. Here people who want to take part do so. For example, a notice board or newspaper advertisement may be used to ask for volunteers to respond.

Generalisability and bias

The findings of the research may lack generalisability because a volunteer sample may consist of mostly helpful, interested, more motivated or obedient people and as such it may not be like the general population. In addition to this a volunteer sample may be open to bias as those who take part may do so in order to seem helpful and so please the researcher – who is often their psychology lecturer, and who therefore has the power to provide good marks on their work!

ISSUES OF RELIABILITY

Reliability refers to how well the research can be replicated at another time, or if two researchers carried out observations how well their opinions of what was observed agree. In short, reliability refers to the consistency of the research. As reliability increases so does our confidence in the results.

OBSERVER RELIABILITY

A widely used method and technique in psychological research is observation. Here the researchers observe behaviour, marking down what they see. Often they are looking for particular types or categories of behaviour. For example, researchers may be interested in investigating whether male children behave more aggressively at play when surrounded by other males or when in a mixed male/female environment. In this case researchers would observe the behaviour of males in a male-only play environment, recording the number of times they saw aggressive behaviour. They would then compare these results with similar observations made in a mixed male/female environment.

The results of research like this will not be of much use if the observations are not done in the same way – i.e. if they are not consistent. We need to ensure that another person observing the behaviour would come up with the same, or at least a very similar, result. For this reason two observers make observations at each session. Their results are compared for similarity. If the two sets of observations are similar then we can say that there is good observer reliability.

Assessing observer reliability

Comparing the observations of two observers is best done using correlation. This relatively simple statistical technique allows us to see how similar two sets of values are. A good positive correlation shows that the two observers provided similar results. If you can show a good positive correlation to two observations of the same event then you can say confidently that you had good observer reliability.

Improving observer reliability

1. Training

If the two observers are able to easily identify the type of behaviours they are watching then they will be able to record those behaviours efficiently. Practice certainly makes perfect as observation is a great skill. Until you try it, you'll not fully believe how fast the different behaviours come and how varied the behaviours that might be recorded are in different situations. For this reason, carefully training the observers before the research begins is advisable as a means of improving observer reliability.

2. Operationalisation

Both observers have to be clear exactly what they are looking for. For instance, if they are observing 'helping' behaviours then they must agree what this means. For instance, how 'helpful' does a behaviour need to be before they record it as 'helpful'? If the definitions of the behaviours to be observed are clear and carefully laid out then observer reliability will be better than if categories of behaviour are badly defined.

3. View of the behaviours

Each observer must have the same ability to see the behaviours being observed. The best way to do this is to have them both base their recordings of behaviour on exactly the same video film of the behaviours being recorded. This ensures that each observer sees exactly the same things. If the observations are recorded 'live', perhaps by watching behaviours in a school playground, then different sight lines may mean that incidences of a behaviour may be missed by one observer but recorded by the other. This would weaken the correlation between their results, and so weaken observer reliability.

TEST RELIABILITY

Various tests are employed in psychological research to measure behaviour. If a researcher is interested in assessing someone's personality then they can use a test carefully designed to do the job. These tests are often in the form of questionnaires and measurement scales that require the participants to indicate opinions of things, and how they might respond or behave in different situations.

Assessing test reliability

Assessing test reliability is very similar to assessing observer reliability. The method used here is *test–retest assessment*, or *test–retest correlation*. Here the test is given to the person again, on a different occasion. The person's responses to the test on each occasion are compared and a correlation carried out. If there is a high correlation between the responses on each presentation of the test then we can say that there is a high level of test–retest reliability. This means that the test measures what it says it does consistently.

Improving test reliability

Altering the test to improve correlation

If the test–retest correlation is low, then the reliability of the test is in question. This means that the researchers must alter the test to improve this reliability. They can do this by looking carefully at the test itself, and identifying the parts of it that did not correlate well on the two occasions it was given. They can then remove those components of the test that are weakening its reliability and replace them with alternative questions and tasks. This new test is then tested again, to see whether reliability has improved, using the test–retest correlation assessment. This process is repeated until a high level of test–retest reliability is shown by a strongly correlated first and second running of the test.

ASSESSING AND IMPROVING VALIDITY

Validity refers to how well the test or research actually measures what it says it measures. A tape measure, for instance, is a very valid tool for measuring height, but a stop-watch is not. Similarly, research that aims to measure the relationship between mathematics ability and age is valid if it does just that, but not if it actually measures the relationship between age and general intelligence.

INTERNAL VALIDITY

Internal validity is the extent to which we can say that our findings are truly to do with what we think they are to do with, rather than something that we have no control over. For instance, we may be interested in investigating whether memory ability is improved by eating chocolate. We must be sure that the person has not secretly drunk lots of cups of tea before the experiment. If the results of our research showed that their memory was indeed improved by chocolate eating we cannot be absolutely sure that the tea did not have something to do with it, and so the task has low internal validity.

Improving internal validity

Anything that influences the dependent variable (in this case the score on a memory task) rather than our independent variable (here it's chocolate eating) reduces the internal validity of our research. In general these things are described as extraneous variables. If an extraneous variable (such as fatigue or practice) influences our findings, then internal validity is reduced. One of the most important aspects of research design is to minimise the influences of extraneous variables. Improved internal validity comes from improved resistance of your research to the influence of extraneous variables. This is best achieved by very careful research design and planning.

EXTERNAL VALIDITY

The results of our research have external validity if they can be generalised from our sample to the general population, in other settings beyond those in which the research was carried out and at different times.

ECOLOGICAL VALIDITY

This is a form of external validity. If the research has high ecological validity then we can say that the results relate to different situations from those in which the research was carried out. Some work carried out in a laboratory, for instance, may lack ecological validity, because outside the carefully controlled environment of the laboratory people may behave differently.

ASK AN EXAMINER

Just because something is done in a lab does not mean that it lacks ecological validity. Findings from laboratory based research may transfer to other settings very well. Be careful of this in the exam. Don't just say 'It lacks ecological validity because it was done in a lab'. This won't get you marks. Be more specific. Explain what circumstances outside the lab might bring the ecological validity into question.

POPULATION VALIDITY

If research has high population validity then the findings can be said to relate to the general population. For instance, if a piece of research has used an opportunity sample of 10,000 people chosen from 20 countries worldwide, its findings are more likely to be relevant to the whole population of the world than if the sample had been three people taken from a cafe in Bristol. If we can say confidently that something has high population validity, then similar research carried out on a different sample elsewhere should provide similar results.

TEST VALIDITY

The validity of a test can be measured in three ways.

1. Content validity

The test being used is carefully scrutinised to see whether it really does test what it says it tests. This is usually done by experts who assess the theory behind the content to ensure that there are no omissions.

2. Face validity

This is like content validity but far less rigorous. Face validity refers to how valid a test seems to be 'on the face of it'. To be more certain, a content validity assessment should be made.

3. Predictive validity

This assesses how well a score obtained on the test at one point in time might predict a score obtained on the test at another time. For instance, if the test is carried out on a person when they are 20 years old, we might like to be able to predict their score on the test if carried out at 30 years old. A high predictive validity would allow us to do this.

ETHICAL CONSIDERATIONS IN PSYCHOLOGICAL RESEARCH

It is extremely important to design and carry out psychological research ethically. You will have covered ethical issues in research during your AS studies. A summary of the major issues is provided in the table below.

There is no reason why research should not be carried out ethically. In places where psychological research is conducted, work does not progress until a summary of the proposed investigation has been considered by a committee drawn from different disciplines and walks of life who look specifically at any ethical issues. This 'ethics committee' can agree to allow the research to go ahead as planned; it may opt to reject it until certain ethical issues have been considered; or it may decide to reject the proposal entirely. The committee acts as a safeguard to stop questionable research from taking place.

If ethically questionable research is conducted the penalties for the researchers can be severe. The work is unlikely to find a publisher, and the psychologists responsible may find themselves without access to further research funding: they may even be expelled from the British Psychological Society, the professional body for psychologists working in the UK. Other countries have their own professional bodies who take a similar view on unethical research.

SUMMARY OF ETHICAL CONSIDERATIONS WHEN DESIGNING A RESEARCH STUDY

ETHICAL ISSUE	EXPLANATION
INFORMED CONSENT	Participants must be told what they will be doing and why they are doing it so they can provide 'informed' consent.
DECEPTION	Participants should not be deceived unless absolutely necessary. If deception is required, great care and careful consideration must be given to the project.
DEBRIEFING	After the experiment is complete, participants must be 'debriefed' and informed of the motivations for the experiment. They must be given the chance to ask any questions they have.
RIGHT TO WITHDRAW	Participants should be free to leave the experiment at any time.
CONFIDENTIALITY	Any information and data provided by the participants must be confidential.
PROTECTION	The safety and well-being of the participants must be protected at all times.

DATA ANALYSIS AND REPORTING INVESTIGATIONS

APPROPRIATE SELECTION OF GRAPHS

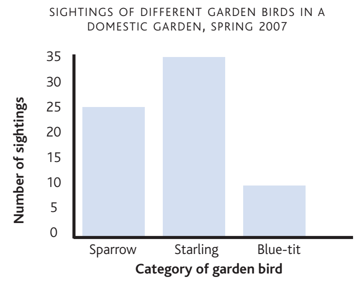
Data can be presented as graphs, tables or statistics. Always remember that the point of presenting data is to make your results easy to read for those interested in your research.

A graph should be simple and clear, and should be the appropriate type of graph for the data you have. A table should not be messy or complicated, or include too much information, and your statistics should be presented simply and as obviously as you can. Experience has shown us that some people, including professional researchers, do not take nearly enough care when presenting their data and it can really spoil what would otherwise have been an elegant and careful piece of work. Professional researchers will tell you that when looking over the many hundreds of papers that are produced in their field each year, they first look at the title of the article, then the summary at the start, then the results section where the research data is presented. If they find the presentation of the data interesting they may spend time reading the whole article from start to finish. If they do not find the presentation of the data clear then they may put the article to one side, maybe not returning to it at all. It is for this reason that the importance of data presentation should not be underestimated.

There are four types of graph to know about. We have already discussed one of them – the scattergraph – when we talked about correlations. Each type of graph is appropriate for different types of data. Do not draw all of them for all data sets, and do not make the mistake of just drawing your favourite type of graph. It may well not be appropriate.

It is important that you are familiar with graphs. It is much more likely that you will be required to read and interpret the information contained in graphs, rather than draw one. Make sure you practice this skill.

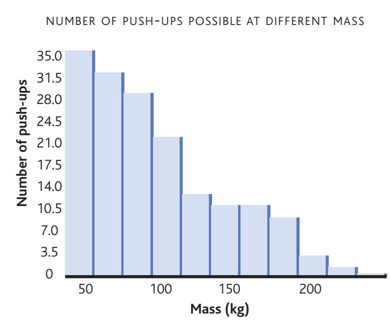
BAR CHART



The bar chart is an extremely useful and common way of presenting data. It is used to depict the number of incidences in a particular 'discrete' category. This means that the categories identified do not overlap in any way. In our example we have 'types of bird'. You cannot have a starling that is also a blue-tit. There is no overlap. You might also use this type of graph if you were depicting data separated by gender – for instance, the number of boys enrolled on a football summer school and the number of girls. You cannot have a boy who is also a girl. The two categories do not overlap: they are discrete. Bar charts have:

» Gaps between the bars
» Frequency (number) on the y-axis
» Category labels on the x-axis.

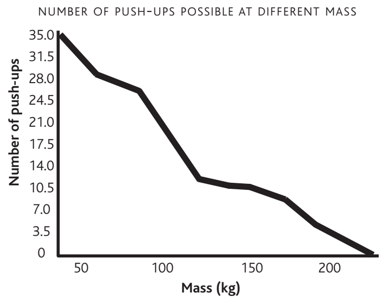
HISTOGRAM



The histogram is similar in many ways to the bar chart. The major difference is that in a histogram the x-axis does not depict discrete categories, rather it shows a continuous scale. In our example that scale is mass (kg). If, on the other hand, your data were concerned with temperature and fatigue, you would measure 'degrees Celsius' on the x-axis. Histograms have:

» No gaps between the bars
» A continuous variable on the x-axis.

FREQUENCY POLYGON



The frequency polygon is very similar to the histogram. Imagine a histogram with a point drawn in the centre of the top of each bar. Now, remove all the bars and join up the points. A frequency polygon has:

» A continuous variable on the x-axis

» A continuous line, no bars.

TABLES

The whole point of a table is to present otherwise complicated information in as simple a way as possible. A table should allow readers to find the information they need with as little effort as possible. Tables can be large, containing lots of data, or they may be small, containing data in the form of summary statistics, such as mean, mode and median. In the case below, we have a table showing the mean scores of male and female students in maths, English and science examinations.

	Male	Female
Maths	56	44
English	43	67
Science	51	52

When drawing a table make sure you follow the most important rule: be as clear and straightforward as you can.

WHICH IS BEST?

Each measure of central tendency and dispersion has its good and bad points. A strong descriptive statistic is generally one that takes in a good deal of the raw data in its calculation, so bear that in mind when thinking about which is most suitable. The strengths and weaknesses of each are described below.

	Mean	Mode	Median	Standard deviation	Range	Semi-interquartile range
Strength	Most powerful measure of central tendency as it uses all of the data	The best measure to use if you want to know how often things happen	Not heavily influenced by rogue scores	Uses every value in the data set, not heavily distorted by extreme values and is the most sensitive	Takes extreme scores into consideration and is simple to calculate	Less distorted scores than the range
Weakness	One rogue score (large or small) can heavily influence it. For instance, the mean of 3, 4 and **8** is 5. The mean of 3, 4, 8 and 1,005 is **255**. The extreme value has seriously influenced the mean	Sometimes a data set does not have a most common value and sometimes it has lots of common values	Not good for using with small data sets. For instance, if you only have the numbers 1, 17 and 2,000 in your data set the median is 17. Not very informative	The most laborious of the measures of central tendency to calculate	If either of the two scores are extreme, range will be distorted. It tells us little about how spread out or clustered together the data are	Uses only 50% of the data in the calculation and is quite laborious to calculate

It's not really useful to ask which of the measures is best and which is worst. Each is appropriate in different circumstances, depending on the data you have and what you are looking for in your research.

DESCRIPTIVE STATISTICS

Descriptive statistics are ways of representing raw data simply. Measures of central tendency and measures of dispersion are the two types of descriptive statistic with which you are familiar from your AS studies.

A measure of central tendency is sometimes referred to as an average. An average is the score that is typical of all the data you have in your set. Measures of dispersion tell us how spread out the data are. The larger the measure of dispersion, the more spread out the data.

Each measure of central tendency (mean, median and mode) has an appropriate measure of dispersion: this is easily described in a table.

"when using a	use a ...."
Mean	Standard deviation
Median	Semi-interquartile range, or range
Mode	Range

Calculating these values is not very difficult: it just takes a little time. This is how you make the calculations. It is worth refreshing your memory.

MEAN

1. Add up the numbers in your data set (call this number A).
2. Count up the number of values in your data set (call this number B).
3. Divide A by B.

MODE

1. Identify the most common value in your data set. This is the mode.
2. If there are two equally common numbers you have a 'bi-modal' set (2 modes).
3. If you have three equally common numbers you have a 'tri-modal' set (3 modes).

MEDIAN

1. Put your data in order (smallest to largest).
2. The data value in the middle is the median.
3. If you have an even number of values, take the mean of the two in the middle.

STANDARD DEVIATION

1. Calculate the mean of the values in your data set.
2. Subtract the mean from each value in turn and square the result in each case.
3. Add up all of the squared values from step 2.

MATHEMATICAL SYMBOLS – A HANDY GUIDE

The problem with maths is that it sometimes uses symbols that make it look more like Ancient Greek than something we might possibly understand. It's not that complicated actually. Here's a handy guide to what each symbol means.

Symbol	Meaning
$\sqrt{}$	square root
Σ	sigma. This just means 'the sum of'
N	the number of items in your data set
x	the value you are working with
$\bar{x}$	x-bar. This just means 'the mean of your data set'
2	multiply by itself. Also known as 'squared'.
σ	lower case sigma. This means standard deviation
=	equals. What's on the left of the symbol equals what is on the right
<	less than. What's on the left is less than what's on the right
>	greater than. What's on the left is greater than what's on the right
<<	a lot less than. What's on the left is a great deal less than what's on the right
>>	a lot greater than. What's on the left is a lot greater than what's on the right
~	approximately. What's on the left is approximately the same as what's on the right

4. Divide the result from step 3 by the number of values in your data set.
5. Take the square root of the figure you calculated in step 4.

RANGE

1. Find the largest and smallest values in your data set.
2. Take the smallest from the largest.

SEMI-INTERQUARTILE RANGE

1. Put your values in order (smallest to largest).
2. Count up how many values you have. Call this 'N'.
3. Calculate N+1 and divide by 4. Call this 'L'.
4. In your ordered data set, find the number at position L.
5. Calculate N+1, multiply it by 3, then divide it all by 4. Call this 'U'.
6. In your ordered data set, find the number at position U.
7. Take the value in step 4 from the value in step 6.
8. Divide the result of step 7 by 2.

When you take each step at a time, the mathematics are not too difficult at all. It's important to be organised when making these calculations. Make sure your working-out paper does not become muddled, and that you keep track of the values. If you do that then life will be much simpler.

EQUATIONS

Before writing this we asked as many students as we could what they would find useful, and what they would like, when learning about research methods and statistics. 82% of them told us that they would like to avoid equations at all costs, and 87% said that learning how to use equations would be extremely useful. It is clear then that most students hate equations and at the same time most of them think that it's a good idea that they know how to use them. How right they are!

$$\sqrt{\frac{\Sigma (x - \bar{x})^2}{N}}$$

This is the equation for working out standard deviation. The trick with these things is to understand what each part means, and to do one thing at a time. Let's split it up into its parts and then work out how to use it. Armed with your knowledge of the more common symbols used in mathematics, let's take a look at the equation again. Look at the line right in the middle for the moment:

$$\Sigma (x - \bar{x})^2$$

1. Calculate the mean value (x) of your data set.
2. Take your value (x) and subtract from it the mean of the data set ($\bar{x}$).
3. Multiply what you get in step 2 by itself.
4. Do this for all of the values in your data set in turn, writing the result down each time. If you have 20 values then you will end up with 20 numbers.
5. Now add up (Σ) all the numbers you calculated in step 4.

You may get a negative number in step 2, but that doesn't matter at all, because once you've squared it (multiplied it by itself) you always have a positive number, much easier to deal with. Tedious and some might even say very boring, but not terribly complicated so far. The rest is even simpler.

6. Take the total you get in step 5 and divide it by the number of values in your data set (N).
7. Put this number into your calculator, find the square root button marked √ and push it. The result you get is σ, the standard deviation for your data set.

If you do one thing at a time, equations are not very complicated at all. The best thing to do is just practise a little, take your time and be organised. Get to know what each part of the equation means. It's a bit like learning another language, but there's not too much to know. Each equation uses pretty much the same information in a slightly different order. If you know the basics, any equation is child's play. Oh, and if you feel a little daunted by this type of thing, so do 82% of our students: and some of them are at university – you are not alone.

PROBABILITY AND SIGNIFICANCE

We've already let you in on a few secrets about statistics. Here's another. All they are for is for showing that your result didn't occur by chance. That's pretty much it. Don't let anyone fool you into thinking they are any more complicated than that. When we do some statistics all they give us after we've finished adding, multiplying, dividing and things is a single important number, which we refer to as 'p'. It stands for 'probability'.

The p value gives us an idea of how likely it is that our results happened by chance or not. The p value can be anywhere between 0 and 1. A p of 1 means that something is definitely, absolutely going to happen. For instance, there is a probability of 1, an absolute certainty, that in the UK Christmas Day will fall on December the 25th. A p of 0 means that something is never ever going to happen. For instance, there is a probability of 0 that this book will turn into a badger.

The nearer to 1 your value of p is, the more likely it is that the results happened by chance. The smaller the value of p, the more likely it is that we can confidently accept our hypothesis, and reject our null hypothesis. A p of 1 means that we are 100% certain that something will happen by chance. A p of 0.9 means that we are 90% certain that something will happen by chance. A p of 0.8 means we are 80% certain that something will happen by chance and so on. The smaller the p the less likely it is that something will happen by chance. Remember:

Small p – Good

Big p – Bad

If we can reject our null hypothesis we can say that our results were 'significant'. By this we mean that our findings did not occur by chance. Statistics let us measure how 'significant' our results are. Statistical tests are used to analyse the data gathered in research to tell us the p value (we will describe these tests in more detail shortly). The smaller the p value, the more significant the result. For example, the statistical analysis might tell us that we have a p value of 0.02. This is the significance level. When we report 'p' in research papers and books it is more often than not written like this: $p \leq 0.02$.

What this means is that the level of p we have found is less than or equal to 0.02. This means that we are at least 98% certain that our results did not happen by chance. Is 98% enough though? For this we need to check the level of significance selected for our research, and this is given by a number referred to as 'alpha'.

ALPHA (∝)

It is important to decide just how sure we need to be that our results did not occur by chance for us to go ahead and conclude that the hypothesis is to be supported and the null hypothesis rejected. In other words, just how small does p need to be?

The level that p needs to be for us to accept it is called the 'alpha' level. In psychology we say that we need to be at least 95% certain that our results did not happen by chance; 95% certainty translates to a p of 0.05. This means that if our p value is less than or equal to 0.05 then we can reject our null hypothesis and say that our results are significant at the 0.05 level.

You may ask why it is that we choose an alpha of 0.05. The answer to this is that this level is an accepted convention right across the behavioural sciences as it gives us the best chance of avoiding both type 1 and type 2 errors (see below for a description of these). In some cases however we might want to be even stricter. For instance, in situations where people's quality of life may be harmed if we are not very sure that our results did not happen by chance we may choose an alpha of 0.01. That is to say, we would need to be at least 99% certain that our results did not happen by chance. This would allow us to say that our results were significant, not only at the 0.05 level but also at the more demanding 0.01 level we require in

THINGS HAPPEN BY CHANCE!

Things happen by chance all the time. For instance, you might recover from a cold the day after you happened to have a banana for breakfast. Does this mean that the banana caused the cold to go away, or did it just happen by chance? You would need to do some research to see. Let's say you have carried out your research, drawn the graph and worked out the descriptive statistics. The results seem conclusive. It really does look as if bananas are the cure for the common cold. More people who had a banana recovered than people who did not have a banana. However, you cannot yet conclude anything as you are not sure that your result did not happen just by chance. It may look as though a chance result is very unlikely, but you need to check, and we do this by significance testing.

such a case. You might employ an alpha like this, for instance, if you had been investigating how well an experimental drug influenced the memory of elderly people with Alzheimer's disease. In order to put your drug onto the market and allow people to start taking it, you need to be very sure that it does what it is supposed to.

If the p value you find does not fall at or below your chosen alpha, then you are unable to accept your hypothesis. Instead you are unable to reject the null hypothesis and you cannot say that your results were significant. In these circumstances, you have to say that your results were non-significant. This is quite different from *insignificant* by the way! Your findings were not at all insignificant. Even a result that does not support the hypothesis is interesting to scientists, but statistically speaking, it is a non-significant result.

Investigation	Hypothesis and null hypothesis	Alpha	Calculated p value	Correct decision	Incorrect decision
Does eating crisps make you feel sick?	Hypothesis: The more crisps you eat the sicker you feel. Null hypothesis: You do not feel sicker by eating more crisps.	0.05	p≤0.02 (p is less than or equal to 0.02)	The p value is less than alpha. We can reject the null hypothesis and say that the result supports our hypothesis. We can confidently sat that we are 95% certain that our results did not occur by chance. Our results are significant.	A type 2 error: We cannot reject the null hypothesis and must reject the hypothesis. The implications of an error like this may result in people eating loads of crisps because they believe it will not make them feel sick when in fact it does.
Does wearing perfume make females more attractive to males?	Hypothesis: Wearing perfume makes females more attractive to males. Null hypothesis: Wearing perfume does not make females more attractive to males.	0.05	p≤0.09 (p is less than or equal to 0.09)	The p value is larger than alpha. We cannot accept our hypothesis and must retain our null hypothesis. We can only say with 91% confidence that our results did not occur by chance. Our results are non-significant.	A type 1 error: We can reject the null hypothesis and accept the hypothesis. The implications of this error may be serious for the perfume industry. Sales of perfume may drop because women may feel that wearing it makes no difference to their perceived attractiveness.
Should we prescribe a potentially dangerous experimental drug to those with schizophrenia?	Hypothesis: Treatment with the experimental drug improves the quality of life of those suffering with schizophrenia. Null hypothesis: Treatment with the experimental drug does not improve the quality of life of those suffering with schizophrenia.	0.01	p≤0.017 (p is less than or equal to 0.017)	The p value is less than alpha. We can reject the null hypothesis and say that the result supports our hypothesis. We can coincidentally say that we are 99% certain that our results did not occur by chance. Our results are significant at the 0.01 level.	A type 2 error: We cannot reject the null hypothesis and must reject the hypothesis. The implications of this error are potentially very serious. Millions who suffer from schizophrenia will not benefit from this wonderful new drug.

TYPE 1 AND TYPE 2 ERRORS

After you have carried out some statistical tests, you must decide whether you accept or reject the null hypothesis. There are two major errors that can be made when doing this. These are described rather confusingly as type 1 and type 2 errors.

TYPE 1 ERROR

This is also known as the 'false positive' error. It is the mistake of rejecting the null hypothesis when it is actually true. You have made the decision to accept your hypothesis by mistake. A very strict alpha value makes this kind of mistake less likely.

TYPE 2 ERROR

This is also known as the 'false negative' error. This is the mistake of accepting the null hypothesis when it is in fact false, and you should have rejected it in favour of your hypothesis. A very strict alpha value makes this kind of mistake more likely.

ASK AN EXAMINER

Do not underestimate the importance of learning about probability, significance, and type1/type 2 errors! They are a crucial aspect of data analysis and you can bet that they will feature frequently in exam questions!

CHOOSING THE CORRECT STATISTICAL TEST

The choice of your test is influenced by the type of research you have done, and the type of measurements you have made.

LEVELS OF MEASUREMENT

There are three types of measurements you can make in psychological research. These are nominal, ordinal and interval. The best way to describe each is with an example.

Nominal level

If you have nominal data you have data that can be classified in categories. By this, we mean that if something is in one category it cannot be in another category also. For instance, if you are counting up the number of men and women at a rugby match, you cannot have someone who counts as a man and also as a woman. Similarly, if you make a trip to a safari park to carry out a survey on animals you may want to count the number of monkeys you see and the number of hippos. You cannot have a monkey that is also a hippo – they exist as discrete categories. If your data is like this, it is described as nominal level data.

Ordinal level

The clue for this one is in the name. Ordinal suggests that there is an *order*. Horse racing is a good example. Horses are recorded as finishing first, second, third, fourth and so on. The order in which they finish is the important thing, not the distance between them. If your data is like this, in some kind of order or rank, then it is described as ordinal level data.

Interval level

If you are measuring something on a scale, perhaps the height of something or the time it takes someone to do something, then you are using an interval scale. Time, temperature, weight and height are all examples of interval levels of measurement.

A knowledge of your experimental design and the level of measurement used will allow you to answer three simple questions, which will lead you to the appropriate statistical test.

EXAMPLE 1

Our research is concerned with the relationship between a person's level of sadness and the amount of chocolate they eat. Here we are measuring sadness ratings and volume of chocolate consumed in 10 different people.

Question 1: Do I have correlation data?

The answer to this is YES. You are looking for a relationship between variables and so your choice of test is the Spearman's Rho.

EXAMPLE 2

Our research is concerned with investigating whether males are better at science subjects than females. We collect science test scores from 10 males and 10 females and want to compare them.

Question 1: Do I have correlation data?

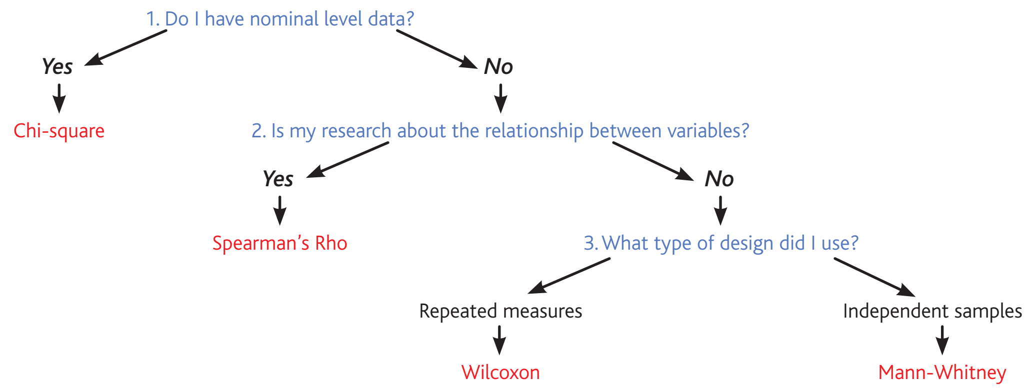
The answer here is NO. You are looking at differences.

Question 2: Am I looking at numbers in categories?

The answer is NO. You are looking at differences.

Question 3: What type of design did I use?

The answer to this is an independent samples design. You have two groups, one male and one female. A comparison is made of the scores of different participants in the different conditions. The test for you is the Mann-Whitney Test.



EXAMPLE 3

Our research is investigating whether memory is better after a high-protein meal than before. Each participant carries out a memory task before a high protein meal of fish and chicken, and again after the meal.

Question 1: Do I have correlation data?

The answer is NO. You are looking at differences.

Question 2: Am I looking at numbers in categories?

The answer is NO. You are looking at differences.

Question 3: What type of design did I use?

The answer to this is a repeated measures design. Each participant provided information before the meal and also after the meal. A comparison is made of the same participant's scores in the two conditions. The test for you is a Wilcoxon test.

EXAMPLE 4

Researchers are investigating whether psychology students and maths students revise differently and whether this influences test scores. The two methods used are cramming information or organised revision. Students can be either psychology or maths students (not both) and they may be organised learners or crammers (not both).

Question 1: Do I have correlation data?

The answer is NO. You are looking at categories of behaviour.

Question 2: Am I looking at numbers in categories?

Yes you are. You are looking at whether test score is influenced by the type of learning and the subject. Each person cannot be in more than one category. They are either psychologists or maths students, and they are either crammers or organised learners. The test for you is a chi-square test.

INFERENTIAL ANALYSIS

DATA ANALYSIS: TEACHERS LOOK AWAY NOW

Let's be absolutely honest about this: most people are not terribly fond of mathematics. In psychology, we spend quite a lot of time using numbers and talking about statistics. Why? Because we need to know whether our findings allow us to conclude anything about our hypotheses. Because statistics are seen as extremely complicated they tend to get more than their fair share of coverage in books and in lessons. This is because people find them hard to understand and so authors and teachers spend a great deal of time explaining them.

It may come as a shock to you, but most psychologists working professionally are also not very fond of statistics. Most of them, however, have realised three very important secrets that we are about to let you in on here. Don't tell anyone though.

1. Statistics are just a tool.

It's as simple as that really. Statistics are the end point of your research. Nearly all of your time as a researcher is taken up deciding what you want to do, how you are going to do it and actually doing it. The statistics come right at the end. They are only a tool for finding out which conclusions you can make from your data. Just as a hammer is a tool for making nails go into walls and an iron is a tool for making clothes flat, statistics are tools to let us weigh up our findings.

The second secret you need to know about statistics is a pretty controversial one. We include it here because we have found it very useful in our learning over the years.

2. Statistics ARE often hard to understand.

There, we've said it. There are two reasons for finding things hard to understand. The first is that you think you are not very smart, and so if something is tricky, then it's your fault. The second, just as valid, reason that something seems hard is that it IS hard. Anyone who tells you that statistics are always simple would be lying to you. They do require some thought. However, the reason most people find them hard is neatly sidestepped when we let you in on secret number 3.

3. You do not *have* to understand it all.
At this point, teachers all over the country are fainting with the realisation that this, the most controversial of their secrets, has been told to their students. Here's an example of what we mean: stick with us here, it's worth it. Sometimes psychologists may use equations when working something out. In these equations they may need to divide by things, multiply numbers or add a lot of numbers up. Here's where secret number 3 comes in.

The reason many students setting out to learn how to use statistics in psychology find it hard is because they are naturally thoughtful and inquisitive people. They are used to asking questions like "WHY am I doing this?" and "WHY do I need to divide by this number?" The problem is that the answers to these questions are often much more complicated that you might expect, and so the student psychologist can become bogged down and sometimes confused. They learn that statistics can be hard and confusing and then everything that involves them is to be hated and avoided.

Now, rewind a little. What might have happened if the student psychologist had not asked the WHY questions? If they had not needed to know why something needed to be divided by something else? In fact, in many cases, most psychologists do not know why certain things are done when applying the tests, they just do them. Here's the most controversial tip of all. Just for the moment, and just where following instructions on statistics are concerned, switch off the bit of your brain that makes you ask 'Why?' Just do it, in the firm knowledge that it works. Remember, statistics are a tool, just like any other. You do not need to understand the physics that go along with the act of hammering a nail into a wall do you? You do not consult textbooks to understand how the thermostat works in an iron when smoothing your clothes do you? Then don't worry about why we divide by something, or why you are told to multiply two numbers together.

An often overused but really good example is cake making. Following a tried and tested recipe one step at a time will result in a good cake. Changing the recipe might improve the cake, but it may well ruin it, so it's safer to stick to the recipe as it always gives successful results. Cooking is basically chemistry with food. When you mix things together and apply heat you change the structure of the ingredients, turning them magically into wonderful cakey goodness. You do not need to be an expert in molecular chemistry to know how to make a cake. You do not need to know how molecules are broken and re-formed during the cooking process and why these molecules are easy to digest. You do not ask 'why' questions in cake making, so there is no need to ask them in statistics.

We have been as careful as we can to set out the statistics you need to know in a form that is easy to follow and reproduce. Where appropriate we'll tell you what's happening and why, but the rule is, if you follow each step carefully you will be provided with the information you need in your research.

USING INFERENTIAL STATISTICS

Descriptive statistics allow us to *describe* the data we are using. They include presenting data in the form of graphs and tables or summarising them as measures of central tendency or dispersion. In this section we will describe how to carry out four different statistical tests that allow us to *infer* something about our data. By that we mean that they allow us to deduce or conclude something about the research we have carried out, not just describe the data. Inferential statistics are often used right at the end of the research process to allow researchers to conclude things about whether their results have occurred by chance or not. Remember, the tests are only tools to allow you to arrive at a p value for your research. The various tests may seem a little confusing at first, but stick with it, and they will become much easier to handle.

ASK AN EXAMINER

You are required to know about the use of four inferential tests: Spearman's Rho, Wilcoxon, Mann-Whitney and chi-square. Don't learn every step of each test – instead, develop an appreciation of the logic of inferential analysis. Understand why you would use a particular test, how the score from each test leads to a p value and how this is interpreted for significance. A good understanding of these things will serve you well in the exam.

SPEARMAN'S RHO

This test is used in correlation research. If you are looking at the relationship between two variables and have drawn a scattergraph, then Spearman's Rho is the test to use. We'll start by summarising some research involving correlation and using the real data in our calculation.

Aim:

To investigate the relationship between ability at mathematics and ability at playing chess.

Hypothesis:

The better you are at maths, the better you will be at chess.

Null hypothesis:

Being good at maths does not mean that you will be good at chess.

Operationalising the variables:

Maths ability was operationalised as 'score out of 100 on a maths test'.

Chess ability was operationalised as 'number of chess problems out of 100 successfully solved'.

The raw data:

	Maths Score	Chess Score
1	58	67
2	34	43
3	77	72
4	59	61
5	46	50
6	90	99
7	22	20
8	78	73
9	89	90
10	67	70

The graph:

This is correlational research investigating the relationship between one variable and another. The appropriate graph to draw is a *scattergram*.

The descriptive statistics:

	Maths Score	Chess Score
Mean	62	64.5
Standard deviation	22.81	22.78

It seems from these data that there is indeed a strong positive correlation between maths ability and chess ability. This is shown most clearly in the graph. As maths score increases so too does chess score, typical of a positive correlation. To be more certain, and to see whether the correlation is significant, we need to carry out the *Spearman's Rho statistical test*.

Doing Spearman's Rho

Step 1: Rank the scores for each variable

Ranking will become extremely straightforward as you have to do it for three of the tests you are required to understand. It is simple enough, but take care to get it right. Take your original table and add two more columns. Ranking means putting the scores in order, smallest to largest, and giving the smallest the rank of 1, the next smallest the rank of 2 and so on. You can see how this is done by looking at the ranks for maths scores in the table below. Notice, the smallest maths score was 22. This gets the rank of 1. The next smallest was 34. This gets the rank of 2 and so on, right up to the highest score of 90, which gets the rank of 10.

	Maths Score	Chess Score	Maths Rank	Chess Rank
1	58	67	4	5
2	34	43	2	2
3	77	72	7	7
4	59	61	5	4
5	46	50	3	3
6	90	99	10	10
7	22	20	1	1
8	78	73	8	8
9	89	90	9	9
10	67	70	6	6

The ranking of our data is uncomplicated. In other circumstances, though, you may find that two participants score the same on the same task. This common problem makes the ranking slightly more difficult. Take a look at the section on tied ranks below for more on this.

Extract from A2 Psychology: The Student's Textbook © Nigel Holt and Rob Lewis ISBN: 9781845901004 www.crownhouse.co.uk

Step 2: Work out differences and square them

In our case we need to work out the difference between maths score and chess score for each of our participants. To start you off:

Participant number 1 was ranked at 4 for maths and 5 for chess.

$4 - 5 = -1$
$-1^2 = 1$

For each participant, the 'difference' number, and then the figure which is the result of squaring that number, are recorded in two new columns, to the right of those shown in the table above. When you've done this for all your participants you should end up with a table like this:

	Maths Score	Chess Score	Maths Rank	Chess Rank	d (difference)	d^2 (d squared)
1	58	67	4	5	-1	1
2	34	43	2	2	0	0
3	77	72	7	7	0	0
4	59	61	5	4	1	1
5	46	50	3	3	0	0
6	90	99	10	10	0	0
7	22	20	1	1	0	0
8	78	73	8	8	0	0
9	89	90	9	9	0	0
10	67	70	6	6	0	0

Step 3: Add up the numbers in the d^2 column ($\sum d^2$ – 'sigma d^2' in English)

You need this number for working out the equation. In our case all we do here is add up the numbers in the far right column:

$$1+0+0+1+0+0+0+0+0+0 = 2$$

Step 4: Identify N and N^2

N refers to the number of pairs you have in your data. In this case we have 10. N^2 is simply that number squared: in your case that's (10 x 10) which is 100.

Step 5: Get organised

Get all the numbers you need for your equation.

$$r_s = 1 - \frac{6(\sum d^2)}{N(N^2-1)}$$

r_s – this is the statistic we are after – Spearman's Rho.

N – the number of people who provided data in your correlation research.

d – The difference between each rank of one variable and the corresponding rank of the other variable.

Your equation requires $\sum d^2$, N^2 and N. Make sure you know what these are and write them down on the top of the paper on which you will do your calculations.

$$\sum d^2 = 2$$
$$N = 10$$
$$N^2 = 100$$

Step 6: Insert your numbers into the equation

With your numbers it looks like this:

$$r_s = 1 - \frac{6(2)}{10(100-1)}$$

This looks a little less frightening and gets even simpler. The first rule is to work out everything within the brackets first. When you do that you get the following:

$$r_s = 1 - \frac{6(2)}{10(99)}$$

You'll know this from maths, but it won't hurt to remind you that you need to multiply the number inside the brackets by the number on the outside: i.e. you need to calculate 6 x 2 and 10 x 99. When you do that, your equation looks like this:

$$r_s = 1 - \frac{12}{990}$$

Next, work out the fraction. Remember, you divide the top number by the bottom number, (in other words, 12÷990) to give you the statistic we need which will indicate the strength of your correlation.

$$r_s = 1 - 0.012$$
$$= 0.988$$

This is your *correlation coefficient*. You will recall that anything positive means that as one variable increases so too does the other (the absence of a sign in front of the number means that it is positive). You will also know that the nearer to 1 your coefficient is, the stronger the correlation. It is clear that we have an extremely strong correlation here, but is it significant? Might it have occurred by chance?

Step 7: Finding the critical value

To see whether the calculated r_s is large enough for you to say that your results did not occur by chance you need to look in statistical tables. At the end of this book you will find an appendix with the table you need for this test. To use it, you need to know:

N The number of participants in your task: 10

r_s The value of Spearman's Rho that you calculated: 0.988

The type of hypothesis
Whether you chose a directional (one-tailed) or non-directional (two-tailed) hypothesis. In this case our hypothesis was directional, as we predicted that better maths scores would go along with better chess scores.

Your level of alpha (∝)
The level of significance (alpha) you are using. Remember, this is usually 0.05, but may in other cases be much lower.

What you need to do now is compare the calculated number (sometimes called the *observed* value) to a number in an appropriate statistical table (sometimes referred to as the *critical* value). You look up the relevant number in the table and see if your observed value is *equal to or larger than the critical value from the table*.

The full table can be found in the appendix, but for simplicity, the section from the table that refers to your research is as follows:

Levels of significance for a one-tailed test

N	0.05	0.025	0.01	0.005
10	0.564	0.648	0.745	0.794

What we can see is that our calculated level of 0.988 is much larger than the level required for a significance (p) level of 0.05, or any other significance level for that matter! We can clearly be 95% certain that our results did not occur by chance. We can definitely reject our null hypothesis and say that our results provided clear support for the hypothesis that there is a strong relationship between maths and chess ability.

ASK AN EXAMINER

Remember, just follow the logic of these tests, don't try to remember how to do them. You start research with a hypothesis and the inferential analysis allows you to see whether or not it is supported by the data you have gathered in your research. If the analysis says that your findings might be due to chance then you have supported the null instead of your hypothesis!

DEALING WITH TIED RANKS
Take a look at the following data. The ranking has already been done for you.

	Spelling Score	Memory Score	Spelling Rank	Memory Rank
1	3	7	3.5	9
2	0	6	1	8
3	8	3	9	4
4	7	**4**	8	**6**
5	6	**4**	7	**6**
6	9	0	10	1
7	5	8	6	10
8	3	1	**3.5**	2
9	4	2	5	3
10	1	**4**	2	**6**

Look at the first column where spelling scores are indicated. Participants 1 and 8 both scored 3 on this test. Now look at the 'Spelling Rank' column. When ranking a column of scores like this, each tied score gets an average of the rank positions that they would normally be in. In this case, the scores would be in rank positions 3 and 4. The mean rank value for these two positions is $(3+4)/2 = 3.5$.

Look now at the 'Memory Score' column. Notice that three participants scored 4 on this test. These scores would occupy rank positions 5, 6 and 7 and so each receives the mean rank of $(5+6+7)/3 = 6$.

The rest of the test progresses in exactly the same way as usual. The only slightly tricky thing here is making sure that you get the ranking right.

THE WILCOXON TEST
The Wilcoxon is used when you are looking for differences, rather than a relationship, with a repeated measures design. This means that the same participant provides data for the different conditions in your research. We can use a simple example of comparing memory in silence and memory with music. In the experiment we could see how people perform on a memory test in silence and how they perform on the memory test while listening to music. The test can be as simple as remembering a list of numbers.

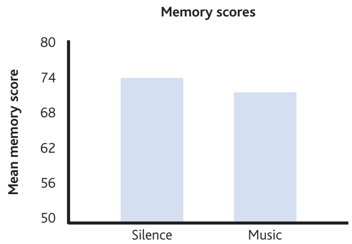
Aim:
To investigate whether memory is influenced by music.

Hypothesis:
Memory for numbers will be worse while listening to music than in silence.

Null hypothesis:
Memory for numbers will NOT be worse while listening to music than when in silence

Extract from A2 Psychology: The Student's Textbook © Nigel Holt and Rob Lewis ISBN: 9781845901004 www.crownhouse.co.uk

Here are some results.



	Silence	Music
Mean	72.55	64.55
Standard deviation	15.15	16.36

It seems that silence does indeed provide better conditions for memory than music. Unfortunately we cannot yet reject our null hypothesis and accept our hypothesis. We cannot be sure that our results did not occur by chance and so we need to test the significance of our findings.

DOING THE WILCOXON TEST

Step 1: Take the scores on one condition from the scores on the other condition

In this case, we simply subtract the memory score while listening to music from the memory score in silence. In this test we must be careful to mark down whether the difference is positive (+) or negative (–).

Step 2: Rank the differences

Here we rank the differences, not the scores themselves. There are two rules to remember for ranking in this test.

1. Ignore all the ties when ranking
2. Ignore the sign when ranking

Once you've done this you should end up with a table that looks like this:

	Silence	Music	d (difference)	Ranks of d	Sign of the rank	Negative ranks	Positive ranks
1	76	80	-4	4	–	4	
2	76	67	+9	10.5	+		10.5
3	55	54	+1	1	+		1
4	89	78	+11	14	+		14
5	76	65	+9	10.5	+		10.5
6	87	81	+6	7	+		7
7	67	64	+3	2	+		2
8	34	27	+7	8.5	+		8.5
9	67	45	+22	19	+		19
10	65	51	+14	17.5	+		17.5
11	59	55	+4	4	+		4
12	91	87	+4	4	+		4
13	99	99	0	(tie)	..		
14	56	51	+5	6	+		6
15	68	55	+13	15.5	+		15.5
16	89	76	+13	15.5	+		15.5
17	78	64	+14	17.5	+		17.5
18	74	67	+7	8.5	+		8.5
19	65	55	+10	12.5	+		12.5
20	80	70	+10	12.5	+		12.5

Notice that in our data there is one tied score, and only one negative difference, ranked at position number 4.

Step 3: Add up the negative and positive ranks, and choose the smaller of the numbers

In our data, the total sum of the negative ranks is 4. The total sum of the positive ranks is a lot more, and comes to 186. The number we want is the smaller of these, 4. This is your Wilcoxon statistic, referred to as *T*. It is your observed value.

Step 4: Finding the critical value

You must now compare the observed value to the critical value in the relevant table. It is included in an appendix at the back of this book. To do this, you need to know:

N The number of participants in your task: 20
T The observed value you calculated: in this case it is 4.

The type of hypothesis

Whether you chose a directional (one- tailed) or non-directional (two-tailed) hypothesis.

Your level of alpha (∝)

The level of significance (alpha) you are using. Remember, this is usually 0.05, but may in other cases be much lower.

The first thing to do is look down the first column for your value of N. Next, read across to the relevant column that refers to your level of alpha. Here you will find the number you are looking for. For simplicity, only the relevant line in the table is reproduced here.

Levels of significance for a one-tailed test

N	0.05	0.025	0.01	0.005
20	60	52	43	26

If *your* value is *smaller than the value in the table* you can say with confidence that your results did not occur by chance. Here, the critical value from the table is 60, for the significance level of 0.05. Your value (4) is definitely smaller than this.

This means that you can say that your results are significant. We report this as follows: 'The difference between memory in silence and when listening to music is significant ($p<0.05$)'.

THE MANN-WHITNEY TEST

If your research investigates the difference in performance on a task by two different groups then this is the test to use. In other words, if you have used an independent samples design, and you are looking for differences, then use the Mann-Whitney. Here's an example of where this test is appropriate.

Aim:

To investigate whether women are more able at doing two tasks at once than men.

Hypothesis:

The ability to do more than one task at once is different in women and men.

Null hypothesis:

The ability to do more than one task at once is *not* different in women and men.

Operationalising the variables:

Two tasks were operationalised by doing sums while copying a rhythm by tapping the feet.

The raw data:

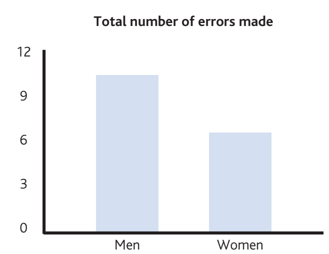
Men (N = 8)	Women (N = 10)
12	9
4	5
17	8
11	4
8	10
12	9
10	4
9	5
	3
	12

Notice, we have eight men and ten women. That's not a problem for the Mann-Whitney test as it conveniently handles both equal and unequal group sizes.

We refer to each different group as group A and B, so the number in group A is N_A (8) and the number in group B is $N_{B.}$ (10)

The graph:

We have two different groups. The appropriate graph to draw here is a simple bar chart.



The descriptive statistics:

	Men	Women
Mean errors	10.38	6.9
Standard deviation	3.74	3.07

From the graph and the descriptive statistics it looks very much as if women make fewer errors when doing more than one task at once than do men. We'll need to do some statistics, in this case a Mann-Whitney, to assess the likelihood that these results have occurred by chance.

Doing the Mann-Whitney Test

Step 1: Rank the data

In the Mann-Whitney, we rank all the data together. By this we mean that the men's scores and the women's scores are all ranked at the same time. We've done this for you here. Notice the usual problem of tied ranks. If you are not sure about this, look back to the section above.

Men (N = 8)	Rank	Women (N = 10)	Rank
12	16	9	10
4	3	5	5.5
17	18	8	7.5
11	14	4	3
8	7.5	10	12.5
12	16	9	10
10	12.5	4	3
9	10	5	5.5
		3	1
		12	16

Step 2: Add up the ranks in each group, calling them Rank A (R_A) and Rank B (R_B)

Sum of ranks for men (R_A)

16+3+18+14+7.5+16+12.5+10 = 97

Sum of ranks for women (R_B)

10+5.5+7.5+3+12.5+10+3+5.5+1+16 = 74

Step 3: Use your values to calculate the results of two formulae

For the Mann-Whitney, we need to calculate two values, referred to as U_A and U_B. Here are the formulae. They are very similar indeed, so be careful you are using the right numbers when you calculate them!

1.

$$U_A = N_A N_B + \frac{N_A(N_A+1)}{2} - R_A$$

2.

$$U_B = N_A N_B + \frac{N_B(N_B+1)}{2} - R_B$$

These look a little frightening at first, but let's take one at a time and see what they look like when we insert our numbers. As usual, begin by identifying the numbers you need and organising yourself carefully.

U_A – the number we are calculating with the formula

U_B – the number we are calculating with the formula

N_A = the number in Group A (men) = 8

N_B = the number in Group B (women) = 10

R_A = the sum of ranks in Group A (men) = 97

R_B = the sum of ranks in Group B (women) = 74

1.

$$U_A = N_A N_B + \frac{N_A(N_A+1)}{2} - R_A$$

Inserting our values, the formula now looks like this:

$$U_A = 8\text{x}10 + \frac{8\ (8+1)}{2} - 97$$

A little basic mathematics can simplify it further to:

$$U_A = 80 + \frac{8\ (9)}{2} - 97$$

This is equal to:

$$U_A = 80 + \frac{72}{2} - 97$$

which equals:

$$U_A = 80 + 36 - 97$$

so finally

$$U_A = 19$$

2.

$$U_B = N_A N_B + \frac{N_B(N_B+1)}{2} - R_B$$

Inserting our values, the formula now looks like this:

$$U_B = 8\text{X}10 + \frac{10\ (10+1)}{2} - 74$$

Again, some straightforward mathematics can simplify it further to:

$$U_B = 80 + \frac{10\ (11)}{2} - 74$$

This is equal to:

$$U_B = 80 + \frac{110}{2} - 74$$

which equals:

$$U_B = 80 + 55 - 74$$

so finally

$$U_B = 61$$

Step 4: Select the smallest value of U and consult the tables for a critical value

In our case, the smallest value of U calculated was U_A: this is our observed value at 19. We now need to consult the table of critical values for the Mann-Whitney. To do this we need to know the number of participants in each group. We had eight males and ten females, so we read across the top of the table until we reach 8, and we then read down to the N=10 line.

A portion of the table is reproduced for you here, but make sure you refer to the real table in the appendix of this book to be sure you know how to read it.

		N2				
		7	**8**	**9**	**10**	**11**
N1	**7**	4	6	7	9	10
	8	6	7	9	11	13
	9	7	9	11	13	16
	10	9	11	13	16	18
	11	10	13	16	18	21

You can see that the critical value of U for N of 8 and N of 10 is 11. If your value of U is equal to or less than the critical value then you can say confidently, with 95% certainty, that your results did not happen by chance.

Our value of U was 19. This is not smaller than 11, so we cannot reject our null hypothesis. Even though the graph and the descriptive statistics suggested that it might be the case, we cannot say with confidence that women are indeed better at doing more than one task at once. This is a very good example of why we should always do the statistics! If we had not done so, we might have made the error of rejecting our null hypothesis by mistake, and accepting our hypothesis even though the difference could well have happened by chance.

CHI-SQUARE (χ^2)

Chi-square test (pronounced 'ky square') is often written as χ^2. We use this test when we have numbers of people in different categories. Here's an example from earlier, to help clarify what we mean:

'Researchers are investigating whether psychology students and maths students revise differently and whether this influences test scores. The two methods used are cramming information or organised revision. Students can be either psychology or maths students (not both) and they may be organised learners or crammers (not both).'

In this case we have a group of psychology students and a group of maths students. The rule is that we cannot have a student studying both. Each of our subject groups is split into two according to the revision style of its members: those that cram for exams (crammers) and those that revise in an organised way (organised).

To summarise, students can be in one of four completely separate categories. These are as follows.

1. Psychology: Crammers
2. Psychology: Organised
3. Maths: Crammers
4. Maths: Organised

What's important here is that no one can be in more than one group. The Chi-square test looks at whether the number of students in each category is different from what we might expect just by chance. In this way the test tells us whether the null hypothesis (that something happened by chance) can be rejected or not, thus allowing us to make a decision about our hypothesis.

Aim:

We are interested in whether the type of revision (cramming or organised) works differently for different subjects.

Hypothesis:

There will be a difference in the effectiveness of different types of revision depending on the subject.

Null hypothesis:

There will *not* be a difference in the effectiveness of different types of revision depending on the subject.

Operationalising the variables:

A self-report questionnaire of psychology and maths students asks them to identify the type of revising they do, cramming or organised revision.

The raw data:

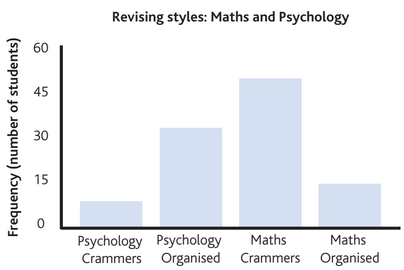
	Psychology	Maths
Crammers	5	50
Organised	35	10

Each square relating to each of the four categories is referred to as a 'cell'. Each 'cell' contains a number which relates to the number of students in that category and is referred to as the *observed frequency*.

The graph:

We have two different groups. The appropriate graph to draw here is a simple bar chart.

The question is, are the numbers of students in these categories different from what we might expect to see just by chance The way to find out is to carry out a Chi-square test.



DOING THE CHI-SQUARE TEST

Step 1: Add up the values in the rows and the columns of the results table, and calculate an overall grand total

The easiest way to do this is to reproduce the results table, adding the totals of the rows and columns and the grand total, like this:

	Psychology	Maths	Total
Crammers	5	50	55 (5+50)
Organised	35	10	45 (35+10)
Total	40 (5+35)	60 (50+10)	Grand total 100 (5+50+35+10)

Step 2: Work out what might be expected just by chance

To work out the number in each cell that might occur by chance is easy, and the result is referred to as the expected frequency. We use the following very basic formula:

$$E = \frac{RC}{T}$$

The expected frequency for the cell (E) equals the row total (R) times the column total (C) divided by the grand total (T).

The four sums are easy to calculate. Make sure you work out the row total times the column total before dividing by the grand total though, and make sure you use the correct row totals and column totals in each calculation.

Observed frequency = 5
Expected frequency = (5x40)/100
= 200/100 = **2**

Observed frequency = 50
Expected frequency = (55x40)/100
= 2200/100 = **22**

Observed frequency = 35
Expected frequency = (45x40)/100
= 1800/100 = **18**

Observed frequency = 10
Expected frequency = (60x45)/100
= 2700/100 = **27**

To summarise, the expected frequencies for our data look like this:

	Psychology	Maths
Crammers	2	22
Organised	18	27

Step 3: Use your observed and expected frequencies to calculate chi-square

Now that we have observed and expected frequencies for each cell we can calculate chi-square using this equation:

$$x^2 = \sum \frac{(0-E)^2}{E}$$

To calculate chi-square (χ^2) we have to calculate $(O-E)^2/E$ for each cell, then add up (Σ) the results of the sums.

We've done this for you here. Each cell contains a single calculation, using the observed and expected frequencies for that cell.

Extract from A2 Psychology: The Student's Textbook © Nigel Holt and Rob Lewis ISBN: 9781845901004 www.crownhouse.co.uk

	Psychology	Maths
Crammers	Observed frequency = 5 Expected frequency = 2 $(5–2)^2/2$ $=3^2/2$ $=9/2$ = **4.5**	Observed frequency = 50 Expected frequency = 22 $(50–22)^2/2$ $=28^2/22$ $=784/22$ = **35.54** (2 decimal places)
Organised	Observed frequency = 35 Expected frequency = 18 $(35–18)^2/18$ $=17^2/18$ $=289/18$ = **16.06** (2 decimal places)	Observed frequency = 10 Expected frequency = 27 $(10–27)^2/27$ $=-17^2/27$ $=289/27$ = **10.70** (2 decimal places)

To complete the calculation of chi-square, add up the results of the individual calculations:

chi-square = 4.5+35.54+16.06+10.70
= 66.8

Step 4: Using the table, find the relevant critical value of Chi-square

To find the correct number you must know whether you chose a directional (one-tailed) or non-directional (two-tailed) hypothesis. In our study we simply stated that there would be a difference between the revising styles: we did not say whether one would be larger than another, so we have a non-directional hypothesis.

Before you can use the table you must calculate a value called 'degrees of freedom', shortened to 'df'. This is done using the following simple formula:

df = (number of rows – 1)
x (number of columns – 1)

For our data it could not be simpler

df = (2–1)x(2–1)
= 1x1
= 1

Finally, consult the relevant table. It's printed at the end of this book for you, but for convenience, the key portion of it is reproduced here. As usual, make sure you consult the table itself, so that you know how it works.

Levels of significance for a two-tailed test

	0.20	0.10	0.05	0.02	0.01
1	1.64	2.71	3.84	5.41	6.64
2	3.22	4.60	5.99	7.82	9.21
3	4.64	6.25	7.82	9.84	11.34

The value for chi-square that you calculated must be the same as, or larger than, the critical value you read from the table. Your calculated value was 66.8. The critical value for the significance level of 0.05 is 3.84. It seems that you can very confidently reject the null hypothesis and say that your data did not occur by chance. Your result is significant ($p<0.05$).

ANALYSING AND INTERPRETING QUALITATIVE DATA

The purpose of research is to provide data of some kind that can be analysed to help support or reject a hypothesis. The data can be in many different forms, but it will either be quantitative or qualitative. The distinction is quite easy to remember. Quantitative data is to do with quantities (numbers, amounts). Qualitative data is to do with quality (opinions, feelings). Which kind you use will depend on the type of research being conducted. In some cases you might even collect both quantitative and qualitative data. For instance, you might be interested in measuring stress levels in relation to maths problems. The result of a maths test would be quantitative; asking people how they feel about the maths test would be qualitative.

How then, do you present qualitative data? It's actually more straightforward than you may think. Consider the following example. A researcher is carrying out a kind of analysis called 'discourse' analysis on interviews with politicians which involves investigating samples of speech. He is interested in seeing how often politicians comment on matters that are local to the United Kingdom, or matters relating to international issues. The speeches of 10 politicians are identified and investigated. The comments recorded by the researcher include:

'The US elections have huge implications for the wars in the Middle East'

'The Scottish parliament can work hand in hand with the parliament in Westminster'

'The floods in the West Country were devastating. We must do all we can to stop this happening again'

'The crisis in central Africa is worsening. I fear for the lives of these people'

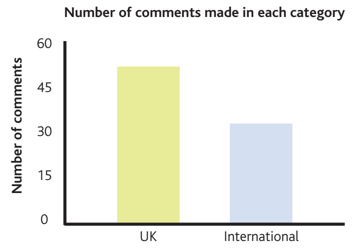
'Rebel fighters in parts of the Russian Federation have grown in strength over the last two years'

'House prices in the UK are, frankly, ridiculous'

While a list of comments like this is interesting, it is not very well organised. We can use a table to help present some of the data in such a way that the reader does not have to sort through a list to work out what has been said about the UK and what has been said about issues outside the UK. There may be many more comments, and these may be included in another table, but a good summary table with samples of the comments will be very useful to readers interested in seeing the kind of comments that have been made. The table may look like this:

Category	Examples
Comments directly relating to UK issues	*'The Scottish parliament can work hand in hand with the parliament in Westminster'* *'House prices in the UK are, frankly, ridiculous'* *'The floods in the West Country were devastating. We must do all we can to stop this happening again'*
Comments directly relating to international issues	*'The US elections have huge implications for the wars in the Middle East'* *'The crisis in central Africa is worsening. I fear for the lives of these people'* *'Rebel fighters in parts of the Russian Federation have grown in strength over the last two years'*

The table makes it much easier and faster to find the relevant information. Tables like this might also include the total number of comments in each category. If we have a total number we can draw a graph, like this:



It is immediately clear from this graph that the speeches of the politicians used in this research had many more comments that directly related to international matters than ones that dealt with UK issues. This is a perfect example of what a graph is for. Graphs allow readers to access a good summary of the data in an immediate, visual way.

CONVENTIONS OF REPORTING ON PSYCHOLOGICAL INVESTIGATIONS

Once we have completed our research we need to write it up before sending it for peer review and (we hope) publication. There are certain conventions that should be followed when reporting psychological research. These are conventions rather than 'rules' because some scientific journals require a slightly different style, but generally speaking, psychological research should be presented in a certain way. Imagine that we wanted to present our research into whether memory is worse with music playing than in silence. Our report should be organised into the sections we identify here.

1. TITLE

It seems obvious to say it, but a strong and clear title is extremely important. Thousands of pieces of research are published every year. Researchers pour through the titles of this new research to see if anything grabs their interest and is relevant to their own work. A title that clearly expresses what the research is investigating is extremely helpful. In our study of whether memory is worse in music or in silence we may choose the following title:

'Memory: An investigation into why music really makes it rubbish'

However, this is not very informative, or formal. A better title might be:

'An investigation into how memory for numbers is influenced by music'

This is more like it. From this title the reader is able to see that the research is into memory: it investigates memory for numbers, and music is used to disrupt it.

An even better title might be:

'The irrelevant sound effect: The disruption of memory for serially presented digits in conditions of music and silence'

The phenomenon we are investigating here is actually called 'the irrelevant sound effect'. It has been widely studied and refers particularly to memory in silence compared to memory with sound presented at the same time. This title clearly indicates to the reader what type of research is in the report.

In this section of the book we have the permission of the authors to use sections of a real paper which appeared in a publication called the *Journal of Applied Psychology* in 2007. The authors were Beaman and Holt, and the copyright for the paper belongs to the publishers, John Wiley and Sons. The title was:

'Reverberant auditory environments: The effects of multiple echoes on distraction by "irrelevant" speech'

In their research, Beaman and Holt looked at how sounds with an echo (reverberation) influenced memory differently than places without an echo.

2. ABSTRACT

The abstract is a short summary of the research paper placed at the very start. A good abstract should provide a little background, some details of the method, a statement of results and a short conclusion. All this in fewer than 150 words. The abstract comes first, straight after the title, but you should always write it last. The reason is that it includes a very brief summary of the information you have included in every section of your report and so it's much easier to write when you have finished the rest of the research report!

Here's an abstract that, unfortunately, is actually better than some we have seen over the years! We really hope you agree that it is not terribly good.

> We did some research that was looking at whether people were more rubbish at remembering stuff when they had music coming through headphones. We gave them things to remember and they remembered them. We worked out some maths and decided whether their remembering was better when they couldn't hear anything. It was. There were about 10 people we tested. At the end of it we decided that it was useful to know this, for instance, that's why we think it should be quiet in libraries mostly. Right at the end we thought about some other research and we think we will now investigate whether listening to old-fashioned music like your dad listens to will make memory worse than listening to stuff you have on your own iPod.

There is very little detail in the abstract. The background to the research is not clear at all. We do not know what the task itself was, what the hypothesis was, what the variables were or how they were operationalised. We do not have a clear understanding of the statistical tests carried out or what exactly was found, and the suggestions at the end for further research are vague to say the least! Not very good at all. Now look at the version below.

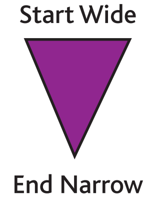
> The research detailed here is an investigation into whether memory for digits presented in series is influenced more by the presentation of music during the task than when participants proceeded in silence. Previous research has suggested that any kind of sound interferes with recall of numbers presented in series, suggesting the hypothesis that memory would be worse when listening to music. An opportunity sample of 10 participants took part in an independent samples design and provided data that supported the hypothesis ($p \leq 0.05$). How the sound influences memory is discussed, with reference to the working memory model. Concluding comments included possibilities for further research.

This abstract is much better. Here we have some background, some detail of our design, the result and even something of our discussion and conclusion – all in fewer that 150 words. Finally, here is the real abstract from the Beaman and Holt research. It's quite complicated, but it does give you an idea of the kind of information and detail you find in professionally produced work.

> Two experiments examine the effect on an immediate recall test of simulating a reverberant auditory environment in which auditory distracters in the form of speech are played to the participants (the 'irrelevant sound effect'). An echo-intensive environment simulated by the addition of reverberation to the speech reduced the extent of 'changes in state' in the irrelevant speech stream by smoothing the profile of the waveform. In both experiments, the reverberant auditory environment produced significantly smaller irrelevant sound distraction effects than an echo-free environment. Results are interpreted in terms of changing-state hypothesis, which states that acoustic content of irrelevant sound, rather than phonology or semantics, determines the extent of the irrelevant sound effect (ISE).

3. THE INTRODUCTION

The introduction is a really important part of the write up. It's here that you get the chance to describe the background to your research and how you really came up with your idea in the first place. This is where you place your work alongside other research in the same field. All introductory sections differ slightly in length, depending on what it is that we need to cover and how much background research there has been in the area. Introductions, however long, should nevertheless follow the golden rule.



Think of your introduction as a funneling system. You should start with all of your background information and details of research that has already been done. You really need to provide a good general overview in the introduction. As you proceed, you should begin to focus your ideas on your research. Towards the end of the introduction, talk about the aims of your research, ending finally with your hypothesis. Take a look at some excerpts from Beaman and Holt (2007).

> The use of virtual environments (VE) in psychology is now well-established. Studies of spatial navigation (Ruddle, Payne, & Jones, 1998), spatial memory (Ruddle, Payne, & Jones, 1999)… and neuropsychological rehabilitation (Wann, Rushton, Smyth, & Jones, 1997) have all benefited from these VE techniques… … However, virtual auditory environments are much less common than virtual visual environments. When virtual auditory environments have been created, however, they have proved useful to the study of the 'sense of place' associated with particular buildings and historic sites… The purpose of the current study is to apply similar techniques to the study of auditory distraction by so-called 'irrelevant sound' known to disrupt working memory.

The introduction begins very generally, and quickly proceeds to a statement of the broad aims of the research. The authors have really told the reader what to expect, the background and the general ideas behind their study. Later in the introduction Beaman and Holt begin to be more specific.

> In a reverberant environment the direct signal is overlayed *(sic)* with multiple reflections, the intensity of each reflection and the delay imposed upon it being a function of the intensity of the original signal and the nature (e.g. the size, surfaces and contents) of the room… It follows that larger spaces with more reverberation may intrinsically be less distracting environments (in acoustic terms) than small rooms with low levels of reverberation. This prediction follows from the changing-state hypothesis of Jones and colleagues…we propose to take a different approach and ask whether extreme values of reverberation, which we expect to represent a powerful manipulation, will affect the results obtained for a much smaller sample size. The advantage of this approach is that it makes positive, rather than null, predictions and provides a direct test of the changing-state hypothesis.

Here, the introduction is really setting the scene, explaining that certain hypotheses and theories described in other research will be investigated. The authors finish by explaining that predictions will be made to test a hypothesis. We can see that the introduction covers a great deal of ground. We've only included a few excerpts here: much more is described in the original article. We can see from the sections we have reproduced that the introduction begins very generally, then focuses, using the 'start wide, end narrow' rule, funneling the information down from a general discussion to a much more focused one.

4. THE METHOD

When writing a method section, you need to ask yourself one simple question: 'Would someone else be able to do exactly what I did from reading my method?'

If the answer to this is 'No', then your method is not good enough. A good method has just the right amount of information for someone who has never watched you work before, and who may never have carried out research in the area, to carry out a 'replication' of your study. Replication is a very important aspect of scientific research. If a study can be replicated then we can be more certain that the findings are to be trusted. Replication adds to the reliability of the research, with unreplicable research being regarded as unreliable. A good method provides other researchers with the information they need to repeat your work closely enough that they can hope to replicate your findings. A method might include a number of sections. Not all methods include all of these sections, but we'll include them here for reference.

Design: A statement of the design used may be included. Here you might indicate the different conditions in your research and if appropriate you might say whether a repeated measures or independent samples design was used, or even include the independent and dependent varieties IV and DV.

> Four conditions were presented, each with 15 trials, giving a total number of 60 experimental trials per participant. Presentation order for each candidate was chosen randomly from the 24 possible condition orders… The four conditions were … three levels of reverberation … and a 'no sound' control condition.

Participants: In this section you would give details of your sample, including how many men and women there were and the age range. You should also indicate how the sample was identified, perhaps using an advertisement for a volunteer sample, or the phone directory for a random sample. It may also be appropriate to indicate where the research was conducted and from what country the participants came. You must be very careful to make sure that the participants remain anonymous as this is their right, and it is also one of the principles of ethical research. This is from Beaman and Holt (2007):

> Eleven students of … University volunteered to participate in this study. All participants reported normal hearing levels and normal or corrected-to-normal vision. All had English as their first language.

Apparatus and stimuli: In this section researchers would indicate the apparatus they used in their research. If they used certain sounds, or visual stimuli, then they would give details of these, perhaps indicating how each was produced.

> Sounds were digitally recorded on MiniDisc and presented from a Sony MiniDisc player at a comfortable listening level to students through Sennheiser HD570 headphones. Listening levels were set to a comfortable level with a test recording before the experiment began and kept at the same level for the duration of the experiment. Stimuli were presented on a Pentium Class PC running Windows XP and MS PowerPoint software. …using Cool Edit sound editing suite (Syntrillium Software Corporation).

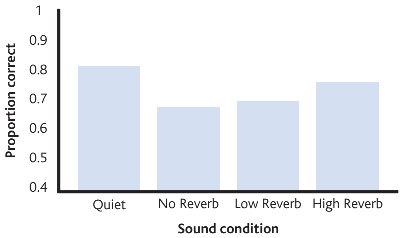
Procedure: This section of the report is particularly important for those wishing to replicate the research. It is here that researchers show exactly what they did.

> Each participant was tested individually in a quiet room. The irrelevant sound was presented continuously … throughout the presentation … directly from the digital recording. Participants were instructed to ignore anything they might hear and were reassured they would not be tested on it in any way. In each trial … the seven numbers were presented in succession on a VDU. After the presentation the instruction 'recall now' was presented. Participants recorded the numbers manually on response sheets provided, writing from left to right. They were asked not to omit any responses and, if they could not remember a particular item, to write down their best guess and move onto the next item. They were asked not to go back and amend previous answers if they felt they had made an error. The visual instruction 'Push space for next trial' cued participants to progress through the experiment.

5. THE RESULTS

After researchers have identified their method, the results are presented. The presentation of results may be in the form of descriptive statistics, tables, graphs and also inferential statistics. In the case of Beaman and Holt (2007) a more complicated statistic called an ANOVA (Analysis of Variance) has been used. Excerpts from the results of this research are shown below.

> … analysis of variance (ANOVA) … showed a significant effect of … speech condition on the mean number correctly recalled…, $p<0.013$.



> Overall, the results do not support the general claim that attempting to absorb noise and reduce echo in workplace environments reduces the likelihood of observing auditory distraction effects.

6. DISCUSSION AND CONCLUSION

Sometimes the discussion and conclusion are split into two separate sections, but more often than not, they are included as a single section in published research.

This section follows the general rule that it should be a discussion.

> 'of the issues presented in the introduction *in light of your results*'

This means that the discussion should look again at what was said in the introduction about the claims of past research and to see if your results allow you to say anything about those claims and hypotheses. Your results may allow you to reject a well-known and widely accepted hypothesis, or they may provide evidence that supports it all the more. You can use the discussion to talk in detail about why you found the results you did.

> The results of these experiments are supportive of the view that the extent of auditory distraction observed is dictated by the number of 'changes in state' that occur within the

irrele-vant sound stream and that anything which acts to reduce the number or extent of these changes will reduce the size of the distraction effect.

The words in the irrelevant stream in these experiments were not, so far as can be ascertained by passive listening, rendered unintelligible by the process of adding reverberation, although no formal tests were carried out on whether the relative levels of intelligibility varied between the conditions and that remains a limitation of the current study.

The present results are inconsistent with the view that it is always helpful, when aiming to optimise cognitive efficiency, to reduce the level of reverberation within the environment……

These experiments are the first to demonstrate that intrinsic factors of particular environments can modify the extent of auditory distraction objectively observed in such environments.

Concluding remarks can be brief and clear. They summarise what the researchers feel can sensibly be concluded from their results.

In conclusion, we have established that reverberation can reduce the auditory distraction effect produced by irrelevant speech.

At the end of the discussion you might decide to identify other things that could now be investigated that relate to your research.

Further research looking at the subjective experience of such noise should, in addition, lead to a greater understanding of how auditory environments can be made more conducive to efficient cognitive function both objectively and subjectively.

7. REFERENCES

The final section is a reference list. Here the researchers give full details of the other research in books or academic journals that they have referred to in their investigation. This is very useful for those reading the work as they may be interested to extend their reading to the articles and books indicated by the researchers who wrote the paper. The reference section from Beaman and Holt (2007) includes the following:

1. Ruddle, R. A., Payne, S. J., & Jones, D. M. (1998). Navigating large-scale 'desk-top' virtual buildings: Effects of orientation aids and familiarity. *Presence: Teleoperators & Virtual Environments*, 7, 179–192.

This is a research paper by Ruddle, Payne and Jones. It was published in 1998, in the seventh volume of an academic journal called *Presence: Teleoperators & Virtual Environments* on pages 179 to 192.

2. Ruddle, R. A., Payne, S. J., & Jones, D. M. (1999). Spatial knowledge and virtual environments. In J. M. Noyes & M. Cook (Eds.), *Interface Technology: The Leading Edge* (pp. 135–146). Baldock, England: Research Studies.

This is a chapter of a book called *Interface Theory: The Leading Edge*. The book was edited by Noyes and Cook, and was published in 1999 by a company called Baldock, based in England. The chapter was written by Ruddle, Payne and Jones and is titled "Spatial knowledge and virtual environments".

3. Wann, J. P., Rushton, S. K., Smyth, M., & Jones, D. (1997). Rehabilitation environments for attention and movement disorders. *Communications of ACM*, 40, 49–52.

This is a short article, written by Wann, Rushton, Smyth and Jones in 1997, called "Rehabilitation environments for attention and movement disorders". It appeared in volume 40 of an academic publication called *Communications of ACM*, on pages 49–52. ACM stands for Association of Computer Machinery.