Starting out: Hypotheses and variables

L

The specification says . . .

Formulation of testable hypotheses: Null hypothesis and alternative hypothesis.

Types of variable: Independent variable and dependent variable.

W

People are obsessed with predicting the future aren't they? Whether it's economists trying to predict the state of the country's finances, weather forecasters telling us how much rain we will have next week, or bookmakers trying to work out the 'odds' of some event taking place.

Predicting the future is what keeps fortune-tellers in work. People are desperate to know whether they can look forward to a happy and successful future. The fortune-teller will gaze into a crystal ball, read tarot cards or even interpret the pattern of tea leaves as a way of predicting the future. Whilst many people swear by these methods, others question whether fortune-telling has any basis in fact and is anything more than *'just a bit of a fun'*.

Having said that, scientists in psychology try and predict the future too – though they use *hypotheses* rather than crystal balls, and would call it *research* rather than fortune-telling.

What is meant by ...

Alternative hypothesis states a relationship (correlation or difference) between variables. It is called 'alternative' as in alternative to the null hypothesis.

Dependent variable (DV) The thing that the researcher measures in an investigation. Any changes in the DV should depend on the IV, and the IV alone.

Hypothesis A clear, precise, testable statement that is written at the beginning of an investigation. It states the relationship between the variables being investigated.

Independent variable (IV) The thing that is varied in an experiment – either deliberately changed by the experimenter or varies naturally. There are different levels of the IV – which are called *conditions* of the experiment.

Null hypothesis A statement of no relationship (correlation or difference) between variables.

Variable Any 'thing' that can vary or change within an investigation.

Hypotheses and variables

We have a theory...

We don't know for certain, it's just a thought – but we reckon being in front of an audience might affect how well someone can perform a task. Even though we've been teaching for many years, we still get nervous when we have to stand up in front of a new group of people and we know it affects our concentration. So, we think being in front of an audience may have some effect on performance.

At this stage, our thought is just a *theory* – a suggested explanation for behaviour. Psychologists come up with theories too – you've read about a number of these already in this book. Psychologists *test* their theories by doing *research*. This chapter is about all the different ways in which psychologists can do research, and all the different techniques they use to make their research objective and unbiased.

Formulating an aim

If we're going to put our theory to the test, we need to make it clear what we are testing. We could give people a simple task to do – let's say, throwing some balls into a bucket – to see how well they get on when other people are watching.

All psychological investigations start with an *aim* which is a general statement that explains the purpose of the study; so here's ours:

> *To investigate whether performance on a task (throwing balls into a bucket) is affected by an audience being present or not.*

The next step is to turn the aim into a **hypothesis** that can be tested. A hypothesis states the relationship between the *variables* being investigated.

So, before we get on to the hypothesis, we need to look at variables.

Variables

A **variable** is any 'thing' that can vary or change within an investigation. In an experiment, there are two key types of variable: the **independent variable (IV)** and **dependent variable (DV).**

Independent variable The IV is the thing that the experimenter deliberately changes. In our investigation, the thing that will change is whether participants are performing in front of an audience or alone. That is what is manipulated (changed) by the researcher.

There are actually two levels of the IV here: some participants perform in front of an audience whereas others perform alone. If we had decided to have just one level (performing in front of the audience) then we have nothing to compare that performance with. How would we know whether it was better or worse? So we have a second level of the IV. Each level of the IV is called a *condition*.

Dependent variable The DV is the variable that is measured by the researcher. In a properly run experiment, the *only* thing that should affect the DV is the change in the IV – the change in the DV *depends* upon the change in the IV. Anything else that might potentially affect the DV should be kept constant and controlled (more on that later in this chapter).

In our investigation, the DV is the number of balls (out of 20) the participant manages to throw into the bucket. This is what we will measure.

Operationalisation of variables

It's really important that variables in an experiment – or any other investigation – are as clearly defined as possible. For example, when we say we want to assess the effect of an audience on performance we need to decide what 'performance' is going to mean.

In our study, we decided to measure performance as the number of balls (out of 20) the participant manages to throw into the bucket. This is clear and precise. We could have decided to have the participants do something else, for example sing a song or recite a poem from memory.

Making the variable clear and precise is called *operationalisation* – identifying some clear operational instructions. We can also do this with the IV and specify the exact size of the audience.

E

Formulating a testable hypothesis

Now we are ready for that hypothesis ...

As we have already said, a hypothesis states the relationship between the variables being investigated.

And as we have also said, the hypothesis should be *clear* and *precise* (operationalised). So we need to be as clear and precise as we can about the variables we are testing and the relationship between them:

- The *variables* are: the IV (two levels) and the DV.
- The *relationship* is: there should be a difference between the two levels of the IV.

So the hypothesis could be:

> *There is a difference in the number of balls thrown into a bucket (out of 20) by participants performing with an audience of 30 people or performing the task alone.*
>
> *[relationship + DV + two levels of the IV]*

That looks pretty precise. Notice how when people take part in psychology studies, they are no longer people – they are now *participants* (well OK, they're still technically people but you get what we mean...).

Alternative hypothesis The hypothesis we have written above is called an **alternative hypothesis** – as in, alternative to the **null hypothesis**, coming next. It is a statement of the relationship between the variables.

Null hypothesis The null hypothesis is a statement that there is no relationship between the variables being tested. For example, it might be that having an audience present during the task makes no difference – or very little difference – to the number of balls participants manage to get into the bucket. Participants might perform just as well, or just as badly, regardless of whether there is an audience watching or not. For this reason, as researchers (*which we've now decided we are!*) we need to write another hypothesis which allows for the possibility that this might happen.

> *There is no difference in the number of balls thrown into a bucket (out of 20) by participants performing with an audience of 30 people or performing the task alone.*

Job done. Now we have our two hypotheses (alternative and null), we can get on with the business of finding out which one of them is true.

A

Apply it – research

Light touch

A researcher uses a confederate* to serve students in a college canteen. When she gives change, for half of the students she lightly touches their hand; for the other half she does not. The researcher then asks each student to rate the likeability of the confederate on a scale from 1 (extremely dislikes) to 7 (extremely likes).

Q

*confederate – an individual in an experiment who is not a real participant and has been instructed how to behave by the researcher.

Questions

1. Write an aim for this study. *[2 marks]*
2. Identify the independent and dependent variables. Write your answers as operationalised variables. *[2 marks + 2 marks]*

P

3. Write suitable null and alternative hypotheses. *[2 marks + 2 marks]*

Trying to get all your balls in the bucket – but would you be better or worse in front of an audience?

Stretch and challenge

Can you see how our alternative hypothesis does not say which *direction* the results will go in – that is, it doesn't say whether participants will perform better or worse in front of an audience?

Sometimes it makes better sense to give the direction. For example we might say:

> *Participants performing with an audience throw more balls into a bucket than those performing the task alone.*
>
> *Participants performing with an audience throw fewer balls into a bucket than those performing the task alone.*
>
> *The number of balls thrown into a bucket is greater if participants are performing with an audience than if participants are performing alone.*

There are lots of different ways to write a hypothesis but they all follow the same basic rule of relationship + DV + two levels of the IV.

Study tip

Don't get mixed up

It is easy to get the IV and DV mixed up so use a simple mnemonic to help you remember the difference between them:
Ice **C**reams **D**o **M**elt

IV **C**onditions – **D**V **M**easure. The IV relates to the conditions of your study whilst the DV is the thing that you measure in both conditions.

Check it

1. Distinguish between a null and alternative hypothesis. *[3 marks]*
2. What is a 'dependent variable'? *[1 mark]*
3. What is an 'independent variable'? *[1 mark]*

Extraneous variables

L

The specification says…

Types of variable: Extraneous variable.

Research procedures: The use of standardised procedures, instructions to participants, randomisation and extraneous variables (including explaining the effect of extraneous variables and how to control for them).

Throughout history, many scientific discoveries have happened by accident. Alexander Fleming left a pile of dirty dishes in his lab, went on holiday and returned to find a strange fungus that had killed off all the surrounding bacteria. This led to the discovery of penicillin.

Percy Spencer walked in front of a magnetron machine in his engineering lab and found that the chocolate bar he was carrying in his pocket had melted. As a result, he went on to invent the microwave oven.

John Pemberton was trying to cure headaches using a combination of coca leaves and cola nuts. When his lab assistant accidentally mixed the two substances with fizzy water, Coca Cola was born.

'Lucky mistakes' like these are very rare. When conducting scientific research, researchers have developed systematic procedures to test their hypotheses – including procedures that ensure that unexpected or unwanted variables do not affect their research.

What is meant by …

Extraneous variable (EV) Any variable, apart from the independent variable (IV) that could have an effect on the dependent variable (DV). If extraneous variables are not controlled this means that the researcher cannot know what truly caused the change in the DV.

Randomisation Using chance – such as tossing a coin or picking names from a hat – to control for the effects of bias when designing a research study.

Standardised procedures Using exactly the same methods and instructions for all participants in a research study.

Extraneous variables

Cause and effect

In a properly run experiment, the only thing that should cause a change in the *dependent variable* (DV) is the *independent variable* (IV). Any other things that may affect the DV should be controlled or removed altogether.

If the only thing that has affected the IV is the DV, then the experimenter has demonstrated *cause and effect* – that any change in the DV is due to the IV.

Extraneous variables

Determining cause and effect is threatened by **extraneous variables** (**EVs**). These are any unwanted 'extra' variables that may interfere with the relationship between the IV and the DV.

The experimenter identifies these when designing the experiment, and tries to control them. For instance, consider a study to investigate whether participants work better if they have had a drink of coffee.

In this study the IV is whether participants have a drink of coffee or not. The DV is 'working better' which we can test by giving participants a quiz. We want to know if the coffee *causes* participants to do better on the quiz, i.e. whether the IV changes the DV.

It is possible that any noise or other distractions might also affect quiz performance so these need to be controlled. Also, the temperature and lighting in the room where the experiment takes place should stay constant.

This is why researchers prefer to do experiments in a *laboratory* because things like noise and lighting are much easier to control. However, other EVs, such as the mood and motivation of the participants, may be much more difficult to control even in a laboratory.

Balls in the bucket experiment

There are lots of potential EVs in the balls in a bucket experiment, and these are things we would need to try and control. The only thing that should change between the two levels of the experiment is the IV – whether an audience is present or not. Everything else should be kept as constant as possible.

This would include the room where the throwing task takes place, the size and type of balls thrown, the type of bucket, etc. All these things should be the same for each participant in the investigation.

We might also want to keep certain things about the participants themselves constant. Perhaps they should all be of a similar age, similar height, have similar eyesight, etc. Hmm…this might be more difficult than we thought!

Stretch and challenge

There are lots of different types of extraneous variable in any experiment.

Aspects of the environment that may affect the performance of participants, such as noise and lighting, are called *situational variables*.

There might also be things about the participants themselves that influence their performance, such as their mood, concentration or skill level. These individual differences (differences between people) are called *participant variables*.

Finally, there may be things the experimenter does that affect the IV, such as giving more information to some participants than others, or being nicer to some participants than others. These are called *experimenter* or *investigator effects*.

A

THINK LINK

There are examples of randomisation in Chapter 1 on Memory. Studies of LTM and STM involving lists of words – such as the study of coding by Baddeley (page 14) and the primacy-recency effect by Murdock (page 20) – would have used a random method to determine the order of the words.

L

W

Research procedures

Over the years, *scientists* have learned that it is very important to design their studies in a systematic way in order to control possible extraneous variables (EVs).

Instructions to participants

All participants should receive exactly the same information throughout an investigation. These are called *standardised instructions*.

Before a study is conducted a researcher will write down exactly what should be said to each participant, word-for-word. This might include:

- Some information about the task that is involved.
- How participants should record their answers.
- What a participant should do if they wish to quit the study (they should be told they can leave at any time).

These standardised instructions ensure that what is said to participants does not act as an EV. For example, a researcher might tell one participant to do their best but not say that to another participant and this could affect their performance.

E

Standardised procedures

The standardised instructions are part of the wider **standardised procedures**. All participants should be given the same information at the start of the study, identical instructions throughout the study and be tested in the same environments. The only thing that should vary is the IV.

This is easier to achieve in a laboratory setting than, for example, assessing the effects of coffee by going up to people in coffee shops and giving them a memory test. In a laboratory, it is easier to make sure everything is the same for each participant. Any differences would act as EVs.

Randomisation

Randomisation is using *chance* during the design of the investigation. This means that key aspects of the procedure are not decided by the researcher, but are decided randomly. This reduces *bias* – a leaning in one direction rather than another.

For instance, when testing the primacy-recency effect on recall of words (page 20), the order of the words on the list should be decided randomly. This ensures that a researcher does not accidently put all the easy words at the start of the list which might mean that participants remember more of these words. The word order might act as an EV.

Q

P

People often don't quite get what 'random' means – for example they think that it would be very unlikely that the lottery balls might be 1, 2, 3, 4, 5, 6, but that is as likely as any other six numbers if each ball has an equal chance of being selected.

Apply it – research

AM or PM?

A researcher wanted to find out if performance on a memory task was better in the morning or afternoon. He selected two groups of participants. One group was tested in the morning and the other group was tested in the afternoon.

Questions

1. Write **one** standardised instruction the researcher could have read to the participants. *[2 marks]*
2. Apart from instructions, identify **one** extraneous variable that could have been a problem in this study. *[2 marks]*
3. Explain how the extraneous variable you identified in question 2 could have affected the outcome of the study. *[1 mark]*
4. Explain **one** way in which the researcher could have used randomisation in this study. *[2 marks]*

If a researcher was testing *primacy-recency effects* when recalling a word list, the order of the words should be decided randomly and not by the researcher.

Not a problem. But where are this researcher's clothes and face? Now that is random.

Study tip

Operationalise!

Don't forget to operationalise the variables when writing your hypothesis – clearly stating how each variable will be measured is really important.

Check it

1. Explain how a researcher could use randomisation to decide the order to use when testing participants. *[3 marks]*
2. Explain why it is important to use standardised procedures with participants in research. *[2 marks]*
3. Using an example, explain what is meant by an 'extraneous variable'. *[3 marks]*

Types of experiment

The specification says...

Designing research: Quantitative and qualitative methods.

Laboratory experiments, field and natural experiments.

Strengths and weaknesses of each research method and types of research for which they are suitable.

Obedience

Sometimes people obey a request and sometimes they don't. What makes the difference?

One study by Leonard Bickman in 1974 showed that a uniform did make a difference. Bickman used three male actors (confederates) dressed either in normal clothes or as a milkman or as a security guard.

The actors made requests to people passing by such as 'Pick up this bag for me', or 'This fellow is over-parked at the meter but doesn't have any change. Give him a dime', or 'Don't you know you have to stand on the other side of the pole?'

The results were clear. Many more people obeyed the orders from the actor dressed as a guard than the other two.

So if you want people to do as you tell them – get a uniform (but make sure it's the right kind).

(Obedience is a topic you will study in Chapter 5.)

What is meant by ...

Field experiment An experiment that takes place in a natural setting. The researcher deliberately changes the independent variable (IV), and measures the effect of the IV on the dependent variable (DV). It is more difficult to control extraneous variables (EVs) in a field experiment than in a laboratory experiment.

Laboratory experiment An experiment that take place in a controlled environment. The researcher deliberately changes the IV, and measures the effect of the IV on the DV. It is easier to keep control of EVs in a laboratory experiment than in a field experiment.

Natural experiment An experiment where the IV is not manipulated by the experimenter but would have changed whether the experimenter was interested or not. The experimenter records the effect of the change on the DV. Natural experiments can take place in real-life settings (in the 'field') or in a laboratory.

Qualitative method Using data that is expressed in words and is non-numerical (although qualitative data may be converted to numbers for the purpose of analysis).

Quantitative method Using data that can be counted, usually given as numbers.

Quantitative and qualitative methods

Psychologists have two ways of collecting data about people. One approach is to collect numbers – to find out how many balls a participant can get in a bucket, or how many answers a person can get right, or to see how tall different people are and so on. Such research is **quantitative** because it deals in quantities (numbers).

The alternative is to collect descriptions, for example to ask people to describe how good their memory is or look at the different ways that men and women present themselves on TV. This is a **qualitative** approach.

Experiments

Experiments are a quantitative approach. All experiments involve an *independent variable* (IV) and a *dependent variable* (DV). The experimenter looks for a measurable change in the DV which has been caused by changes in the IV. However, *how* the IV changes, and in what *situations*, depends on the type of experiment.

Laboratory experiments

A *laboratory* is a controlled environment. The experimenter has a high degree of control of everything that happens in that space – for example what participants hear and what they see. A **laboratory experiment** is an experiment that takes place in a laboratory.

Strengths

One strength of a laboratory experiment is that *extraneous variables* (EVs) can be controlled. This means the researcher can be more certain that any changes in the DV are due to the IV and not some other variable. This means that we can be more confident about cause and effect conclusions.

Another strength is that laboratory experiments can use *standardised procedures* because of the control that is possible. This controls EVs but also means the experiment can be repeated (*replicated*) by another researcher to see if the results are the same. If they are, this confirms the **validity** of the findings.

Weaknesses

One weakness is that laboratory experiments may not be like everyday life. When throwing a ball in a bucket in a highly controlled setting, participants may not feel the same as they do when, say, doing a talk in front of a 'real' audience or playing sport in front of a 'real' crowd. This means that the behaviour produced in a laboratory may not generalise that well to the wider world. Laboratory experiments may have low validity.

Another weakness is that in a laboratory experiment participants know they are being tested. This may cause them to change their behaviour to try to help the experimenter get the results they want. This means that participants' behaviour is not like it would be in everyday life and the data collected will lack validity.

Validity

A number of the strengths and weaknesses on this spread refer to increased or decreased validity – in fact we have also been using the concept of validity throughout this book when evaluating research.

What is it? Put simply it refers to whether a result is 'true'. It does not mean that the researcher got the 'right' answer – it means that the researcher got a realistic answer..

Valid research represents something that is real. When a researcher conducts a study, they want their results to relate to everyday life.

Validity is probably THE MOST IMPORTANT CONCEPT in research methods. It concerns the question of whether any observed effect is a genuine one and tells us something about real people in everyday life – not just about participants in a research study.

Field experiments

Field experiments take place 'in the field' – not an actual field but just a natural setting such as a gym, a café or a New York street (see the study in the introduction on the facing page). However, as in a laboratory experiment, the researcher still changes the IV to see what effect this has on the DV.

Strengths

One strength of field experiments is that they are often more realistic than laboratory experiments because they are conducted in a natural environment. Participants often don't realise they are being studied in a field experiment so their behaviour may be more natural too. This means field experiments usually have good validity.

Another strength of field experiments is that, like laboratory experiments, they follow standardised procedures. This helps to control some extraneous

A

Weaknesses

One weakness with field experiments is the researcher may lose control of extraneous variables in the real-life setting. This means it is much more difficult to show cause and effect in field experiments. For example, in the Bickman study (facing page), it might happen that there was increased traffic during the milkman *condition* which distracted participants and made them less obedient.

Another weakness is that there may be special *ethical issues* in field experiments. One ethical issue is whether people know they are involved in an experiment, and whether they have given their permission to be involved in all aspects of it. The people who were given instructions in the Bickman study became participants in a field experiment without their knowledge or *informed consent*. It could be argued that this was unethical.

E

Natural experiments

A **natural experiment** is when the change in the IV is not brought about by the experimenter. A change in the IV is happening anyway – whether the experimenter is interested or not – he or she just records the effect on the DV.

For example, a researcher might want to know if younger or older people are happier. Age is the IV but this would not be changed by the experimenter!

Here's another example. Let's say, as part of their PE lessons, some of our ball throwing participants had extra training on how to throw a ball. We might be interested to see if students who had extra training were less affected by the presence of an audience. Now the IV would be extra training or no extra training, and the DV would be performance on the task in front of an audience. However, in this case, the IV is not controlled by the experimenter – the experimenter is making use of something that varies anyway. Therefore this is a natural experiment.

Strengths

One strength is that natural experiments usually have high validity because, for example, they involve real-life changes and these normally occur in a natural setting. In such a study, we would still eventually test the ball throwing abilities of our participants in a laboratory-like environment.

Another strength of natural experiments is that when the DV is measured in a laboratory there can be a high level of control over EVs.

Weaknesses

Q

One weakness is that the natural event that the psychologist is interested in studying may only happen very rarely, such as the effect of a hurricane on stress levels. This means there may be fewer opportunities for this kind of research.

Another weakness is that there may be unique characteristics of the participants. In the example above, those students who took part in extra PE lessons may have been chosen because they were better (or worse) at throwing a ball anyway. The researcher can't *randomly allocate* participants to the conditions of the IV. This may act as an EV and affect the results of the investigation.

P

Does sunny weather put people in a better mood than cloudy weather? Now that sounds like a cue for a natural experiment. Weather is the IV.

Apply it – research

Decisions

Researchers wanted to investigate whether...

- Being deprived of sleep for 48 hours has a negative effect on memory test performance compared to sleeping 'normally'.
- People who have been the victims of a serious crime are more likely to experience stress-related illnesses than people who have not been victims.

Questions

Which of the above aims would lead to:

1. A natural experiment? Explain your answer. *[2 marks]*
2. A laboratory experiment? Explain your answer. *[2 marks]*
3. A field experiment? Explain your answer. *[2 marks]*

Stretch and challenge

Not all laboratory studies are laboratory *experiments*. A researcher may set up a controlled environment in which to view behaviour. However, there may be no manipulation of an IV – there may be no IV at all. The researcher will just observe behaviour as it happens, and record what they see.

For instance, a researcher may observe interactions between a parent and child through a one-way mirror and record different categories of behaviour. This would be a *controlled observation* rather than a laboratory experiment.

THINK LINK

Many of the experiments in the Memory and Perception chapters are carried out in the laboratory – for instance, those that investigate the effect of factors on perceptual set in Chapter 2 (pages 52–57).

There are examples of field experiments, such as Godden and Baddeley in Chapter 1 (page 28) and of natural experiments, such as Bennett in Chapter 3 (page 77) or Kaij's twin study in Chapter 8 (page 220).

Check it

1. Distinguish between quantitative and qualitative methods. *[3 marks]*
2. Natural experiments are said to have more validity than laboratory experiments. Explain why this might be so. *[3 marks]*
3. Explain **one** strength and **one** weakness of conducting a field experiment. *[4 marks]*

Experimental designs

The specification says...

Designing research: The experimental method.

Experimental designs: Independent groups, repeated measures, matched pairs, including strengths and weaknesses of each experimental design.

Allocation to conditions, counterbalancing.

One of the first reported attempts to design experiments was by Robert Fisher in 1935. In his book called (would you believe) *The Design of Experiments* he described the *lady tasting tea* experiment.

The experiment had a very small sample. It involved one lady who claimed she could tell whether the tea or the milk had been added to the cup first. Fisher set up eight trials, four where tea had been added first, and four where it was milk first. The cups were presented to the lady in a random order.

In the language of modern psychology, this is a *repeated measures design*. It was also one of the first reported experiments to make use of a *null hypothesis*. For the record, the lady got all eight cups correct.

Experimental designs

One aspect of the design of experiments is the type of experiment used by a researcher – *laboratory*, *field* or *natural*.

There is another aspect of the design of experiments, which is called (*yes, you've guessed it*) **experimental design**. This is a little confusing because you might think that all aspects of the design of experiments would be called 'experimental design' but, no, this refers specifically to which *conditions* of the experiment participants will take part in (i.e. which level of the IV a participant is given). There are three options described below.

Independent groups

The clue's in the name! An **independent groups** design is when there are separate groups of people for each level of the independent variable (IV) in the experiment (i.e. each *condition* of the experiment). Often there are just two levels of the IV, one of which acts as a control or baseline for making comparisons. So one group is called the **control group** and the other is the experimental group.

In the balls in the bucket experiment, one group of participants would throw balls with an audience (*experimental group/condition*), and another group would throw balls without an audience (*control group/condition*). The performance of the two groups is then compared.

Strength

One strength of this design is that **order effects** are not a problem – these are described on the facing page. Each participant only does the task once and therefore, for example, there is no *practice effect*. Order acts as an *extraneous variable* (EV) which is good to avoid.

Weakness

One weakness of this design is that the participants in each group may differ. What if all the really good ball throwers are in one group? Or all the confident people? These are called *participant variables*. The difference in performance may be more to do with participant variables than the change in the IV. This acts as an EV and reduces the *validity* of the results.

Dealing with participant variables: Allocation to conditions

The participant differences described above can be dealt with using an unbiased method of **allocation to conditions**.

This can be done in a variety of ways. For example, the first person to arrive to the experiment can be placed in one group, the next in another group, and so on. This is a systematic method.

Or a random method can be used such as flipping a coin – participants who get heads go in the experimental group and those who get tails go in the control group (this is another form of *randomisation* – see page 97). You may end up with more participants in one group but that is not a problem.

This procedure is done so the researcher does not influence who goes in which group. It also means that, chances are, participant variables will be evened out between groups/conditions.

What is meant by ...

Allocation to conditions In an independent groups design, each group receives one level of the IV (one condition of the experiment). Random allocation is an unbiased unbiased method used to control for participant variables. If random methods are used this ensures that each participant has the same chance of being in one group as any other.

Control group A group of participants who receive no 'treatment'. Their behaviour acts as a baseline against which the effect of the independent variable may be measured.

Control condition The condition in a repeated measures design that provides a baseline measure of behaviour without the experimental treatment (IV).

Counterbalancing Used in repeated measures designs to control for order effects. Half the participants complete the conditions in one order, and the other half in the opposite order.

Experimental design The different ways in which participants can be organised in relation to the conditions in an experiment.

Independent groups Participants are allocated to different groups where each group represents one experimental condition (level of the independent variable), e.g. condition A and condition B.

Matched pairs Pairs of participants are matched in terms of variables relevant to the study, such as age or IQ. One member of each pair takes part in condition A of the experiment and the other takes part in condition B.

Order effects In a repeated measures design, an extraneous variable arising from the order in which conditions are presented.

Repeated measures When all participants take part in all the conditions of the experiment.

A

Apply it – research

Designing

A team of researchers conducted two studies and discovered that...

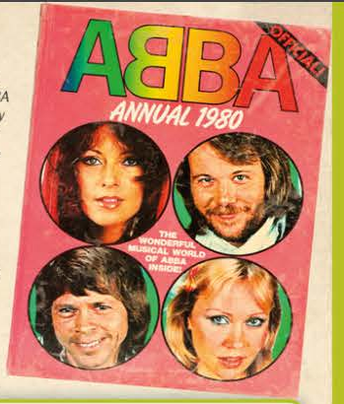
- People who had high stress levels were more likely to develop colds than people with low stress levels.
- People performed better on memory tasks after they took a caffeine tablet than they did before they took it.

Questions

For each of these studies:

1. Identify the experimental design. *[1 marks]*
2. Explain **one** strength of the design. *[2 marks]*
3. Explain **one** weakness of the design. *[2 marks]*
4. Explain how this weakness could be dealt with. *[2 marks]*

It's a little known fact that the Swedish pop group ABBA took their name from a way of reducing order effects in a repeated measures design experiment. The ABBA design is a sophisticated form of counterbalancing – the two conditions are called A and B. Each participant does four trials, A, B, B then A.



E

Repeated measures

Perhaps it might be better to compare *like with like*. In a **repeated measures** design, all the participants receive all levels of the IV (i.e. they are involved in all conditions of the experiment). So, participants would throw balls in the bucket with an audience present, and also throw balls in the bucket without an audience present. The two sets of results (data) for each participant would then be compared.

Strengths

One strength of repeated measures is that the problem of participant variables is removed. Each participant is compared against themselves rather than other people. This means that the participant variables are controlled.

Another strength is that fewer participants are needed in this design than in an independent groups design. If you have 20 participants this produces 20 sets of comparisons whereas in an independent groups design you need 40 participants to get 20 sets of comparisons. This means the study may be more expensive because more participants are needed.

Weakness

One weakness is order effects. When participants are tested twice the order in which they do the tasks may make a difference. In the balls in the bucket study, participants do the task twice – they throw the balls in front of an audience and then do the same thing with no audience. The second time their performance may be better because of practice rather than because of having no audience. This is called (again no surprise) a *practice effect* which acts as an EV. This is likely to reduce the validity of the results.

Dealing with order effects: Counterbalancing

Counterbalancing tries to control order effects in repeated measures designs. Because all participants will eventually take part in both conditions – half the participants should complete the conditions in one order, and the other half should complete the conditions in the opposite order.

So, the first participant takes part in the **control condition** (A) and then the experimental condition (B), the second participant takes part in the experimental condition (B) and then the control condition (A), and so on. If there is an issue with participants completing the conditions in a particular order, this will be evened out across the whole investigation.

Matched pairs

A **matched pairs** design is a kind of 'halfway house' between the other two designs. In a matched pairs design, participants only take part in one condition of the experiment – as in an independent groups design. The participants are in two separate groups; however, these groups are not made up of completely different people. The participants are tested on key participant variables (such as ability to throw a ball or self-confidence). Then, they are put into pairs by matching the person with the best ball throwing ability with the person with the second best and so on. One of each pair is allocated to a different group of the experiment.

Strengths

Matched pairs is a 'happy medium' between the other two designs and so deals with some of the problems of both. Because the participants are only tested once, this means there are no order effects as there are in repeated measures designs.

As well as this, participants are matched on a variable that is important for the experiment. This partly solves the problem of participant variables which are a feature of an independent groups design.

Weakness

One weakness is that matching participants takes time and is never exact. It takes time to get the data to do the matching. Plus, it is also likely that only one or two key variables will be matched so some participant variables remain. This means the method sometimes involves quite a lot of effort for little gain.

Q

P

Stretch and challenge

We discussed extraneous variables on page 96. One important type of extraneous variable is *demand characteristics*. These are features of the experimental situation that may give away the aim of the investigation and 'demand' a particular response from participants. Demand characteristics can affect the validity of the results because participants are unlikely to act naturally if they work out the aim of the study.

The more aspects of the investigation the participants see, the more likely they are to work out the aim. This means that demand characteristics tend to be a bigger problem in repeated measures designs as participants are involved in all conditions of the experiment.

Check it

1. Explain how counterbalancing is used with a repeated measures design. *[3 marks]*
2. Describe and evaluate the matched pairs design. *[6 marks]*
3. Explain **one** strength of using an independent groups design. *[3 marks]*

Sampling methods

The specification says…

Sampling methods: Target populations, samples and sampling methods and how to select samples using these methods: random, opportunity, systematic and stratified.

Strengths and weaknesses of each sampling method.

Understanding principles of sampling as applied to scientific data.

2015–16 was the year in which 'the opinion polls got it wrong'. The Conservative Party were widely predicted to lose in the 2015 UK General Election – but they won.

It was assumed that those supporting Brexit would lose out in the EU Referendum in 2016 – but they won.

Donald Trump was given little chance in the opinion polls against Hillary Clinton for the US Presidency – but he won.

The issue is that no matter how large an opinion poll is, it can never include everybody who is going to vote in an election. Opinion polls are always a 'best guess' of what might happen based on the *sample* of the population.

What is meant by …

Opportunity sample Produced by selecting people who are willing and available at the time.

Random sample Produced by using a random technique in which every member of the target population has an equal chance of being selected.

Sample A subset of the target population which aims to be representative of that population.

Sampling method The system used to produce a sample.

Stratified sample Produced by selecting participants in proportion to their frequency in the target population.

Systematic sample Produced by selecting every *n*th person on a list of the target population.

Target population The group that the researcher is interested in studying, from which a smaller sample is selected.

The principles of sampling

Target populations and samples

The point of psychological theories is that they are meant to tell us something about people in general – for instance, whether people put on a better performance with an audience than without. It's not possible to test everybody in the world (obviously) – or even all the people in your town. Therefore psychologists choose a smaller group of people to study known as a **sample**.

The sample is drawn from a larger group called the **target population**. This is the group of people about whom we wish to make a statement – so we might wish to draw conclusions about all the people in the world, or all men in the world, or just men in our town.

Generalisation

The purpose of the sample is to be able to *generalise* the results to the target population, i.e. make a statement about *all* of the people in the target population. This means the sample must properly *represent* the target population. If a psychologist wanted to say something generally about people from the town of Oldham, then the sample they choose must be made up of a cross-section of the different people who live in Oldham – including for instance, a mix of different genders, ages, educational attainment and ethnic groups.

Target population

Participants are drawn from a **target population** (the group of people the researcher is interested in).

Bias

Obtaining a representative sample is quite difficult to do. Even if the psychologist did manage to find a good mix of people, it still might be that they all live in the same village, or that they all happen to be students. Most samples in psychological research are *biased* to some extent. It can be almost impossible to select a group of participants for an investigation that perfectly reflects the target population – particularly if the target population is very large.

The representativeness of the sample will also depend on what **sampling method** is used – and some of these methods tend to be less biased than others.

Apply it – research

You choose

Two psychologists conducted a series of studies using different sampling methods to select their participants. They used…

- Random sampling to choose 20 nurses from a local hospital for a study of stress.
- Opportunity sampling to choose 15 students from a local sixth-form college to complete a questionnaire.
- Systematic sampling to choose 50 households in the county of Somerset to conduct face-to-face interviews.
- Stratified sampling to compare the attitudes of teenagers and elderly people in a small town towards illegal drugs.

Questions

For each study:

1. Identify the target population. *[1 mark]*
2. Explain **one** way in which the sample could be selected. *[2 marks]*
3. Explain **one** strength of the method used. *[2 marks]*
4. Explain **one** weakness of the method used. *[2 marks]*

L E W

Sampling methods

Random sampling

In a **random sample**, every person in the target population has an equal chance of being selected.

How to do it The researcher needs a list of everybody in the target population. For example, if they were doing this in a particular school, they would need a list of all the teaching staff, support staff, students, cleaners, etc. All the names would be given a number and then the sample is selected using a random method – for instance, picking numbers from a hat or using a random number app.

Evaluation One strength of random sampling is there is no bias. Every person in the target population has an equal chance of being selected. If there is no bias the sample should be representative of the target population.

One weakness is that this form of sampling takes more time and effort than other methods. You need to obtain a list of all the members of the target population, identify the sample and then contact the people identified and ask if they will take part. In the end many people may not be willing to participate so the sample ends up being biased.

Opportunity sampling

In an **opportunity sample**, the researcher selects people who happen to be there at the time.

How to do it For example, the researcher goes into a school common room and selects students who are willing and available to take part in an investigation.

Evaluation One strength is that opportunity sampling is relatively easy to do because you simply go to people who are nearby. This makes it quicker and therefore cheaper than other methods.

One weakness is that the sample is likely to be unrepresentative of the population, particularly if it is drawn from one place, like a common room. This means that the results from the study can only be generalised to the target population from which the sample was drawn.

Systematic sampling

In a **systematic sample** a numerical formula is used, such as selecting every 50th name from the phone book.

How to do it The researcher again needs a complete list of the target population. They will then select, say, every third, or fifth, or tenth (or any other number) person on the list until they have a complete sample of participants.

Evaluation One strength is that, like random sampling, systematic sampling avoids researcher bias. Once the system of selection has been decided the researcher has no say over who is selected. This increases the representativeness of the sample.

One weakness is that, although it is usually representative, it may still be biased, for instance you might end up with an all-men or all-women sample using this method (the same could happen with random and opportunity sampling). This means that representativeness is not guaranteed.

Stratified sampling

In a **stratified sampling** participants are selected from different subgroups (strata) in the target population in proportion to the subgroup's frequency in that population.

How to do it For example, if a researcher wishes to sample different age groups in a school, they first of all have to identify how many students are, for example, aged 10–12, 13–15 and 16–18.

Let us say there are 1000 students in the school and 300 are 10–12 years old, 500 are 13–15 and 200 are 16–18. Therefore the frequency of the subgroups is 30%, 50% and 20% respectively.

If the researcher wishes to have a sample of 50 participants then 30% should be aged 10–12 – which is 15. The researcher than randomly selects 15 participants from the 300 students in that age group.

Evaluation One strength is that this is the most representative of all the sampling methods. The researcher identifies important sub-groups and makes sure that these sub-groups are represented in proportion to the numbers in the target population. However, the decision about which sub-groups to use may be biased, thus reducing the representativeness of the sample.

One weakness is that this is a very lengthy process and those participants selected may not always agree to take part. This means that it is not a method that is used much in psychological research.

Is it possible to include all the ways in which people are different in a psychology sample?

Stretch and challenge

There are other types of sampling. One common type is *volunteer sampling*. This is an easy option for the researcher as the participants will come to the researcher, so this requires little effort to set up.

The researcher may advertise for volunteers by pinning a notice on a common room wall, for instance, or putting an ad in a local paper.

Volunteers are not likely to be very representative of the target population as they are *self-selecting*. There may be a high number of 'unusually helpful' people who happen to be free when the investigation is taking place.

In fact, many of the methods on the left end up being volunteer samples because people say no when asked to participate – so the researcher basically ends up with the most willing.

Q P

Study tip

Go compare

Comparing the different sampling methods is a good way to evaluate them. For example 'Opportunity sampling is easy to carry out but not representative whereas stratified is more difficult to carry out but is more representative'.

Check it

1. Describe how a psychologist could use random sampling to select 20 participants from a local primary school. *[3 marks]*
2. Outline **one** strength and **one** weakness of using random sampling. *[6 marks]*
3. Explain what is meant by 'target population'. *[2 marks]*
4. What is a systematic sample? *[2 marks]*

Ethical considerations

L

The specification says...

Ethical considerations: Students should demonstrate a knowledge and understanding of:

- **Ethical issues in psychological research as outlined in the British Psychological Society guidelines.**
- **Ways of dealing with each of these issues.**

Don't stand too close

Yes they are called *public* toilets, but you would expect to have some privacy there – right?

Not if you happened to be a participant in a study by Dennis Middlemist and colleagues in 1976. The researchers wanted to know how men would react to their personal space being invaded. They decided to carry out a field experiment in a men's urinal.

There three IVs/conditions: a confederate (ally of the researcher) stood either immediately next to a participant, one urinal away, or was absent (the control condition).

An observer recorded how long it took before the participant started to urinate. This was taken as a measure of how comfortable the participant felt. The longer it took for them to urinate, the more anxious they were judged to be.

Not surprisingly, participants took longer to start to urinate when their personal space was invaded. Now that's just taking....things a little too far.

What is meant by ...

British Psychological Society (BPS) guidelines A code of conduct that every professional psychologist in the UK has to follow. It explains how participants should be dealt with, and is based on four main principles: respect, competence, responsibility and integrity.

Ethical issue When there is a conflict between the rights of participants to be safe and the goals of research to produce valuable data.

Ethical issues in psychology

Obviously, psychologists want to find out as many interesting things as they can about people (and animals) in lots of different situations. But how far should psychologists be allowed to go in getting their data?

An **ethical issue** comes about when there is a conflict between participants' rights and well-being on the one hand, and the needs of researchers to gain valuable results on the other hand. The well-being of participants should be protected at all times.

We shall consider five ethical issues in psychological research. Note that, in all of these issues, children under the age of 16 are a special case and should not be exposed to any risks.

Informed consent

At the beginning of a study, participants should be given comprehensive information concerning the nature and purpose of a study and their role in it. This should include the aims of the research, the procedures and the fact they can leave the investigation at any time if they wish (*right to withdraw*). Participants can then make an informed decision about whether they want to take part – that is, give their *informed consent*.

The researcher may not wish to reveal the aims of the study. If participants know the aims, this might alter their behaviour and affect the *validity* of the study. In such cases participants must be informed at the end of the study what the true aims were.

If some information is withheld at the start of the study, participants still need to know enough information about the procedures to make their informed decision.

Deception

Participants should not be lied to or misled about the aims of a study. Participants who have been deceived, or not given full information, cannot not give fully informed consent.

However, there is a difference between lying and simply withholding some information. Both are forms of *deception* but *mild* deception can be justified. If there were two groups taking part in the balls in the bucket experiment (*experimental* and *control condition*), then we may not tell either group about the other until the end of the study. This would be considered mild deception.

Sometimes psychologists do conduct studies after much consideration that involve more major deception. But this is only permitted if the benefits of the research justify this action.

Protection from harm

Participants should not be placed at risk, and their physical and psychological safety should be protected at all times. This includes not being made to feel embarrassed or stressed. Several famous studies in psychology have placed participants under severe stress (such as the study of personal space invasion on the left). Part of protection from harm is reminding participants that they are free to leave the investigation at any time.

Privacy

People expect to be able to control information about themselves. In some research (such as the personal space study on the left) participants are in a public place. It has generally been regarded as acceptable to make *observations* of people in public – except there isn't universal agreement on what exactly counts as a public place. For example, some people think that listening to someone's conversation on a bus and using the information in research is not ethically acceptable.

Confidentiality

If a person's privacy has been invaded, their identity should be protected. Data collected about a person, even the results of an experiment, belong to that person. To protect their privacy their personal details should not be accessible. In fact, in most cases, participants are anonymous.

Ethical decisions involve weighing up the *benefits* of research, set against the potential *costs* to the well-being of participants.

L E W A Q P

THINK LINK

Ethical issues arise in every piece of research in this book. Milgram's study of obedience in Chapter 5 (see page 136) is probably one of the most discussed. Our concern is mainly for ethical issues with humans but psychological research also involves animals, for example Seligman's study of learned helplessness (see page 227).

Ways of dealing with ethical issues

The BPS guidelines

The British Psychological Society (BPS) is the professional group who, amongst other duties, monitor the behaviour of psychologists in the UK.

Professional bodies such as those for lawyers, doctors, etc., produce ethical guidelines for their membership – the group of professionals 'police' the behaviour of their group. Ethical guidelines are a set of rules about how professionals should deal with ethical issues. The **BPS guidelines** have to be followed by all professional psychologists in the UK. If a psychologist fails to do this, they are not sent to prison but they do risk losing their job.

When a psychologist is planning a research study they have to decide how they will deal with the ethical issues that arise. They have to present their plans to an *ethics committee*. This is a panel of experts who will decide whether the proposed investigation meets the ethical guidelines or not.

The BPS guidelines include the following three issues as well as more general advice (see 'Stretch and challenge' on the right).

Dealing with informed consent

To deal with informed consent, researchers should produce a letter for participants (or their guardians) explaining what the study involves and, if appropriate, the aims. This should explain everything participants need to know about the investigation so they can make an informed choice about whether to take part.

Participants are then asked to sign a consent form. If participants are under 16, this form should be signed by a parent or guardian.

In some field experiments or observations, it may not be possible for participants to give their consent until the end of the study (retrospective consent) or it may not be possible to gain any consent. For example, in Piliavin *et al.*'s study (see page 142) the participants were all travellers on the New York subway. They never knew they had been participants in a psychological study.

A researcher must have a strong justification for withholding the opportunity for consent.

Dealing with deception and protection from harm

At the end of a study, participants should be given a full *debriefing*. This should inform them of anything they were not told about during the study – such as the true aims, the existence of other groups of participants, etc. The debriefing should also let participants know they have the right to withhold data if they are unhappy about aspects of the study.

If participants were made to feel uncomfortable or anxious during the study or question their own abilities after the study, the psychologist has a responsibility to put this right. Participants should be reassured that their behaviour was 'normal' or 'typical', as this is something they may feel concerned about. Participants also might be offered counselling.

Dealing with privacy and confidentiality

All participants should be anonymous. When writing up the research report, researchers should not refer to participants by name. Instead, they should call them Participant 1, Participant 2, etc. Sometimes, psychologists will use the initials of the person, for instance, the case study of HM (see page 17). Anonymity can include not disclosing where the study took place in the final report.

Psychologists should not share their data with other researchers – unless this has been made clear to the participants in advance.

Finally, participants should be reminded that their data will be protected and remain confidential.

Following psychological research, a full *debrief* is essential so participants are made aware of their rights, and what their data will be used for.

Apply it – research

Puzzled

Researchers gave their participants a puzzle that deliberately did not have a solution. Several participants spent a long time trying to solve it and became quite stressed.

Questions

1. Name **two** ethical issues the researchers should have considered when designing this study. *[2 marks]*
2. For each ethical issue you identified in question 1 explain why it is an issue. *[1 mark + 1 mark]*
3. For each ethical issue you identified in question 1 explain **one** way of dealing with the issue. *[2 marks + 2 marks]*

Stretch and challenge

The British Psychological Society (BPS) ethical guidelines are organised around four main principles:

1. *Respect* covers several areas including confidentiality, informed consent, avoiding deception and the right to withdraw.
2. *Competence* refers to the fact that psychologists should be properly qualified and professional at all times.
3. *Responsibility* includes protecting participants from harm.
4. *Integrity* means that psychologists should give honest and accurate advice when dealing with participants.

Check it

1. Using an example, explain what is meant by an 'ethical issue'. *[3 marks]*
2. Outline **one** ethical issue and explain how it can be dealt with. *[2 marks + 2 marks]*
3. Explain the purpose of the British Psychological Society guidelines. *[2 marks]*

Interviews and questionnaires

The specification says...

Designing research: interviews, questionnaires.

Strengths and weaknesses of each research method and types of research for which they are suitable.

A famous study by Elizabeth Loftus and John Palmer in 1974 investigated the effect of leading questions on memory. Participants were shown film clips and were then asked, '*How fast was the car going when it _____ the other car?*'

When the gap in this question was filled by the word *contacted*, the average estimate of speed was around 32 mph. With a different group of participants the word *smashed* produced an average speed of 41mph.

This may not sound like a big difference, but legally, 10mph could be the difference between a fine and the loss of a driving licence.

Loftus' research showed that different words can prompt a different answer – changing wording like this has been called 'leading questions'. Understanding the big difference that a word or two can make has been very influential in changing the way police interview witnesses in real-life situations.

Leading questions can also be a problem for psychologists when they are designing questionnaires and interviews. They may create bias in the way

What is meant by ...

Closed question One that has a fixed range of possible answers. They produce *quantitative data*.

Interview An interaction between an interviewer and an interviewee. The interviewer asks the interviewee questions to determine their thoughts and feelings on a particular topic. The questions may be decided in advance, or develop as the interview goes along. An interview may be used as a method of data collection on its own, or as a way of measuring the dependent variable in an experiment.

Open question One that invites respondents to provide their own answer rather than select one of those provided. They tend to produce *qualitative data*.

Questionnaire A set of written questions (also referred to as 'items') used to find out a person's thoughts or attitudes on a particular topic. A questionnaire may be used as a method of data collection on its own, or as a way of measuring the dependent variable in an experiment.

Interviews

Interviews and questionnaires are both examples of *self-report methods*. We can learn about human behaviour by asking people for their thoughts, feelings and opinions.

Most **interviews** involve a face-to-face conversation between the interviewer and the interviewee, though they can also be conducted over the phone or by text. The main thing is that interviewer and interviewee are talking in real time.

There are three types of interview.

Structured interviews

In *structured interviews* the interviewer reads out a list of prepared questions to the interviewee. The interviewer works through the list. The interviewee may be asked follow-up questions such as 'Why do you think that?' or 'Can you give some examples?' but these follow-up questions are also prepared beforehand. In other words the interviewer follows an exact script.

Unstructured interviews

An *unstructured interview* is a lot like a conversation. The interviewer has few if any questions prepared in advance, though there will be a general aim at the beginning. In order to maximise the information gathered, the interviewer develops new questions based on the answers given. For example:

Interviewer: 'Tell me about your childhood.'

Interviewee: 'It was generally very happy except we moved a lot.'

Interviewer: 'Did moving around make it less happy?'

Semi-structured interviews

In practice, most interviews are *semi-structured*. A number of questions have been decided in advance but the interviewer will ask follow-up questions at certain points.

Job or college interviews tend to be semi-structured as do interviews you hear on the television or radio – the presenter has some questions prepared in advance but often 'goes with the flow' and asks questions based on the answers given and when that line of questioning runs out the interviewer returns to the script for a new question.

Evaluation

Strengths

One strength of interviews (particularly unstructured interviews) is that they produce extensive information. An experienced interviewer can ask follow-up questions that provide extra insights into a person's thoughts and feelings. Such data is useful if the researcher wants detailed insight into the way the interviewee sees the world.

Another strength of interviews is that it is the only way to obtain information from people who can't write or find it difficult to express their feelings on paper. It isn't always easy to express your thoughts, so a good interviewer can help by asking probing questions or even rephrasing a question better. This improves the quality of the information collected.

Weaknesses

One weakness is that interview data can be difficult to analyse. The researcher may end up with lots of information – some of it irrelevant – so drawing general conclusions may be complex. The conclusions that are drawn may be more based on the researcher's preconceived ideas than what the interviewee originally meant.

Another weakness is that people may feel less comfortable about revealing personal information than when writing answers to a questionnaire. Some interviews may involve sensitive or distressing topics and interviewees may not want to reveal their true feelings to another person sitting in front of them. This limits the information collected.

A

THINK LINK

Questionnaires may be used as part of an experiment as a way of measuring the dependent variable. See Chapter 6 for an example of how a rating scale is used in Yuki *et al.*'s study to measure perceptions of happiness (page 180).

In Chapter 3 children in the conservation studies (page 72) were questioned to obtain their responses.

L

W

Questionnaires

Questionnaires (also called surveys) are made up of a prepared list of written questions to which a person responds. This can be done face-to-face (as in a structured interview) or completed over the phone, on the Internet or by post.

Questionnaires (or interviews) may be used as a method of data collection on their own, or as a way of measuring the dependent variable in an experiment.

Open and closed questions

In both questionnaires and interviews questions can be open or closed but **open questions** are more likely in an interview and **closed questions** are more likely in a questionnaire. Most questionnaires have a mixture of open and closed questions.

An *open question* does not have a fixed range of possible answers. Respondents are free to answer in any way they wish. For example, we might ask participants in the balls in a bucket experiment to explain how they felt whilst performing in front of the audience. Open questions tend to produce *qualitative* data, i.e. data that is not numerical.

A *closed question* has a fixed range of possible answers. We could ask participants if they felt more nervous when throwing balls in front of the audience and give them two options: 'yes' or 'no'. Alternatively, we could ask them to rate their nervousness on a scale of 1 to 10 (where 1 is not nervous at all and 10 is extremely nervous). Closed questions produce *quantitative* (numerical) data.

Evaluation

Strengths

One strength of questionnaires is that they can gather information from lots of people relatively quickly. It takes a lot of time to prepare a questionnaire but once it is designed it can be sent to thousands of people at the same time. If the researcher has gathered a large number of responses, generalisations are easier to make.

Another strength is that questionnaires produce data that is often easier to analyse than interviews. This is because questionnaires tend to be made up of closed questions that can be converted to quantitative data. This means comparisons can be made more easily between groups.

Weaknesses

One weakness is that respondents may not always answer questionnaires (and interviews) truthfully. They may give the answer they think is appropriate rather than their real feelings. This is called a *social desirability bias* – people tend to give answers that put them in a better light. This means that the data collected is low in *validity*.

Another weakness is that questions in a questionnaire (or an interview) may [illegible] and Palmer's research). This again affects the validity of the answer given.

Q

P

Study tip

Comparisons

When you are giving a strength or a weakness of an interview it always helps to make a comparison to a questionnaire – 'one strength of an interview is you don't have to be able to read and write, which is required in a questionnaire'.

Apply it – research

Feedback

A psychologist studied the effectiveness of a new therapy for depression. A group of participants received ten weeks of the therapy and were given a questionnaire to complete afterwards about how they were feeling. They were also interviewed about their experiences.

Questions

1. Write **one** open question and **one** closed question that could have appeared in the questionnaire. *[1 mark + 1 mark]*
2. Briefly explain how the psychologist could have carried out a semi-structured interview with the participants. *[2 marks]*
3. Explain **one** strength and **one** weakness of using questionnaires to find out how the participants were feeling. *[2 marks + 2 marks]*
4. Explain **one** strength and **one** weakness of using interviews to assess their experience. *[2 marks + 2 marks]*

"That's true. We did advertise for someone who 'works well under pressure'..."

Stretch and challenge

Fixed-choice questions are closed questions because they have a fixed list of options and the respondent has to choose which applies to them.

How many balls did you throw into the bucket in front of the audience?

☐ *0–2* ☐ *3–5* ☐ *6–8* ☐ *More than 8*

A *Likert scale* is a type of fixed-choice question in which the respondent reads a particular statement and indicates their agreement or disagreement.

It is more difficult to throw balls in a bucket in front of an audience.

1	2	3	4	5
Strongly agree	*Agree*	*Neutral*	*Disagree*	*Strongly disagree*

A *rating scale* is similar, where participants represent their views in numbers.

☐ *How much stress did you feel during the experiment?*
A lot of stress = 3 Very little stress = 2 No stress = 0

Check it

1. Explain the difference between a questionnaire and an interview. *[3 marks]*
2. Identify **three** steps in conducting an interview. *[3 marks]*
3. Describe and evaluate the use of a questionnaire to conduct research. *[9 marks]*
4. Explain **one** weakness of using an interview to carry out research. *[3 marks]*

Observation studies

The specification says...

Designing research: Observation studies (including categories of behaviour and interobserver reliability).

Strengths and weaknesses of the research method and types of research for which it is suitable.

One of the most famous observation studies in psychology was by Leon Festinger and his colleagues (1956). Festinger followed the activities of a woman called Marian Keech who led a religious group called the Seekers.

Keech believed that she was in contact with alien beings from a planet called Clarion who warned her that a flood was coming which would destroy America. However a flying saucer would rescue Keech and her followers on the day of the flood.

The Seekers had such faith that the messages were real that they left their jobs and, in some cases, their families to prepare to escape the flood. On the day the flood was supposed to happen, the Seekers gathered to wait for the alien craft that would transport them to safety. When the flood (or the flying saucer) never came, Keech explained she had received another 'communication'. Because the Seekers had 'spread so much light', God had decided to save the world from destruction.

Rather than abandon their beliefs, the Seekers faith in Keech and her ability to communicate with aliens actually got even stronger.

What is meant by ...

Categories of behaviour When a target behaviour (for example, aggression) is broken down into units that can be observed and recorded, such as number of kicks, number of punches, etc.

Interobserver reliability The extent to which there is agreement between two or more observers involved in observations of a behaviour.

Observation studies A researcher watches or listens to participants engaging in whatever behaviour is being studied. The observations are recorded. Observational methods are often used to measure the dependent variable in an experiment.

Observation studies

Rather than asking people what they might do (as in *questionnaires* and *interviews*), another way of investigating behaviour is to watch what people *actually* do. We all 'people-watch' on a daily basis. **Observation studies** give researchers a chance to record behaviour just by sitting back and taking notes.

Types of observation

Observational research is quite flexible. Researchers have several options in terms of the type of observation they might carry out.

1. Naturalistic versus controlled Often observational research takes place in a natural setting. In a *naturalistic observation*, behaviour is recorded in the place where it would normally occur and nothing is changed in the environment. Sometimes a researcher might decide that they want to control some aspects of the situation, so a *controlled observation* might be better. The Zimbardo prison study (described in the introduction of this book, see pages 4–5) is an example of a controlled observation as many aspects of the environment (the mock prison) were controlled.

2. Covert versus overt A *covert observation* is when participants are not aware that their behaviour is being recorded. For this to be *ethical*, the behaviour must be happening anyway in a public context. The opposite is an *overt observation* in which participants are told in advance that their behaviour will be watched and recorded.

3. Participant versus non-participant A researcher has to decide whether to become part of the group they are studying or not. In the study on the left, Festinger and his colleagues actually joined the Seekers in order to observe them and therefore this is a *participant observation*. Alternatively, the researchers can remain separate from the people they are studying – a *non-participant observation*.

Categories of behaviour

There are different ways to record behaviour. A researcher can simply write down everything that they see. This might be difficult though because so much is happening – so a researcher usually makes the collecting of data a bit more systematic. Often, the observer will break up the target behaviour into different **categories of behaviour**. For instance, if the target behaviour to be studied is *flirting behaviour* this could be sub-divided into several categories:

Eye contact Smiling Laughing Touching Hugging

Each of these behaviours should be observable and obvious so the researcher is able to count these each time they occur. All possible categories of behaviour should be included in the list – so, for the target behaviour *flirting behaviour*, the list above might be a bit short. We have extended it in a behaviour checklist at the bottom of the facing page.

Interobserver reliability

Observational studies should be carried out by more than one researcher because *observer bias* can be a problem. A single researcher may overlook important details or only record data that fits with their *expectations*.

If data is to be collected by a pair or a group of researchers, it is important they establish **interobserver reliability**. If each observer's record is accurate then it should be the same as another observer's record.

To check interobserver reliability:

- The researcher creates categories of behaviour (a *behaviour checklist*).
- All observers should watch the same sequence of behaviour.
- Observers should compare the data they have recorded and talk about any differences. If there are just two observers correlation can be used to work out how similar the two sets of observations are.
- If the observations made are not very similar, the categories of behaviour need to be altered or rewritten altogether. It is possible that the categories were were not easy to identify or there were overlaps between the categories.

THINK LINK

There are many different types of observation. For example, in Chapter 6 von Frisch's study of honey bee communication is an example of a naturalistic observation (see pages 166–167). Kendon's study (page 170) in the same chapter is a controlled observation where participants interacting were watched through a one-way mirror.

E Evaluation

Strengths

One strength of observation studies is that they may provide a better indication of behaviour than questionniares/interviews. If a researcher asks people to report what they think and do, the answers may not be truthful. In fact sometimes people don't actually know what they would do in a particular situation! So observing behaviour may actually provide data with high *validity*.

Another strength is that observation studies look at real-life behaviour. In particular, with covert observation people shouldn't know they are being studied so they act naturally. In naturalistic observation there is no manipulation, unlike in a field experiment where there are deliberate changes. This means the data collected may have more validity than using other methods.

Weaknesses

One weakness of observation studies is that there are *ethical issues*. In order to watch or listen to natural behaviour, people shouldn't know they are being observed (covert observation). The problem is that people may object to being observed, even if they are in a public place. The researcher is invading their right to control information about themselves (i.e. their privacy).

Another weakness is that observers have expectations and this can influence what they see or hear. This is called *observer bias*. For example, let's imagine that a psychologist believes that baby boys are more aggressive than baby girls and decides to investigate this by observing the behaviour of children in a nursery. The psychologist's expectations that boys are more aggressive may mean that he (or she) is more likely to interpret the boys' behaviour as aggressive.

(This can be dealt with if more than one person makes the observations and the observations are then compared – interobserver reliability.)

Stretch and challenge

Below is a *behaviour checklist* for collecting data in an observation. The target behaviour *flirting* is broken down into observable categories. These should be tallied in the correct box every time one of these behaviours is seen in the pair being observed.

You might want to conduct an observation of pairs of people in a film using this checklist, to see whether gender has an effect on flirting behaviour.

When you have carried out your observation, have a think about the categories of behaviour. Were they clear and precise? Was the data easy to collect? Which categories would you change if you did the observation again?

Gender of pair	Eye contact	Smiling	Laughing	Touching other person	Playing with own hair	Hugging	Turning away	Other
Boy–girl								
Boy–boy								
Girl–girl								

A behaviour checklist to record flirting behaviour between pairs of students (boy–girl, boy–boy, girl–girl)

Harvey went to great lengths to conduct his covert observation.

Apply it – research

Child's play

Two psychologists wondered if girls and boys gained more enjoyment from playing with some toys more than others. They arranged for several different toys to be available in a room (e.g. gun, ball, doll, hammer, etc.). They sat in the corner of the room and observed the behaviour of four-year-old children with the toys.

Questions

1. Explain why this observation study could be described as controlled, overt and non-participant. *[1 mark + 1 mark + 1 mark]*
2. Identify **three** categories of behaviour the psychologist could have used to record each child's toy-related enjoyment. *[3 marks]*
3. The psychologists realised their study had low interobserver reliability. Explain what this means. *[2 marks]*
4. Apart from the issue of interobserver reliability, explain **one** strength and **one** weakness of this observational study. *[2 marks + 2 marks]*

Study tip

Observations are not field experiments

Students often get these two research methods confused so make sure you know the difference. Both take place in natural settings, but a field experiment is an experiment which means there is an IV and a DV.

In an observation the researcher may control aspects of the environment but cause and effect are not being investigated.

Of course observation may be used as *part* of an experiment as a way of measuring the dependent variable.

Check it

1. A psychologist wants to ensure he has inter-observer reliability in the observation study he is conducting. Outline what is meant by 'interobserver reliability' and explain how he can check it. *[5 marks]*
2. Outline **one** strength and **one** weakness of an observation study. *[4 marks]*
3. Using an example, explain what is meant by 'categories of behaviour'. *[3 marks]*

Correlations

The specification says ...

Correlation: An understanding of association between two variables and the use of scatter diagrams to show possible correlational relationships.

The strengths and weaknesses of correlations.

Computation of formulae is not required.

Many people are superstitious. They won't walk under a ladder. They think that a black cat crossing their path is a sign of good fortune. They won't leave the house on Friday the 13th for fear that something terrible might happen to them.

Superstitions are when two things that aren't really connected – like black cats and luck – become linked together in a person's mind. A famous example of superstitious thinking is the *Curse of the Bambino*.

The American baseball team Boston Red Sox were one of the most successful teams in the early years of the sport. They won every American league championship from 1914 to 1918. Then, before the start of the 1919 season, the Red Sox sold their best player to a rival team. His name was Babe Ruth (pictured above), also known as *The Bambino* – and he was one of the most talented players in the history of the sport. And so the curse began…

It would be another 86 years before the Red Sox won the American championship (known as the *World Series*). The 'curse' was lifted in 2004 and, finally, Red Sox fans could stop blaming their bad fortune on the selling of The Bambino.

Psychologists are also interested in how things are linked. However, the things they investigate tend to be more scientific than curses …

What is meant by ...

Correlation A mathematical technique used to investigate the relationship or association between two variables, called co-variables.

Scatter diagram A type of graph that represents the strength and direction of a relationship between co-variables in a correlation.

Correlations

Association between two variables

Correlations tell us how things are linked together. The things being measured are properly called *co-variables*, and we are interested in the relationship or *association* between the co-variables.

Correlations not only tell us whether two things are related, they also tell us the *strength* and *direction* of that relationship.

Co-variables

Correlations involve the analysis of *quantitative* (numerical) data. This means the two co-variables need to be represented numerically.

For instance, if we were interested in whether there was a relationship between temperature and aggression, we would need some way of measuring these in numerical form. Temperature is straightforward enough – we just record the degrees Centigrade. Aggression might be trickier. We could get people to rate how aggressive they feel on a scale (from 1 to 10), or count how many times they raise their voice during a set period of time.

For each person in our study we record what the temperature was when they were tested and a score for aggressiveness.

Scatter diagrams

Correlations are plotted on a special type of *graph* called a **scatter diagram**. One of the co-variables being measured is represented on the x-axis – the one that goes across. The other co-variable is represented on the y-axis – the one that goes up (see examples below).

For each person studied a dot is placed on the graph to represent the x and y position of the co-variables. In our example this would be the temperature and aggressiveness score for each participant.

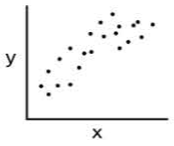
The scatter diagram means it is possible to see, at a glance, whether there is a relationship between the two variables. The closer the dots are to a diagonal line, the stronger the correlation. The direction of the line also tells us the type of correlation.

Types of correlation

The three main types of correlation are *positive*, *negative* and *zero correlation*.

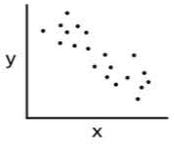
A positive correlation is when one co-variable increases as the other increases, or one co-variables decreases as the other decreases. For example, the number of people in a room and noise level may be positively correlated – as the number of people in the room increases, so does the level of noise.

The dots will be arranged from bottom left to top right.



A negative correlation is when one co-variable increases, the other decreases. For example, the number of people in a room and personal space may be negatively correlated – the more people there are in the room, the less personal space there is.

The dots will be arranged from top left to bottom right.



A zero correlation is when there is no relationship between the two co-variables. For instance, there may be no relationship between the time it takes a group of people to complete a crossword puzzle and the number of bags of crisps that the local shop sells over the course of a week.

The dots show no clear pattern.

THINK LINK

In Chapter 5 Social influence, the bystander effect is discussed which is based on a negative correlation – the more people there are that are able to help, the less likely help is to occur (see page 142).

Evaluation

Strengths

One strength of correlations is that they are a good starting point for research. Correlations tell us that two variables are related, and this may give researchers ideas for future investigations. For instance, if there does appear to be a relationship between temperature and aggression levels, a researcher might want to investigate this further – by carrying out an experiment in which temperature (independent variable) is changed to measure the effect on aggression levels (dependent variable). Correlations help identify patterns between variables which might lead to new research and theories.

Another strength of correlations is that they can be used to investigate more complex relationships than those described on the facing page. Most people's general level of alertness will rise during the day until it reaches a peak – and then it will fall steadily towards bedtime. This is called a *curvilinear relationship*. This means that correlation has many uses as a technique to understand how variables are associated.

Weaknesses

One weakness is that, even though correlations tell us that two variables are related, they do not tell us *how* those variables are related. For instance, a researcher might find a positive correlation between stress and illness (co-variables are: stress score and number of days off work). Even if we know these co-variables are positively correlated, we do not know which *causes* the other. Are high levels of stress causing illness? Or, is illness causing stress – the person is stressed about getting behind at work because they are often absent? It is not possible to show cause and effect between two co-variables, which limits the conclusions that can be drawn.

Another weakness is that important *intervening variables* may be overlooked. The apparent link between stress and illness may not be because one causes the other but because of a third factor that comes in between – hence 'intervening'. For example, if a person has a stressful job, they might sleep less, smoke and drink more, and not look after themselves properly. This might increase their chances of becoming ill. The point is that because there is no control of extraneous variables in a correlation, we can't be as certain as we would be if we conducted an experiment. This means that, in a correlation, conclusions may be drawn that are wrong.

Study tip

Correlational hypotheses are different

If you are asked to write a hypothesis for a study looking for a correlation or an association between two variables then it needs to be a correlational hypothesis. This means that you need to identify whether there is a positive or negative correlation between the covariables that are being investigated and write 'There is a positive correlation between X and Y' or 'There is a negative correlation between X and Y'.

Between the year 2000 and 2009 there was a strong positive correlation between the amount of cheese consumed each year in the United States, and the number of people who died by becoming entangled in their bed sheets.

Coincidence??

Yeah, probably.

Apply it – research

Linking variables

A team of psychologists carried out two research studies and discovered the following:

Study 1: When elderly people felt more in control of their lives, they became more active and healthier.

Study 2: As people increased their alcohol intake, their performance on a memory task decreased.

Questions

For each of these studies:

1. Identify the type of correlation and explain your choice. *[1 mark + 2 marks]*
2. Roughly sketch an appropriate graph of the correlation. Label the axes carefully and give the graph a suitable title. *[4 marks]*
3. Explain **one** strength of using a correlation in this study. *[3 marks]*
4. Explain **one** weakness of using a correlation in this study. *[3 marks]*

Stretch and challenge

You can analyse your own correlations and create your own scatter diagrams using Microsoft Excel.

First, enter your data in two columns – one column for each co-variable.

To see how your data looks on a scatter diagram click and drag your cursor from the top left to the bottom right of the two columns. From the toolbar menu select 'chart' and select 'XY scatter'. Your scatter diagram will appear.

Alter one of the values and see how this affects the relationship between the co-variables on the scatter diagram. Sometimes just a single value can make quite a difference!

Check it

1. What is meant by a 'correlation'? *[2 marks]*
2. Name the graph used to show the relationship between two variables in a correlation. *[1 mark]*
3. Explain **one** weakness of using a correlation. *[2 marks]*

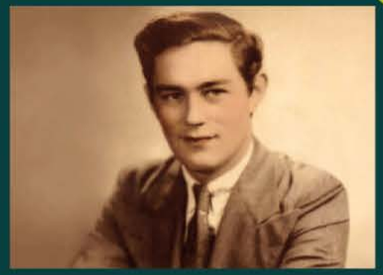
Case studies

The specification says

Designing research: Case studies.

Strengths and weaknesses of the research method and types of research for which it is suitable.

HM and his brain



One the most famous case studies in the history of psychology was Henry Mollaison, known to the world as HM until his death. We mentioned his case in Chapter 1 (page 17).

At the age of 7, Henry was involved in a bicycle accident. As a result of this, he experienced epileptic seizures for many years until these became so severe that he was referred for surgery at the age of 27.

The surgeon, William Scoville, removed a part of Henry's brain called the hippocampus (he sucked it out with a straw). The operation was successful in controlling the seizures but had a terrible side-effect – it destroyed parts of Henry's memory. Henry could remember things from the past but he was unable to form new memories. He could no longer remember new facts, faces or places he had been. He read the same magazines with no memory of having read them before. For many years he thought he was 27 but eventually started to guess his own age because he could see that he was older.

Years of observation and research revealed that, despite his amnesia, Henry was able to learn some simple motor skills. This showed brain scientists that memory for facts and events is stored in a different part of the brain than *procedural memories* (memory for motor skills).

By the time of his death in 2008 at the age of 82, Henry was considered to be the most studied, and the most important, patient in the history of brain science.

His brain was perserved for further scientific study – what are the ethics of that? In fact, what are the ethics of the case study in general? Henry was never able to provide his informed consent to any of the procedures.

You can read about Henry in the book *Memory's Ghost* by Philip Hilts (1996).

Case studies

A **case study** is an in-depth investigation of something. The *something* could be a single individual, a group, an event or an institution – like a school.

Case studies tend to involve people or events that are unusual or unexpected – such as the case of Phineas Gage (see page 200) or the St. Pauls riots (see page 148). However, this doesn't always have to be the case. Research may focus on more 'everyday' experiences too – such as how a new parent adjusts to going back to work. The study of one person in depth can provide special insights not gained from, for example, a *questionnaire* asking lots of mothers about their experiences.

A qualitative method

As a general rule, case studies collect mostly *qualitative data*, i.e. data that is not numerical and expresses people's experiences in words.

The data in a case study may be a *case history* describing past events in the individual's life or actions leading up to a particular event. Often case studies involve *interviews* with the person or friends and relatives, or in a case study of an event, interviews with people who were involved in an event. All of these will produce *qualitative data*.

Some case studies may involve experimental testing that produces *quantitative data* to see what the person can and cannot do. An important feature of the case study of HM (see left) is that researchers performed experiments on his long-term memory, to see which parts were still working. Psychological tests, such as intelligence tests, may also be used which produce a numerical score.

Longitudinal

Case studies can be conducted in a very short period of time but they tend to be *longitudinal*, in that they take place over a long period of time. The reason why case studies are longitudinal is so the researcher can see how behaviour changes over months or even years. This may involve looking back in time (a case history) or following an event over time. For example, Henry Mollasion (HM) was studied by neuroscientists for 55 years following his brain surgery.

The subject(s) of a case study is often chosen for its *uniqueness*. Such people (or animals) may be able to tell us important things about 'typical' behaviour.

What is meant by ...

Case study An in-depth investigation of a single individual, group, institution or event.

THINK LINK

We have mentioned three case studies on this spread: HM (page 17), Phineas Gage (page 200) and the St. Pauls riots (page 148). We have mentioned many individuals throughout the book, such as the Hopi Indian, Naquayouma (page 164) or Angela Cavallo (page 192) but these are not case studies because they are too narrow in their focus.

E Evaluation

Strengths

One strength of a case study is that researchers are often more open-minded when conducting a case study – they don't have a particular *hypothesis* or focused *aim*. This means that unexpected results may be produced and the researcher can gain unusual insights into the behaviour they are studying. This is a good thing for psychology in general because researchers can otherwise become quite blinkered by what they hope to find out.

Another strength of case studies is they are the best way of studying unusual forms of behaviour. Unusual behaviours are rare and therefore can't be studied using experiments requiring many people. For example, the case study of HM looked at the experience of one man who had part of his brain removed. Such research may help our understanding of 'normal' behaviour. In the case of HM the research gave insights into how memory works in everyone's brain. In this way, case studies provide important knowledge about general functioning.

Weaknesses

One weakness of case studies is that they often concern unique people or events. This means it is often not possible to generalise the results beyond the particular person or event being studied. In addition it is difficult to make before and after comparisons. For example in the case of Phineas Gage we only have anecdotal evidence of what he was like before the accident. Therefore case studies may not produce useful information.

Another weakness is that case studies are subjective. The conclusions drawn from the case study will often be based on the researcher's own 'reading' of the case. This may not necessarily be the same conclusion that another researcher would come to. This challenges the *validity* of conclusions drawn from case studies.

A Apply it – research

Deprived of sleep

A sleep psychologist studied an individual who slept only one hour every night. The psychologist conducted many structured and unstructured interviews to collect data from the participant and also interviewed his relatives. The study included experimental testing to assess his memory functioning and continued over a two-year period.

Questions

1. Explain **two** reasons why a case study was a suitable research method for the psychologist to use. *[2 marks + 2 marks]*
2. Explain **one** way in which the psychologist collected qualitative data. *[2 marks]*
3. Explain **one** strength and **one** weakness of doing this case study. *[2 marks + 2 marks]*

Q Study tip

Use underlining in the stem

Research methods questions are mainly 'application questions' with a stem or scenario, such as the one above. It pays to underline key words in the stem as these will often be needed when answering the questions.

P

When studying something like gambling behaviour we may gain different insights from the detailed study of one person's experience than, for example, a survey about gamblers' attitudes. A case study about David is described below.

Stretch and challenge

Here's another case study – this one related to addiction, a psychological problem we will study in Chapter 8.

Mark Griffiths (1993) described the case study of an individual who experienced fruit machine addiction.

'David' an 18-year-old, and his mother, were the subject of a series of interviews about David's addiction to fruit machines. David's parents explained his addiction as coming about because of problems he experienced as a teenager.

His mother described how she had followed David, suspicious of where he was spending his time. When she saw him entering an amusement arcade, she was at first relieved and assumed this was just 'harmless fun'.

Shortly after, his mother and sister began to notice money missing from the family home. As the arguments with his parents increased, the more extreme David's addiction became. David himself described the fruit machine as being like a drug that made him feel 'high'. As his addiction became more serious, the feeling of winning was no longer what drove him to continue. The lights, colour, excitement and adrenaline rush were what kept David coming back to the amusement arcade.

Case studies offer a very different way of studying addiction compared to the other methods discussed in Chapter 8. They provide useful insights into addictive behaviour which can't be provided by more general descriptions or lists of numerical data.

Check it

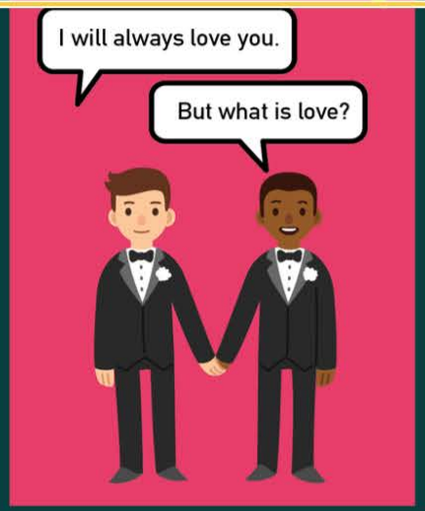
1. What is a 'case study'? *[2 marks]*
2. Explain **one** strength and **one** weakness of conducting a case study. *[4 marks]*
3. Explain what type of research a case study is most likely to be suited to and why. *[3 marks]*

Reliability and validity

The specification says...

How research should be planned, taking into consideration the reliability and/or validity of:

- **Sampling methods.**
- **Experimental designs.**
- **Quantitative and qualitative methods.**



Meet Mr Reliability and Mr Validity.

Mr Reliability is consistent. He's never late. He's in the same mood every time you meet him. He even wears the same suit. Mr Reliability doesn't change.

Mr Validity always tells the truth. He is genuine and honest. People say of Mr Validity *'what you see is what you get'* and *'he is exactly what it says on the tin'*. Mr Validity never lies.

When planning research, Mr or Ms Psychologist need to make friends with Mr Reliability and Mr Validity.

What is meant by ...

Reliability concerns the consistency of a measurement. Every time a thing is measured the result should be the same – unless the thing itself has changed.

Validity concerns whether a result is 'true'. Valid research represents something that is real. When a researcher conducts a study, they want their results to represent everyday life.

Reliability and validity

If research is going to be of any use the results can't just be a fluke. We expect that, if we repeated a study, we would get the same result. Research needs to be *reliable*. We also expect that the study actually represents what people do in their everyday lives, not just in a *laboratory*. Research needs to be *valid*.

Reliability

Reliability is a measure of consistency. If a particular measurement can be repeated, then psychologists would describe this measurement as being reliable. If a set of weighing scales is working properly, it would give the same weight measurement of the same (unchanged) object every time it was measured. This is reliability.

Psychologists don't often use weighing scales or tape measures to make their measurements. They use *questionnaires* and *observations* and case studies to 'measure' people. These methods of measurement need to be consistent otherwise they are meaningless.

Quantitative methods

Generally, methods that collect *quantitative* data tend to be the most reliable.

Experiments *Laboratory experiments* are often described as reliable because the researcher can control and *standardise* many aspects of the procedure. This means the procedure can be quite easily repeated (replicated) in exactly the same way for every participant.

There is less control in a *field experiment*. Less control of extraneous variables in the real-life setting makes replication of a field experiment more difficult.

Interviews and questionnaires A questionnaire (or *interview*) is a tool for finding out about a person's thoughts and feelings. Each time that tool is used the person should provide the same answers, i.e. the tool should be reliable. (Note that you might get different answers *if* the person has changed their views – but we wouldn't expect that if they did the questionnaire twice in a short space of time.)

It is probably easier to achieve reliability if a questionnaire is made up of closed questions, where the choice of answers is fixed, than open questions where a person might answer the question slightly differently each time.

The reliability of a questionnaire can be checked by comparing the answers a person gives on two separate occasions (called *test-retest* for obvious reasons). The answers should be the same if the questionnaire is reliable.

Observations In an observation, reliability can be about making sure the same observer would record the same observations twice – for example the observer could watch a video and compare the observations they made on each occasion.

Generally the more important form of reliability for observations is consistency between two (or more) observers, which can be checked using *interobserver reliability* (discussed on page 108).

Reliability of observations is helped by having detailed categories of behaviour that cover all the different aspects of a target behaviour. Also, by making sure that all observers are properly trained.

Qualitative methods

Qualitative methods are generally regarded as less reliable than *quantitative* methods. Methods such as *unstructured interviews* and *case studies* are difficult to repeat in exactly the same way. In an unstructured interview, for instance, interviewees are encouraged to explore a particular topic and talk in detail about their thoughts and feelings. This means that no two interviews – even with the same person – are likely to be the same. Therefore there is a lack of consistency (reliability).

It is also unlikely that even a *structured interview* with two different interviewers would be carried out in the same way. Even if the questions were the same every time, different interviewers will have a different interviewing style which may prompt quite different replies.

The data produced in interviews and case studies is also much more open to interpretation. This means that two different researchers might draw different conclusions and the conclusions would therefore lack consistency (reliability).

Study tip

Reliability or validity?

These concepts are easily confused. It is helpful to understand that reliability is part of validity. This is because, if a measure is more consistent (reliable) it is more likely to be valid as there is a greater chance of it measuring what it should be measuring (validity).

If in doubt, use validity. It is always better to comment on validity rather than reliability.

Validity

Validity refers to whether a result reflects what is really 'out there' in the real world. We briefly discussed validity on page 98 of this chapter. Basically the whole point of research design is to improve the validity of research results.

Sampling methods

Sampling methods are supposed to produce a group of participants who represent what is really 'out there' in the *target population* – but not all of them do. We have seen how some methods, such as opportunity sampling, are unlikely to produce a valid, representative *sample*. Whereas others, such as *stratified sampling*, are specially designed to do just that.

Experimental designs

Each of the three *experimental designs* can cause problems with validity. The results of a repeated measures design can be influenced by order effects. *Conditions* of the experiment should be counterbalanced as a way to address this.

When using an *independent groups design*, *participant variables* may act as an *extraneous variable*, affecting the validity of the findings. Participants should be *randomly allocated* to each *experimental condition*

Matched pairs design overcomes both problems though matching isn't perfect.

Quantitative methods

In general, the experimental method has validity issues. The more artificial the task or the setting in an experiment, the less validity there is. This is not necessarily solved by doing a field experiment (rather than a laboratory experiment). It is still possible to have artificial tasks in field experiments that do not generalise very well to everyday life.

In addition, increased issues with extraneous variables in a field experiment means we cannot be certain that the change in the *dependent variable* was due to the *independent variable*.

On the other hand, in a laboratory experiment participants are aware that their behaviour is being measured, which may reduce the naturalness.

Methods that produce numerical data, such as questionnaires and experiments, are often accused of lacking validity. This is because researchers may be investigating a very narrow aspect of human behaviour – for example, the capacity of *short-term memory* in a

a single score, and we get little or no insight into other aspects of their behaviour.

Qualitative methods

In contrast, qualitative methods such as interviews and case studies, are designed to gather the detailed information that quantitative methods lack. A case study, for instance, gives the researcher lots of insight into how the participant sees the world, their thoughts and behaviour. In this way, qualitative methods can be classed as high in validity.

Having said that, qualitative data is difficult to analyse, and is often fairly subjective. A researcher's own *expectations* are likely to influence the analysis and this reduces the validity of the conclusions drawn.

Are people with bigger heads more intelligent? That would probably be a reliable method of measuring intelligence – but is it a *valid* test of intelligence?

Stretch and challenge

It has been suggested that the circumference of a person's head could be used as a measure of intelligence. This is likely to be a fairly *reliable* measure of intelligence because adult head size is consistent from one year to the next.

You may even feel this is a *valid* measure of intelligence. After all, if you have a bigger brain then you might have more intelligence. However, research doesn't support this idea. Intelligence is not related to brain or head size. This means this measure of intelligence lacks validity.

Apply it – research

Consistent and true?

A class of GCSE students planned a number of research activities. They wanted to use:

A. Unstructured interviews to investigate people's attitudes towards mental illness.

B. An observational method to study obedience in a real-life situation.

Question

For each of these studies, explain **one** way in which the study may lack reliability. *[2 marks]*

Check it

1. Research should be planned to take into account validity.
   a) Outline **one** thing that can cause problems with the validity of an independent groups design. *[2 marks]*
   b) Explain **one** way the validity of an independent groups design could be increased. *[2 marks]*
2. Distinguish between reliability and validity. *[3 marks]*
3. Explain why qualitative methods are seen as less reliable than quantitative methods. *[3 marks]*

Types of data

The specification says...

Quantitative and qualitative data: The difference between quantitative and qualitative data.

Primary and secondary data: The difference between primary and secondary data.

Here's a word you might not have heard before – *zettabyte*. According to the business writer, Taylor Mallory Holland, 2016 was the year in which a zettabyte of data was used worldwide.

So what does a zettabyte of information look like? It's 1 billion terabytes worth of data. Or, to put it another way, 1 trillion gigabytes. According to Holland, worldwide Internet traffic has increased by five times over the past five years – and will grow by another eight times by the year 2020.

The increasing use of mobile devices is what has caused this massive explosion of data. Now with the touch of a button or swipe of a screen, data can be sent anywhere around the world at any time. Pretty mindboggling eh?

In psychology, data isn't just about the numbers. It's about the words as well.

What is meant by ...

Primary data Information that has been obtained first hand by the researcher for the purposes of a research project.

Qualitative data Information that is expressed in words and not numbers (although qualitative data can be converted to numbers for the purpose of analysis).

Quantitative data Information that can be counted, usually given in the form of numbers or tallies.

Secondary data Information that has been gathered by someone other than the researcher before the current investigation. Such second-hand data might include results from other psychological studies or official records such as government statistics.

Quantitative data

We have referred to **quantitative** and **qualitative data** and methods throughout this chapter. Some people think that quantitative data is just about numbers and qualitative data is about words and about what people think and feel – but quantitative data can involve words and it can involve data about what people think and feel.

For example, you might be asked the question 'Are you superstitious? Yes or No'. Even though the question is made up of words, the data it produces is quantitative. This is because there is a fixed set of responses and we can easily count up how many people said yes and how many said no.

Evaluation

One strength of quantitative data is that it can be quite easy to analyse. Scores can be converted to averages which can be represented on *graphs* and charts. This means it is straightforward to make comparisons between groups of participants. Statistics are open to less interpretation than words so there is less chance of bias in quantitative data.

One weakness of quantitative data is that it lacks the depth and detail of qualitative data. We learn very little about participants when their thoughts or abilities are reduced to just a single score. Quantitative data tends not to have the same *validity* that qualitative data does, i.e. it doesn't reflect how complex things are in the real world.

Qualitative data

Qualitative data starts out as data in words or pictures. This is not just a few words but lengthy paragraphs.

Such data can be turned into numbers. For example, a researcher might *interview* ten people about their early childhood experiences. The researcher could then go through the data and count how many times each person mentioned their mother, or mentioned their father. In this way numerical data would be produced.

Basically qualitative data starts out in a form that is not easily counted.

Evaluation

One strength of qualitative data is that it has more depth and detail than quantitative data. The participant or respondent is free to fully express their thoughts and feelings. The researcher gets much more insight into the participant's view of the world than is gained with quantitative data. For this reason, qualitative data is seen as having more validity than quantitative data.

One weakness of qualitative data is it is more difficult to analyse. The researcher may have pages and pages of material from an interview for instance, and it may be difficult to summarise this material and therefore difficult to draw conclusions. This means that the conclusions that are drawn may be based on the researcher's own opinions. For this reason, qualitative data can be open to bias.

Stretch and challenge

In practice, there is not always a clear cut-off point between qualitative and quantitative methods in modern psychological research. For example, if we were to interview the participants after the balls in the bucket experiment to see how they felt during the study, we would collect *both* qualitative and quantitative data.

In a *questionnaire*, there may be questions involving rating scales (quantitative data) but also open questions that ask for a much more detailed written response (qualitative data).

THINK LINK

Most studies collect both qualitative and quantitative data. For example, the results of Milgram's study of obedience (page 136) focused on the numbers of participants who obeyed (quantitative data) but Milgram also reported the verbal responses of the participants (qualitative data).

Primary data

Primary data can be qualitative or quantitative. The key thing about primary data is that it has been collected for the purpose of the research being carried out. Primary data is collected first hand by the researchers, so the data collected matches the *aims* of the study.

Primary data can be gathered by using any research method as long as the experiment, interview, etc., has been designed for the purpose of the investigation.

Evaluation

One strength of primary data is that it suits the aims of the research. Because the researcher knows the type of data they need for their investigation, they can design the method to access this information. Primary data is authentic, as it comes first-hand from the participants themselves. This means the data may be more useful.

One weakness is that it takes more time and effort to collect primary data than secondary data. Designing, say, an experiment is costly and time-consuming. The researcher must design the study, test the procedures, find some participants, analyse the data, etc. It is much easier and quicker to use data that is already tried and tested (validated).

Secondary data

Secondary data is data that is *second hand*. It is data that has been collected by someone else – so it already exists before the researcher comes to do their investigation. Secondary data may have been analysed by another researcher so statistics already exist such as the mean and graphical analyses.

A researcher may access secondary data by looking in books, journal articles or on websites. Secondary data may also include official records held by employers, the police or government statistics.

Evaluation

One strength of secondary data is it is easy to access. The time taken to collect primary data is not needed if someone has carried out a very similar study before. The data is already checked. This means the research involves very little effort and possibly much less expense.

One weakness is that the study or studies that have already been conducted may not quite fit what the researcher wants. Secondary data may be out-of-date, not quite complete or of poor quality. This means that using secondary data may waste valuable time – time that could have been spent designing a primary data investigation.

Study tip

Interpret when you conclude

When you are writing an evaluation always try to finish your point with a conclusion. Remember a conclusion is an interpretation of what you have said so far. Some useful phrases to use are 'This suggests that ...' or 'This means that ...'.

Quantity or quality? Try describing it in numbers or words.

Apply it – research

Data differences

A psychologist and his assistant worked together to investigate gambling addiction. The assistant searched the Internet and printed materials for government statistics on the number of people with addictions to various types of gambling. The psychologist carried out unstructured interviews with gambling addicts, using many open questions.

Questions

1. Identify the sources of primary and secondary data in this study. Explain your answer. *[1 mark + 1 mark + 2 marks]*
2. Explain **one** strength of using primary data in this study. *[2 marks]*
3. Identify the sources of quantitative and qualitative data in the study. Explain your answer. *[1 mark + 1 mark + 2 marks]*
4. Explain **one** strength of using qualitative data in this study. *[2 marks]*

Pre-schoolers and their devices. Is this what they mean by primary data?

Check it

1. Using quantitative data in research has been criticised. Use your knowledge of psychology to explain why the use of quantitative data has been criticised. *[5 marks]*
2. Distinguish between primary and secondary data. *[3 marks]*
3. Outline **one** strength of using primary data. Refer to secondary data in your answer. *[3 marks]*

Descriptive statistics

The specification says...

Descriptive statistics: Understand and calculate mean, median, mode and range.

A report by the Office of National Statistics in 2010 described 'Mr and Ms Average' (not to be confused with 'Mr Reliability and Mr Validity'!).

In 2010, the average (mean) age for a British man was 38, and he would go on to live for another 41 years. The mean age for a British woman however, was two years older, and she would go on to outlive Mr Average by three years.

Mr Average was 5 feet 9 inches tall and weighed nearly 14 stone. Ms Average tipped the scales at 11 stone and was 6 inches shorter than her husband.

Finally, the average age of marriage for Mr Average was 30.8 and 28.8 for Ms Average.

Statistics should be treated carefully though as they can sometimes give a misleading picture of the general population. It is a statistical fact that the average British person has slightly less than two arms. This is because the overall number of arms in the population includes people with one arm or no arms. As there aren't any people with more than two arms, this slightly reduces the overall average.

Stretch and challenge

In psychology, when working out the range, once researchers have subtracted the smallest value from the largest value, they will often add 1 (so the range on the right becomes *9* not *8*).

This is applied as a *mathematical correction*. When quantitative data is recorded in research, numbers are *rounded up*. Someone may have taken 34.6 seconds to complete a task in an experiment, but this would be recorded as 35 seconds. Adding 1 is a way of accounting for this *margin of error*.

What is meant by ...

Descriptive statistics The use of graphs, tables and summary statistics (such as averages) to identify trends and analyse sets of data.

Mean The mathematical average. Calculated by adding up all the numbers/scores in the data set and dividing by the number of scores that there are.

Median The middle value in a set of data when they have been put in order from lowest to highest.

Mode The most common value (or values) in a set of data.

Range A simple measure of the spread/dispersion in a set of data. The lowest score is subtracted from the highest score.

Descriptive statistics

Descriptive statistics express numbers in a way that gives an immediate impression of the overall pattern of results.

There are many ways to do this:

- *Spread (or dispersion)* tells us the spread out-ness of the data. The **range** is an example of this.
- *Averages* provide a typical value for a set of data. The **mean**, **median** and **mode** are all different ways to express an average.
- *Graphs* and *tables* provide a visual summary of a set of data. We can see at a glance what the differences and similarities are. Graphs are dealt with on the next spread.

Only quantitative data can be expressed using descriptive statistics – though qualitative data can be turned into numbers.

A way to represent spread

Range

The range measures *dispersion* or *spread*. This tells us whether a set of data are quite close together (e.g. everyone got almost the same score and there were few individual differences) or spread out a lot, maybe with some participants who had quite extreme scores.

The range is a simple calculation which involves arranging the data in order and then subtracting the lowest score from the highest score in the data set.

So, let's imagine there were 10 people in the audience condition for the balls in the bucket experiment and they were throwing 20 balls each, so the maximum score is 20:

9, 10, 8, 16, 9, 13, 8, 12, 15, 9

We first need to put the numbers in order:

8, 8, 9, 9, 9, 10, 12, 13, 15, 16

And then we can see that the highest score is 16, and the lowest score is 8, so the range is 16 – 8 = 8.

Evaluation One strength is that the range is very straightforward to calculate as you just subtract the lowest value from the highest value.

One weakness is the range can be distorted by an extreme score. Had one of the participants scored 1, the range would now be 15 which does not give a clear picture of the overall spread of scores. Most of the participants did not do anywhere this badly!

The full range – from biggest to smallest, and everything in between...

Ways to represent the average

Mean

The mean is the mathematical average of a set of scores. We work out the mean by adding up all the scores, and dividing the total by the number of scores that there are.

So, if there were 10 people in the audience condition for the balls in the bucket experiment, and their scores were:

9, 10, 8, 16, 9, 13, 8, 12, 15, 9

The total is 109 and there are 10 items of data, so the mean is 109/10 = 10.9

So typically participants got 10.9 balls in the bucket.

Evaluation One strength of the mean is it is the only average that includes all the scores in the final calculation. For this reason, it is the most sensitive measure as it represents the whole of the data set.

One weakness with the mean is it can be easily distorted by an extreme score. Let's say an additional participant scored one. The mean would come out as 110/11 = 10. The one score has a strong effect on the mean and it is now less representative of the overall data.

Median

The median is the middle score in a set of data. To work out the median, the scores need to be arranged from the lowest score to the highest score – and then the middle score can be identified. This is easy to do if there is an odd number of scores. If there is an even number of scores (as there is in our set of scores), this is slightly more difficult.

8, 8, 9, 9, 9, 10, 12, 13, 15, 16

Here there are two middle scores. In this situation, we add these together and divide by 2. So, the median is (9 + 10)/2 = 9.5

Evaluation One strength of the median is that it is not affected by extreme scores as they will be at either end when the scores are put in order. So, unlike in the example of the mean above, one participant scoring 1 would not affect the median very much .

One weakness of the median is that it is less sensitive than the mean. Not all of the scores are included in the final calculation. This means that a few very low values would affect the mean whereas they wouldn't affect the median – but should if the median truly reflects the 'average'.

Mode

The mode is the value that occurs the most in a set of scores – in other words, the number that is most common.

We can see that in our data set the score of 9 appears three times which is more than any other score. Therefore the mode is 9.

If another participant joined in and scored 8 we would have two modes – 8 and 9. This is *bimodal*. There could be three modes – *trimodal* – or even more – *multimodal*.

Evaluation One strength of the mode is that it is very easy to calculate (it's not really a calculation!) compared to the other two measures of average.

One weakness is that the mode can be quite unrepresentative of the set of scores overall. In our example the mode is quite a bit lower than the mean.

Study tip

Remember your workings

It is always good practice to show your working out, even if you used a calculator for a maths question. This way, even if you made an error in your calculation, you may still gain some

If you're a parent, there's no such thing as an 'average child'.

Apply it – research

Eyewitness!

Six participants witnessed a staged incident in which a 'suspect' threatened someone with a knife. Another six participants witnessed the same incident but it did not involve a knife. The participants had to correctly identify ten features of the suspect's face (e.g. blue eyes). The table shows each participant's score out of 10.

Group 1 Knife present		Group 2 Knife not present	
Participant	Score	Participant	Score
1	3	7	6
2	3	8	5
3	5	9	8
4	4	10	7
5	3	11	8
6	2	12	4

Questions

For each group of participants calculate (and show your workings):

1. The mean. *[1 mark + 1 mark]*
2. The median. *[1 mark + 1 mark]*
3. The mode. *[1 mark + 1 mark]*
4. The range. *[1 mark + 1 mark]*

Stretch and challenge

When deciding which measure of average is the most suitable, it's worth having a quick look at the scores in the data set.

Researchers generally try and use the mean when they can as it's the most sensitive measure, and it provides a good comparison of performance between *conditions*. However, if there is one extreme score (either unusually large or unusually small) this will distort the mean unrealistically – as explained on the left – so it becomes less representative.

In this situation, the median is the better option.

For some data, the mode is the only measure of central tendency that can be used. When data has been recorded in categories (see the behaviour checklist for flirting behaviour on page 108), the only way to identify the most typical category is to use the mode.

Check it

1. Briefly outline **one** weakness of using the range to calculate the spread of data. *[2 marks]*
2. Explain how to calculate the range of a set of scores. *[2 marks]*
3. With reference to the data below explain how the median is calculated. *[3 marks]*

   12, 11, 15, 10, 11, 12, 16, 10 12, 9

Interpretation and display of quantitative data

The specification says ...

Interpretation and display of quantitative data: Construct and interpret frequency tables and diagrams, bar charts, histograms and scatter diagrams for correlation.

Normal distribution: The characteristics of normal distribution.

We looked at averages on the previous spread. Scientists have calculated that, on average, women speak more words over the course of a day than men. Women rack up as many as 20,000 words in a day, whereas for men it's around 13,000.

But which words are we using the *most*? Language experts have identified the top 100 most high frequency words – that is, those that crop up most in spoken or written communication. Can you guess which they are..?

You might struggle to guess all 100, but how about the first 5? You won't be surprised to learn that high frequency words tend to be the little link words that hold sentences together. The top five are: the, and, a, said, in.

So there you go: 'the' is the most high frequency word there is. Probably true. I've already used it 11 times in this little box!

What is meant by ...

Bar chart A type of graph in which the frequency of each variable is represented by the height of the bar. The categories on the x-axis have no fixed order and there is no true zero.

Frequency table A 'table' is a systematic way of representing data so it is organised in rows and columns. A frequency table displays a record of how often an event occurs.

Histogram A type of graph where the frequency of each category of continuous data is represented by the height of the bar. In contrast to a bar chart, the data in a histogram have a true zero and a logical sequence. There are also no spaces between the bars.

Normal distribution A symmetrical spread of frequency data that forms a bell-shaped curve. The mean, median and the mode are all the same point – the highest peak.

Interpretation and display of quantitative data

On the previous spread we looked at some *descriptive statistics* – range and averages. Tables and *graphs* are also examples of descriptive statistics. They provide a visual analysis of a set of data.

Quantitative data is data in numbers and one of the strengths of quantitative data is that it can be summarised in tables and graphs so that researchers can draw conclusions fairly easily.

Scatter diagrams are one type of graph used specifically for correlations. These are discussed on page 110.

Frequency data

On this spread we focus on tables and graphs that show frequency – how often something happens. In our balls in the bucket experiment we might test 100 participants and give each participant 20 balls to throw. The data might look like this:

12, 9, 2, 9, 5, 11, 11, 4, 4, 14, 7, 3, 8, 10, 6, 14, 2,11, 8, 13, 14, 2, 9, 10, 7, 9, 3, 9, 11, 4, 9, 6, 10, 5, 10, 8, 17,10, 8, 10, 13, 7, 7, 14, 10, 9,10, 11, 11, 6, 11, 9, 11, 10,11, 7, 11, 2, 13, 2, 11, 5, 11, 8, 11, 8, 11, 11, 2, 12, 9, 12, 7, 7, 12, 11, 12, 3, 12, 5, 12, 13, 13, 6, 13, 14, 10, 4, 9, 14, 15, 8, 14, 5, 15, 9, 15, 9, 9, 11.

Looking at the data you can see that some scores are more frequent than others but it is difficult to see any pattern and difficult to draw conclusions. We could put the numbers in order, which would help, but even better we should use frequency tables and graphs to be able to 'eyeball' our data.

Frequency tables

The *frequency* of a score is the number of times the score occurs. For example, if three students score 65 marks in their psychology mock exam, then the score of 65 is said to have a frequency of 3.

The scores for the balls in a bucket experiment (above) are shown in the **frequency table** below where we have tallied each occurrence of a score – doing a tally means we can systematically go through the 100 scores above and then produce the frequency total for each score.

Score: Number of balls in the bucket (out of 20)	Tally	Frequency
2	𝍸 I	6
3	III	3
4	IIII	4
5	𝍸	5
6	IIII	4
7	𝍸 II	7
8	𝍸 II	7
9	𝍸 𝍸 III	13
10	𝍸 𝍸	10
11	𝍸 𝍸 𝍸 II	17
12	𝍸 II	7
13	𝍸 I	6
14	𝍸 II	7
15	III	3
16		0
17	I	1

You can clearly see here that the most common score was 11 out of 20 balls in the bucket, with 17 participants gaining this.

This means that 11 is the *mode* as this was the one that was most frequently occurring.

We can also calculate the *median* by finding what numbers came exactly in the middle. The 50th and 51st numbers were 10, so that is the median.

We can calculate the *mean* if we multiply each score by its frequency and add the total – which is 917. Divide this by 100 to get the mean = 9.17

The *range* is calculated by subtracting the lowest score from the highest score. The lowest score was 2 and the highest score was 17. Therefore, the range for this set of scores is 15.

E

Frequency diagrams

A frequency diagram is basically a graph that displays the frequency of data.

Histogram

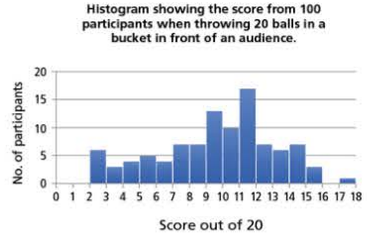
We can represent the data for the balls in a bucket experiment in the graph on the right (data is on the facing page).

- The horizontal line (x-axis) is usually used for the number categories (in this case scores).
- The vertical line (y-axis) is usually used for the frequency, i.e. the height of the bar represents how many participants achieved that score.

If we also had data from the same participants performing with no audience it would make it easy to compare the two IVs/conditions.

This kind of frequency graph is called a **histogram**.

Note that the data starts from 0 and all categories are shown.



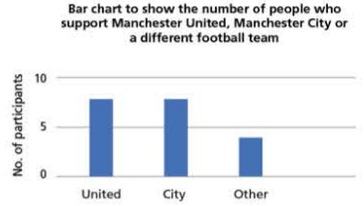
Bar chart

You might think that the graph at top right is a **bar chart**. Bar charts are similar to histograms – they both represent frequency.

However, in a bar chart the data on the x-axis is not continuous – the bars can be arranged in any order. In the histogram above it wouldn't make sense to start with 6 and then 3 and then 8 – there is a fixed order.

In a bar chart there are spaces between each bar, as in the example on the right.

Bar charts are used when data is in *categories*. For example, let's say we asked 20 football fans in Manchester to name their favourite team. The possible answers are shown on the x-axis. The frequency is shown on the y-axis.



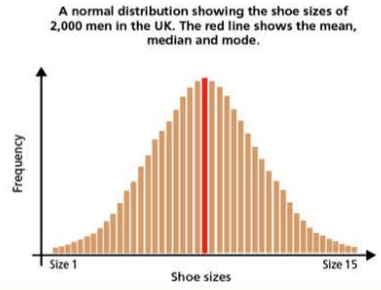
Normal distribution

A **normal distribution** is a special type of histogram. If you take certain human characteristics such as shoe size and plot the values for say 1000 people, you would find that most people have a fairly average shoe size with much fewer at the extreme ends.

A histogram of this data forms a 'bell curve'. Our data at the top right in the histogram is a bit like a normal distribution but the two 'tails' are not equal. In a perfect normal distribution the mean, the median and the mode are all at the peak of the curve – shown by the red vertical line on the picture on the right.

The two halves of the distribution – either side of the vertical line – are perfectly *symmetrical*. In other words, 50% of people are on one side and 50% of people are on the other.

Finally, in an ideal normal distribution the tails of the curve never touch the x-axis. This is because, in any population, more extreme scores are always possible.



A

Apply it – research

Displaying memory

A researcher tested the memories of a group of 50 football supporters. They had to correctly recall the scores of 10 English Premier League matches played the previous weekend. Here are the results of the study (each score is out of 10): 5, 3, 9, 6, 2, 7, 7, 5, 0, 1, 7, 4, 8, 3, 9, 8, 4, 4, 7, 7, 1, 0, 6, 2, 5, 1, 4, 8, 6, 3, 5, 7, 2, 9, 8,

Q

Questions

1. Construct a frequency table of these results. Use appropriate headings and give it a suitable title. *[4 marks]*
2. Draw a histogram of the data. Label the axes carefully and give it a suitable title. *[4 marks]*
3. Explain **one** difference between the histogram you have drawn

P

Study tip

Remember to label your graph

When drawing a graph it is vital to have a title stating what the graph is showing. The title should reflect the two variables that the study is investigating. You must also label the *x* and *y* axes.

Tables of data need a title too.

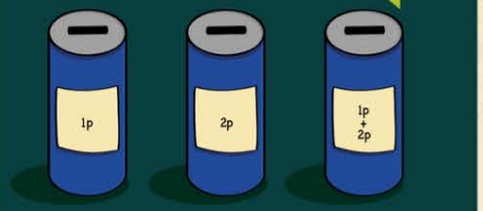
Check it

1. Identify **two** features of a normal distribution curve. *[2 marks]*
2. Outline **one** difference between a histogram and a bar chart. *[2 marks]*
3. What is meant by a 'frequency table'? *[2 marks]*

Computation

The specification says

Computation: Recognise and use expressions in decimal and standard form: use ratios, fractions and percentages, estimate results, find arithmetic means and use an appropriate number of significant figures.



Maths is all about logic so here's a logic puzzle.

Look at the jars in the picture above. One jar contains a mixture of 1p and 2p coins, one jar contains only 1p coins, the other jar contains only 2p coins. *But the labels on the jars are mixed up, so that none of the jars have their correct label.*

Your task is to work out which labels go on which jars. What is the least number of coins you would need to take out of the jars to do this? (Answer below).

Arithmetic, eh? Tricky innit? But follow the rules explained here and you can't go wrong …

Answer: You would actually only need to take one coin to work out where the labels should go - but this must come from the jar which is currently labelled '1p and 2p'.

Let's say the coin you took from the '1p and 2p' jar was a 1p piece. This means it has to be the '1p' jar. *It cannot be the '1p and 2p' jar because the jar has the wrong label on it. It cannot be the '2p' jar because you have pulled out a 1p coin.*

This leaves two jars: one labelled '2p' and the other now without a label. All the jars must change labels, so the '2p' jar must become the '1p and 2p' jar, and the remaining jar is '2p'.

Easy when you know how!

Computation

Yes, it's the arithmetic bit!

Decimals

A *decimal number* is any number written with a decimal point – so the digits to the left of the point are whole numbers (e.g. 25.0 or 213.0) and the digits to the right are parts of a whole number (e.g. 0.12 or 0.03). Each position has a value. In the number 25.0 the 2 represents two tens whereas in 213.0 the 2 represents two hundreds and in 0.12 the 2 represents two hundredths (2/100).

Fractions

The decimal 0.12 is simply another way of writing a *fraction*; 0.12 is 12 out of 100.

The important thing with fractions is that you can reduce them to a simpler form which may be easier to understand.

You look for a number that divides evenly into the top and bottom part of the fraction (the numerator and the denominator). You are looking for the *highest common factor*.

In our case of 12 out of 100 or 12/100, 4 is the highest common factor and we get the fraction 3/25.

To change any fraction to a decimal divide the denominator into the numerator.

Ratios

Ratios are really just another way to express a fraction. Consider this example: we conduct our balls in the bucket experiment with 10 participants where each participant does both the audience and alone conditions. We find that eight participants do better in the audience condition and two do better in the alone condition.

8 out of 10 did better in the audience condition = 8/10 (or 4/5).

2 out of 10 did better in the alone condition = 2/10 (or 1/5).

We have used fractions to represent the results but we could use a ratio. We could say that the ratio of audience to alone was 8:2

We can reduce this using the highest common factor to 4:1

Percentages

Yet another way to represent a fraction is a *percentage*. These are fractions out of 100.

So 12% means 12 out of 100. It is the same as 0.12 or 12/100 or 3/25.

Finding the arithmetic mean

Instructions on how to calculate the *arithmetic mean* (which is the same as the *mean*) are described on page 119. It's straightforward – just add up all the scores and divide by the number of scores there are.

Standard form

Very large numbers or very small numbers can be difficult to understand. For example:

325,000,000,000,000,000 or 0.00000000325

Standard form helps make the size of the number more understandable.

First you reduce the digits to a value between 1 and 10, so 325 becomes 3.25

In order to accurately represent our very large and very small numbers we have to give the *size* – this is expressed by saying how many times we need to multiply 3.25 by 10 to get to our original number.

325,000,000,000,000,000 is 3.25 times 17 tens which is written as 3.25×10^{17}

A

You may wonder how we worked this out.

325,000,000,000,000,000 is 325,000,000,000,000,000.0 (can you see the decimal point?). To get 3.25 we have to move the decimal point 17 times to the left.

To get 3.25 from 0.00000000325 we have to move the decimal point 9 places to the right. Therefore we write it as 3.25×10^{-9} (note the negative sign to show we went to the right).

E

Significant figures

Another way to deal with very large or very small numbers is by using *significant figures*.

When we are faced with a very large number, we might round it up – to the nearest thousand, ten thousand, hundred thousand, etc. So, for example, 216,592 to two significant figures is 220,000 (it is not 210,000 because when we remove '6592' we are over halfway to 220,000 so we round up).

When numbers come to the right of the decimal point, we can do something similar. For example, 0.000216592 to two significant figures is 0.00022 – the zeros are just place holders and not 'significant'.

Note that it is easy to confuse significant figures with decimal places. 'Decimal places' refers to the number of digits that come to the right of the decimal point. So 0.00022 is five decimal places.

Estimate results

If you are working with large-ish numbers it is quicker to do a rough calculation. Working with one significant figure is the way to go.

For example if you wanted to work out 526 times 189, you could use 500 × 200.

You don't have to be this guy to master the arithmetic!

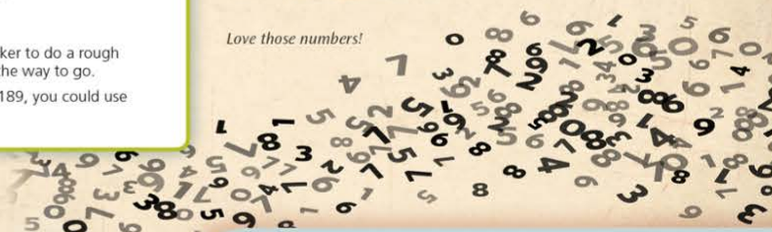
Apply it – research

Obedience by the numbers

A psychologist investigated the effect of wearing a uniform on people's willingness to obey orders. A confederate dressed as a security guard approached 72 people in the street and told them to pick up a piece of rubbish; 54 of these people obeyed. Another confederate dressed in everyday clothes issued the same order to 64 other people; 16 of these people obeyed.

Questions

1. What percentage of people ordered by the security guard obeyed? Show your workings. *[2 marks]*
2. Express this finding as a fraction in its lowest form and as a ratio also in its lowest form. *[1 mark + 1 mark]*
3. What percentage of people ordered by the 'everyday' confederate obeyed? Show your workings. *[1 mark]*
4. Express this finding as a fraction in its lowest form and as a ratio also in its lowest form. *[1 mark + 1 mark]*



Love those numbers!

Stretch and challenge

Why is maths so important in psychology?

If you continue to study psychology to a higher level, you will come across statistical tests (*We're not mentioning this to put you off!*) Statistical tests are extremely important in psychological research. What they allow researchers to do is to work out whether the results they have found are *significant*, or whether they are no more than could have occurred by *chance* (a fluke).

This is why, going back to page 95, researchers write an alternative hypothesis *and* a null hypothesis. The statistical test tells the psychologist whether they have found support for the alternative hypothesis, or whether they have found support for the null hypothesis.

Q

P

Study tip

Don't worry

Students sometimes have bad past experiences with maths and so they panic when they are asked to do calculations. The anxiety they experience then makes them struggle to think clearly, which makes the question more difficult than it actually is. It is important to bear this in mind – when you start to panic, practise relaxing (close your eyes and think of a lovely summer's day).

Check it

1. Convert the fraction ¾ into a decimal. Show your workings. *[2 marks]*
2. Using the data below, work out the arithmetic mean. Show your workings. *[3 marks]*

   13 14 12 13 13 2 12 11 15 14 0 12
3. Express 0.025913 in standard form. *[2 marks]*
4. Express 4,570,000 to two significant figures. *[1 mark]*
5. Estimate the following calculation: 58,231,526 × 321. Show your workings. *[2 marks]*

Practice questions, answers and feedback

A researcher was interested in whether gender affected the likelihood of help being given in an emergency situation. She designed a study where a confederate would stand next to a 'broken down' car at the side of a road to see whether people would stop to help. The car had the bonnet up and the hazard lights on to give the impression that it had broken down.

The researcher noted the number of times in a two-hour period other drivers would stop to help the confederate. The study was conducted on two different days so that on one day the confederate who stood by their car for two hours was a man, whilst on the second day the confederate was a woman.

Table 1: Results of the study with a confederate (a man or a woman) standing next to a broken down car

	Confederate	
	Man	Woman
Number of times a driver stopped to offer help in a two-hour period	5	15

Question 1. Identify the dependent variable in this study. *[1 mark]*

Amanda's answer

Whether the confederate standing next to the car was a man or a woman.

Amanda has identified the independent variable rather than the dependent variable.

Jay's answer

Number of times the other cars stopped in a two-hour period.

Jay has correctly identified the appropriate variable.

Question 2. Write a null hypothesis for this study. *[2 marks]*

Amanda's answer

There is no difference in the number of drivers who stop to help the male and female confederate with their broken down car.

This is a suitable null hypothesis as both variables are operationalised and both levels of the IV are mentioned.

Jay's answer

More female confederates receive help as men tend to know more about cars so could fix it themselves.

This is not a null hypothesis as Jay is stating what he thinks is going to happen. No justification is needed.

Question 3. This study was a field experiment. Identify and explain **one** strength of this type of study. *[2 marks]*

Amanda's answer

Field experiments are more controlled than lab experiments as they take place in natural settings so the researcher can stop all of the independent variables affecting the study.

Amanda has produced a muddled response as field experiments actually have less control of extraneous (not independent) variables than laboratory experiments do.

Jay's answer

Field experiments are more realistic as participants aren't aware they are being studied, which means they have good validity.

Jay has identified a relevant strength and explained clearly why it would be a positive of a field experiment.

Selema's answer

Field experiments follow standardised procedures and they are more ethical than lab studies as the participants don't have to give their consent so are happier to take part.

Selema has identified a relevant strength but her answer is muddled as she starts talking about ethical issues instead of explaining why standardised procedures are a strength.

Question 4. Draw a conclusion from the results of the study. *[2 marks]*

Amanda's answer

The results of the study show that only 5 drivers helped the male confederate whereas 15 drivers helped the female confederate.

Amanda has described the results of the study rather than drawn a suitable conclusion.

Jay's answer

Female drivers are more likely to receive help than male drivers.

Jay has produced a relevant conclusion but should refer back to the results in order to justify his claim.

Selema's answer

Female drivers are more likely to receive help than male drivers when their car breaks down. This is shown by the fact that three times as many drivers stopped to help the female confederate.

Selema has produced a response where she has drawn an accurate conclusion and then referred to the results of the study to justify her answer.

Question 5. Identify whether primary or secondary data was collected in this study. Give a reason for your answer. *[2 marks]*

Amanda's answer

Primary data was used as the researcher used the results from other studies when collecting her data.

Amanda has identified the right type of data but her reason is incorrect.

Jay's answer

Secondary data was used as the researcher used other people in their study to collect the data.

Jay has given the wrong type of data and his justification is also incorrect.

Selema's answer

Primary data was used as the researcher collected the data herself as she noted the number of drivers that stopped to help.

Selema's has produced an accurate and detailed response.

Question 6. Suggest **one** extraneous variable that could have affected the results of the study and explain how the researchers could have controlled it. *[1 mark + 2 marks]*

Amanda's answer

The time of day that the study took place was important as if the traffic was very busy then the drivers would have been less likely to stop whereas when the road is not busy people would be more likely to stop. This means that the business of the road is the thing that causes the drivers to stop rather than the gender of the person standing next to the car.

Amanda has produced an in-depth description of the extraneous variable but has not answered the second half of the question as she has not mentioned how it would be controlled.

Jay's answer

The weather could influence whether people would stop or not to offer help as if it is raining then they are less likely to. The researcher would therefore need to make sure that she did both conditions of the study on a sunny day.

Jay has suggested a relevant extraneous variable and has explained how it could be controlled.

Selema's answer

They would need to make sure that the same car was used in both conditions, for example the same make and colour.

Selema has explained how an extraneous variable would be controlled but needed to explain what the extraneous variable was first (using a different car in each condition).

Question 7. Identify the sampling method used in this study. Explain a weakness with this method. *[1 mark + 2 marks]*

Amanda's answer

Random allocation has been used as the researchers cannot control which cars drive past on the road.

Amanda has not identified a relevant sampling method or weakness.

Selema's answer

They used opportunity sampling which is an issue as the sample may not represent all drivers so can't be generalised.

Selema has identified the correct sampling method and has briefly explained a weakness but needs to add more detail to her answer, for example, by stating that 'the results cannot be used to explain the behaviour of all drivers' rather than just stating that 'the results cannot be generalised' which is a little too brief.

Question 8. Identify a suitable graph for the data in the table and briefly explain why this graph would be appropriate. *[2 marks]*

Jay's answer

A histogram would be good to use as the variable on the x-axis is continuous as it relates to the same thing (confederates).

Jay has identified the wrong type of graph as the variable on the x-axis is not continuous – 'relating to the same thing' doesn't mean continuous.

Selema's answer

A bar chart would be the best graph to use as there are separate categories on the x-axis which are a male and female confederate.

Selema has chosen the correct graph but the justification is not appropriate. Both a histogram and a bar chart have separate categories – the key issue is that the data on gender behaviour has no set order. It is not continuous.

Question 9. What is the ratio of the number of drivers who helped the female confederate in comparison to the number of drivers who helped the male confederate? Show how you calculated your answer. *[2 marks]*

Amanda's answer

$5/15 = 0.33 \times 100 = 33\%$ so there are two thirds more female drivers than male drivers

Amanda is along the right lines in her thinking but has not expressed her answer as a ratio.

Selema's answer

$15/5 = 3$ and $5/5 = 1$ so the ratio is 3:1

Selema got the correct answer and has also provided clear calculations.