Chapter 1 Experiments

Page 9 No. 1.1

1. a. The aim is to see if noise affects concentration.
   b. Loud noise and soft noise (silence).
   c. If Group B did best this suggests that noise is not good for concentration.

2. e.g.
   a. 10 mph and 50 mph.
   b. Cheap and expensive.
   c. Ripe and unripe.
   d. Young and old.

3. a. Teacher’s response.
   b. Joke and no joke.

4. a. To see if alcohol affected the goldfishes’ memory.
   b. Alcohol content of water.
   c. Water solution high in alcohol, or no alcohol.
   d. Ability to remember a maze task.

5. a. To see if performance improves just because you expect to be better.
   b. The quality/successfulness of the composer.
   c. Whether the composer was inferior or superior.
   d. How well the musicians performed.
   e. You might conclude that expectations did improve performance.
   f. Increase your expectations, believe in yourself.

6. a. To see if one side of the face is perceived as more friendly/conveys emotion better.
   b. Side of face.
   c. Composite face composed of both left sides or both right sides.
   d. Rating the face for friendliness.
   e. You might conclude that the right side of the face is better at displaying emotion.
   f. This would ensure it wasn’t just a fluke that the right side is better. Testing a number of faces would make this more certain.
1. a. Aims.
b. Hypothesis.
c. Aims.
d. Hypothesis.

2. e.g. Participants doing the task with noise show lower levels of concentration than those doing the task with no noise. (Note: both levels of the IV included.)
   a. Directional.
   b. Directional: Participants doing the task with no noise do the concentration task better than those doing the task with noise.
   Non-directional: There is a difference in concentration levels between the noise and no noise group.
   c. Directional: Goldfishes’ memory for a maze is worse after being in a water solution high in alcohol compared to having been in a water solution with no alcohol.
   Non-directional: Goldfishes’ memory for a maze is different after being in a water solution high in alcohol compared to having been in a water solution with no alcohol.
   d. Directional: A composite picture of the left side of the face is rated as more friendly than one of the right side of the face.
   Non-directional: There is a difference between ratings of a composite picture of the left side of the face compared to one of the right side of the face.

3. a. Non-directional.
b. Directional.
c. Directional.
d. Directional.
e. Directional.
f. Non-directional.

4. a. Directional: Girls watch more hours of TV a day than boys.
   Non-directional: There is a difference in the number of hours of TV watched per day by boys and girls.
   b. Directional: Teachers give higher marks on essays to more attractive students than to students who are less attractive.
   Non-directional: The essay marks awarded by teachers to attractive and unattractive students are different.
   c. Directional: People who sleep for more hours do better on class tests than those who sleep for fewer hours.
   Non-directional: People who sleep more perform differently on class tests than those who sleep less.

5. An experimental hypothesis is written specifically for an experiment whereas a research hypothesis would be a hypothesis written for any kind of study, not just an experiment.

6. The alternative hypothesis is a statement of the relationship between two variables whereas the null hypothesis is a statement of no relationship between two variables. For example, an alternative hypothesis might say ‘There is a difference between the scores from Group A and Group B’ whereas the null hypothesis would be ‘There is no difference between the scores from Group A and Group B’.

7. a. Null: There is no difference in the number of hours of TV watched per day by boys and girls.
b. Null: The essay marks awarded by teachers to attractive and unattractive students are not different.
c. Null: There is no difference in class test performance of people who have slept more than those who have slept for fewer hours.

Page 13 No. 1.3

1. a. To investigate the best ways to revise, in particular whether visual techniques work better than words.
b. If they thought visual techniques were better they might try harder to do well (or the opposite would apply if they didn’t think visual techniques were best – then they might make less effort). This might mean that their success would be due to their efforts rather than the techniques themselves.
c. It would seem to be reasonable as the task did not involve any physical or psychological harm to participants.
d. You should tell them some basic information about what is involved (e.g. tell participants they would be doing a revision task using various techniques). You could also tell them that at the end of the study they would be debriefed and given full details.
e. (1) Tell participants the true aims of the study, (2) discuss any concerns they may have.

2. a. You would collect a small group of participants (maybe just 2), and run through the task with them with special focus on the instructions and the stimulus materials. This would involve both the practice list and lists 1–4.
b. (1) You can see whether participants understand the instructions or whether these need refining so you are sure you have clear instructions, (2) you can see whether the materials are clear enough, for example the colours may not be strong enough so there was little conflict between word and colour.

3. a. The confederate would sit in the room with the participant.
b. (1) Making people take a test that might them anxious could create psychological harm, (2) participants would need to be deceived about the purpose of the confederate.

Page 15 No. 1.4

1. Independent groups.
a. Group A would do the task with no noise followed by noise, and Group B would do the task with noise first and then no noise. Or counterbalance by doing ABBA (noise, no noise, no noise, noise) and then averaging scores for both noise conditions and both no noise conditions.
b. e.g. This would be quite a complicated procedure.

2. One group did AB, the other group did BA.

3. a. One group would read lists where the colour of word and meaning of word were the same, the other group would read the conflicting list (colour of word conflicts with word meaning).
b. Participants in one group might be better at reading tasks and therefore they would do better because of this.
c. Reading ability, ability to cope with colour/word conflict.
d. They are controlled completely because the same person is used twice.

4. a. Comparing scores of two groups of people. Independent groups.
b. Comparing scores of two groups of hamsters. Independent groups.
c. Comparing scores from the same person. Repeated measures.
d. Comparing scores from matched pairs. Repeated measures.
e. Comparing scores of three groups of people. Independent groups.
f. Comparing scores from the same person. Repeated measures.

5. a. Students gain different scores on maths test A and maths test B.
b. The teacher has no idea which test might be harder so it wouldn’t make sense to state the direction of the difference.
c. When the students do the second test they may have had some practice from doing the first test. Therefore this might explain why they do better on the second test rather than because it was easier.
d. She could use counterbalancing, where one group of students did test A and then test B and the other group did test B followed by test A.

Page 17 No. 1.5

1. The intention is to be able to generalise the findings of a study to the target population. In order to do this the sample should be similar to the target population i.e. represent it.

2. It might be described as biased because there are too many older people compared to the young people (though it could be argued that there are sufficient of each type).

3. a. Random sample.
b. Opportunity sample.
c. Volunteer sample.
d. Systematic sample.
e. Stratified sample.

4. **Random**
   Strength: Potentially unbiased because everyone has an equal chance of being selected.
   Limitation: Takes considerable time to organise all the names, then select them and contact those selected.

   **Opportunity**
   Strength: Easiest method because you take the people who are immediately there.
   Limitation: Sample biased because drawn from a small target population.

   **Volunteer**
   Strength: Access to a wider variety of participants than opportunity sampling.
   Limitation: Sample biased because participants likely to be highly motivated and may respond strongly to researcher’s cues.

   **Systematic**
   Strength: Unbiased as an objective system is used.
   Limitation: Not truly unbiased.

   **Stratified**
   Strength: More representative than other methods because of the equal representation of subgroups.
   Limitation: Very time consuming to select subgroups and random selection within this.

5. The difference is that a random sample is truly unbiased whereas a systematic sample isn’t because each person in the target population does not have an equal chance of being selected.

6. a. Girls remember more words than boys.
b. Boys and girls aged between 5 and 12 years.
c. e.g. Opportunity sample. You could go to a primary school and select 15 boys and 15 girls from one class.
d. One strength is that it would be quick because, for example, you just go to the school and get your sample without having to wait for volunteers.
e. Independent groups.
f. One weakness would be that you can’t control for participant variables so it might be that the boys were more highly motivated and that’s why they would do better.

7. a. College students.
b. A random sample would give an unbiased sample whereas a volunteer sample is likely to be biased, for example, the most motivated students might volunteer.
c. You would select a few local colleges and first of all identify all those students who climb mountains. Put their names in a hat and select the number you need. Do the same with the names of the other students in the college.
d. You could put a noticeboard up in the college asking for interested students to take part, and ask especially for students who like to climb mountains.
e. The volunteer sample.

Page 19 No. 1.6

1. a. e.g. The IV might be having one other person present. The DV might be taking blood pressure.
b. People who have another person present when they get an injection have lower blood pressure than those on their own (without social support).
c. Just having a confederate might not provide ‘social support’ as the confederate would just be a stranger, so you would not be testing the value of a friend.

2. a. Morning operationalised as between 8 and 10am, afternoon operationalised as between 2 and 4pm.
b. Not drunk is having a blood alcohol level below 40 mg of alcohol per 100 ml of blood and drunk would be having more than 80 mg of alcohol per 100 ml of blood (the UK drink driving limit).
c. Ask people to rate photographs for attractiveness on a scale of 1 to 5 where 5 is very attractive. The ‘not attractive’ condition would be operationalised as an average score of less than 2 and ‘attractive’ would be an average score of more than 4.
d. A young child is under 8 years old and an older child is over 8 years old.
e. Being hungry is not having eaten for 12 hours and not hungry is having eaten within the last 2 hours.

3. a. IV is the film (helpful or neutral), and DV is how helpful the children were afterwards.
b. You could give the children a questionnaire on helpfulness and their score on the questionnaire would be the measure of helpfulness. Or observe the children’s behaviour at playtime and count how many times they did something helpful.
c. The children who watch the helpful film are more helpful afterwards than the children who watch the neutral film.
d. We expect the children who watch the helpful film would be more helpful afterwards than the children who watch the neutral film.
e. Directional (could equally have written a non-directional hypothesis).
f. e.g. Independent groups (answer could be repeated measures or matched pairs), strength would be that participants’ behaviour would not be affected by an order effect e.g. if they watch the helpful film first they might still be affected by that film when they watch the other film.
g. One weakness is that participant variables may act as an extraneous (and confounding) variable e.g. one group of participants might simply be more helpful than the others.

h. Children in the UK or in a certain town.

i. Draw names out of a hat, or take the first ones who volunteered.

4. a. By placing the names of all the children from preschool in one hat and all the names of the children in school in another hat and drawing out the number needed. (Note that you must mention two things – putting in all the names, and then drawing them out.)

b. So that everyone in the target population has an equal chance of being selected.

c. The IV is what school the child is at (preschool or ‘real’ school) and the DV is the ability to remember symbols that look like letters.

d. Children at school have a better ability to remember letter-like symbols than preschool children.

e. The researcher might choose a directional hypothesis because past research indicated that children at school are better at remembering letter-like symbols than preschool children.

f. If you only used 2 symbols, a lot of children might either get both right or score zero; you would discriminate between them better by having lots of symbols so that the children display a wide range of results.

g. It would be quicker instead of testing lots of symbols and the children might not get so bored.

Page 21 No. 1.7

1. e.g. Having extra practice on a task, or personal characteristics.

2. a. e.g. One classroom might be a more pleasant environment, and/or one teacher might encourage the students more.

b. e.g. Some people could have had a drink before the experiment and therefore their reaction times would be more affected by having two more glasses of wine, and/or some wine might be stronger.

c. e.g. The age of the students would act as a confounding variable, and/or the teacher might have been more positive with the younger students.

d. e.g. One group might have brighter students, and/or some students might not have been as motivated when taking the IQ test so that would affect their performance rather than the pill.

3. a. A situational variable, an investigator effect.

b. A participant variable, a situational variable.

c. A participant variable, an investigator effect.

d. Both participant variables.

4. a. e.g. Memory ability, age.

b. Memory because they are being tested on recall, age because older students might do the task better.

c. Testing memory on nouns and on verbs.

d. All participant names could be put in a hat and the first 10 names put in group A, the second 10 names put in group B.

e. This is to make sure that recall was not affected by the length of the word – it might be easier to recall one syllable/short words.

f. Because every child in the school does not have an equal chance of being selected.
5. Extraneous variables do not vary systematically with the independent variable (IV) and therefore do not act as an alternative IV. They are simply nuisance (extraneous) variables that muddy the waters and make it more difficult to detect a significant effect. Confounding variables do vary systematically with the IV.

Page 23 No. 1.8

1. 

<table>
<thead>
<tr>
<th>Data set</th>
<th>(a) Mean</th>
<th>(b) Median</th>
<th>(c) Mode</th>
<th>(d) Which is best and why?</th>
<th>(e) Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9.91</td>
<td>8</td>
<td>6</td>
<td>Median may be best because mean affected by extreme values at one end.</td>
<td>2–22</td>
</tr>
<tr>
<td>2</td>
<td>11.7</td>
<td>12</td>
<td>13 and 14</td>
<td>Median because unaffected by extreme value.</td>
<td>2–29</td>
</tr>
<tr>
<td>3</td>
<td>5.92</td>
<td>6</td>
<td>5 and 8</td>
<td>Mean best because takes all values into account and there are no extreme scores.</td>
<td>2–10</td>
</tr>
<tr>
<td>4</td>
<td>n/a</td>
<td>n/a</td>
<td>cat</td>
<td>Only mode can be used because nominal data (in categories)</td>
<td>n/a</td>
</tr>
</tbody>
</table>

2. Together the mean and range give you a more complete description of the data set. Two data sets may have the same mean but a different range.

3. The standard deviation is more precise because it takes all of the values into account.

4. Set B would have the smaller standard deviation because more of the numbers are closer to the mean.

5/6 For data set A the mean is 12 and the standard deviation is 8.46.
For data set B the mean is 12 and the standard deviation is 6.12.

7. a. The mean response times suggest that people were faster to respond when the victim had a cane rather than when the victim appeared drunk.
   b. The range of the scores for the ‘cane’ group is smaller, which suggests a more consistent response. The larger range suggests that some people were especially reluctant to help the apparently drunk victim.
   c. The standard deviation shows the same trend as the range – more consistent response when responding to the ‘cane’ victim – which means that all people are more likely to go to help relatively quickly whereas some people are less willing in the drunk condition.

Page 25 No. 1.9

1. Difference 1: In a bar chart there is no fixed order of items along the x-axis whereas for a histogram the order must be ascending or descending.
Difference 2: In a bar chart there is a space between each bar because the data are not continuous, whereas in a histogram the bars touch.

2. A normal distribution has a characteristic bell shape.
The highest point of the curve is where the mean, median and mode occur. They are all the same value.
The distribution of the sample is important too. 34.13% of the same lie one standard deviation below the mean and 34.13% lie one standard deviation above the mean.
3. In a skewed distribution the ‘bulge’ of the curve is not in the middle. This is because the modal score is not at the same point as the mean, whereas the mode is the same as the mean in a normal distribution. In a skewed distribution the mode is either lower (positive skew) or higher (negative skew) than the mean i.e. the most common score is not the same as the arithmetic average.

4. In a negative skew the bulk of the scores lie at the top end of the range as would be the case if we look at how people score on a very easy test. Most people would do quite well. However some people don’t and, overall, the arithmetic average will be lower the most common result.

5. A ceiling effect produces a negative skew, as explained above.

6. A floor effect produces a positive skew – when most people do very poorly on a test the modal score is likely to be very low but some higher scores will pull the mean higher than the mode.

7. The percentage would be 2 × 34.13 i.e. 68.26%.

8. Graph A: Students prefer a dog as a pet – and rats and snakes are rare amongst pets. Graph B: The older students would appear to have better memories than younger ones. Graph C: Same conclusion as for Graph B. Graph D: Same conclusion as for Graph A.

9. Graph A is meaningless. Comparing pairs of students tells us nothing about overall trends so it is impossible to draw any conclusions about which condition (organised versus random lists) produced the best scores. Whereas we can see this immediately looking at Graph B.

Chapter 2 Different kinds of experiments

Page 33 No. 2.1

1. Study A
   a. IV: Appearing blind or drunk (but not both). DV: Whether help was forthcoming, and how quickly.
   b. Task was not artificial, helping someone is a natural/everyday thing to do.
   c. Subway train is natural setting, but someone collapsing may be regarded as unnatural.
   d. High in mundane realism.
   e. No they didn’t know they were being studied.
   f. Experimenter went to them (although aspects of the situation were contrived).
   g. How many times participants had witnessed something similar, how carefully participants were watching the event.
   h. A field experiment.
   i. Strength: This study enables helping behaviour to be investigated in conditions more similar to everyday life e.g. lots of other distractions, ambiguity about the cause, other people watching you – and therefore has more everyday meaning than if studied in controlled lab conditions. Weakness: The number of variables (as listed before) means that clear conclusions are not possible because so many different things may have contributed to more or less helping.
Study B
  a. IV: Good or bad news. DV: Rating given to stranger.
  b. Task was artificial; they were taking part in a psychology study though waiting in a waiting room was more natural.
  c. Waiting room is a natural setting, though waiting for an experiment is a bit contrived.
  d. Medium in mundane realism.
  e. Yes, they knew they were being studied when they answered the questionnaire.
  f. They were brought into a contrived situation.
  g. If they guessed the purpose of the study, and/or if they didn’t like the confederate regardless of news.
  h. A lab experiment.
  i. Strength: Could control the participants’ ‘state of mind’ prior to the main part of the study i.e. hearing something good or bad. Weakness: Participants might be looking for cues about the experiment and this could affect their behaviour.

Study C
  a. IV: The model who behaved aggressively or not. DV: Whether the child behaved more aggressively when given toys to play with.
  b. Task was somewhat artificial – asking children to watch someone being aggressive with a Bobo doll.
  c. Playroom could be a natural setting but the children may have realised this was part of a study.
  d. Medium in mundane realism except that they were aware of participating and may also have felt the model’s aggressiveness was not realistic.
  e. They knew they were part of a study, though they may not have realised they were being observed at the end.
  f. They were brought into a contrived situation.
  g. How aggressive each child was anyway, and/or whether they had played with Bobo before.
  h. A lab experiment.
  i. Strength: Could compare experimental conditions to see if seeing an adult behaving aggressively had a different effect to a non-aggressive adult model. Weakness: Children may not have responded as they would in everyday life because they realised the task was contrived and were looking for cues about how to behave.

Study D
  a. IV: Peers told that they had done well or poorly on a maths test. DV: How well participants did on the maths test.
  b. Task was not artificial, doing a maths test is a usual part of school life.
  c. Classroom is a natural setting.
  d. High in mundane realism.
  e. No, they didn’t know they were being studied.
  f. The experimenter went to them.
  g. If they guessed the purpose of the study, and/or if they didn’t listen to how well the other pupils did.
  h. A field experiment.
  i. Strength: The pupils’ behaviour was studied in the context of their everyday life and therefore they would not be paying particular attention to the information from the maths test, they would respond to that in a natural manner. Weakness: There may have been other cues from the teachers about how they should behave, acting as extraneous variables.
Study E
a. IV: Increased or decreased lighting. DV: Worker productivity.
b. Task was not artificial.
c. Their workplace was their everyday setting.
d. High in mundane realism.
e. Yes, they knew they were being studied.
f. The experimenter went to them.
g. There was an important variable that was not controlled – the fact that they knew they were being studied, which led to high expectations and acted as a confounding variable.
h. A field experiment.
i. Strength: This study was able to look at the effects of lighting on workers’ behaviour in their usual setting, with all the other things that would affect their behaviour. Therefore they would behave more naturally. Weakness: One key confounding variable was not controlled (the attention they received) and this led to a lack of validity of the study.

Study F
a. IV: The counting interval (3 or 18 seconds). DV: The number of trigrams correctly recalled.
b. Task was very artificial/contrived – remembering trigrams.
c. Normal teaching room is a natural setting.
d. Low in mundane realism – the rest of the task was very artificial.
e. Yes, they knew they were being studied.
f. Neither.
g. If they guessed the purpose of the study, and/or if some of them had better memories.
h. A lab experiment.
i. Strength: The varying levels of IV meant the researchers could assess the effect this had on recall, especially as other variables could be reasonably controlled (e.g. distractions that might have diverted attention and reduced recall). Weakness: The task did not reflect the kind of things people are usually remembering so it may have led to a very poor measure of memory.

2. a. To see if the clothes a person wears affects how helpful people are.
b. Because there is an IV that is manipulated, and a DV to measure the effects of changes in the IV.
c. It has been conducted in a more natural environment, one that is not contrived/ artificial. The task is not contrived/artificial and the participants (passers-by) are not aware that they are being observed.
d. People are more helpful when approached by a confederate who is dressed in a business suit than by a confederate who is dressed in a tracksuit (directional). (Remember that the hypothesis should be about people and not participants, should be in the present tense and should be testable/operationalized.)
e. A confederate is someone who is briefed by the experimenter to behave in specific ways and in this case acts as an IV.
f. It would be better to have just one confederate for both conditions because this controls for other aspects of the person – one confederate might be friendlier than the other and this would affect how helpful participants are rather than the clothing worn by the confederate i.e. this controls a possible confounding variable.
g. Answer is likely to be opportunity sample – hard to see that it could be anything else.
h. Strength: Easiest method because you just use the first participants you can find, also most feasible if you are stopping people in the street (couldn’t use a random sample or a volunteer sample).
Weakness: Inevitably biased because the sample is drawn from a small part of the target population.

i. Independent groups.

j. Weakness: The two groups of participants may not be equivalent, for example it might just happen that one group had people who were more helpful.

k. Deception, or lack of informed consent.

l. Mode because the data is in categories (helpful or not helpful), i.e. nominal.

m. Answers as appropriate.

n. Answers as appropriate.

o. Answers as appropriate. (Remember that a conclusion should be about people and not participants, and it should be in the present tense.)

3. a. Strengths: You would be able to control extraneous variables such as what other things were going on at the same time as the confederate asked for some help. Such distractions might affect a participant’s helpfulness.

b. Weaknesses: Participants would be aware that they were being studied, which might lead them to search for cues and realise that the task involved seeing how helpful(227,265),(718,307), which might lead them to adjust their behaviour.

Page 35 No. 2.2

1. e.g. The form to fill in at the start, the panic button.

2. Cues presented by experimenters e.g. speaking more slowly would help the participants work out what the researcher expects to find.

3. a. The participants are likely to respond to the experimental cues e.g. the experimenter expects participants to do better in the morning condition and communicates cues to encourage morning participants to do better.

b. You should use standardised instructions to stop any inadvertent cues being given about which condition should be easier, or operate double blind (have an experimenter who is not the investigator).

c. Experimenter/investigator effects are the reverse side of participant effects. The experimenter/investigator knows the research aims/hypothesis or may have ideas about the research aims/hypothesis. These may inadvertently be communicated to participants – in this case the experimenter/investigator would communicate expectations to those doing the morning condition (if that was the hypothesis).

d. Ensure they don’t guess the aims by telling them that the study is actually about something else. Or counterbalance by using different experimenters and tell them that a different hypothesis is being tested.

e. Being given two tests in one day might be a cue.

f. Use an independent groups design.

4. a. Using lists of adjectives might be a cue to the participants that this is an important feature of the experiment.

b. It would be better to embed the adjectives in some more complicated statements so the aim of the experiment was less transparent.

c. The experimenter might treat the two groups of participants differently because of his/her own expectations, and encourage a more positive response from the group given the positive adjectives first.

d. This could be dealt with by ensuring that the experimenter did not know what list of adjectives each participant was given.
e. Having adjectives might lead participants to guess that the study is about liking Mr Smith (or not), and might lead participants to notice all the positive ones first. (Note that, in this case, demand characteristics are similar to participant effects.)

f. It would be better to embed the adjectives in some more complicated statements so the aim of the experiment was less transparent.

Page 37 No. 2.3

Study A
1. Issue 1: Extraneous variables.
   Issue 2: Order effects.

2. Issue 1: One extraneous variable might be time of day which would vary with the before/after lunch conditions (people being more tired in the afternoon, which would affect concentration).
   Issue 2: Being tested twice may mean participants make less effort the second time around.

3. Issue 1: Use a control group.

4. Have a control group who have nothing to eat – though hunger may act as a further extraneous variable!

Study B
1. Issue 1: Participant variables.
   Issue 2: Participant effects.

2. Issue 1: The group tested after the stress task may have higher blood pressures than the group tested before.
   Issue 2: Participants may realise that the impossible puzzle was a trick and react to this by deliberately remaining calmer.

3. Issue 1: Control the participant variable using random allocation to groups.

4. Test all participants’ blood pressure beforehand and make sure both groups have similar mean values for blood pressure.

Study C
1. Issue 1: Investigator effects.
   Issue 2: Extraneous variables.

2. Issue 1: The researcher had expectations about women being more helpful that he might have communicated to the participants, affecting the way they responded.
   Issue 2: If it was raining some of the time people might be less willing to take part in the research.

3. Issue 1: Use a different researcher to conduct the study, which means the study would be double blind.

4. The person who designed the study should get someone else to conduct the study, someone who has no expectations about which gender is more friendly.

Study D
1. Issue 1: Investigator effects.
   Issue 2: Demand characteristics.
2. Issue 1: It might be that the method of judging similarity was affected by researchers’ expectations. The final drawings may not have looked like either label but judges/investigators had to put them in one category or another. Issue 2: Participants may have guessed what the experiment was about and therefore intentionally drew something that looked like the label they were given.

3. Issue 1: Control investigator effects.

4. Have a third category, ‘doesn’t look like either label’.

Page 39 No. 2.5

1. a. internal validity
   b. population validity and generalisability/ecological validity
   c. internal validity
   d. external/ecological validity/generalisability
   e. investigator effects (or demand characteristics)
   f. ecological validity; external validity
   g. external/ecological validity/mundane realism; internal validity
   h. Mundane realism

2. Study A
   Low ecological validity because there is low mundane realism/representativeness (recall of nonsense syllables doesn’t mirror the real world).
   But high ecological validity because it has been replicated by subsequent research.

   Study B
   High ecological validity because there is high mundane realism (real world task) and also the study has been replicated by Study A.
   Low ecological validity because learning postcodes may not generalise to other memory tasks, findings may apply only to certain kinds of recall.

Page 41 No. 2.5

1. a. IV: Before and after introduction of TV. DV: Measures of pro- and anti-social behaviour.
   b. Manipulated independently of the research study, TV was introduced to the island.
   c. Other factors might lead to an increase in pro- or anti-social behaviour such as changes in the government on the island.
   d. Can’t generalise because a unique sample of people with special social norms.
   e. To see if the introduction of Western TV altered the behaviour of people on the island.
   f. The introduction of Western TV leads people to display fewer pro-social behaviours and more anti-social behaviours.
   g. The children on St Helena.
   h. Repeated measures.

2. It was a field experiment because pupils were allocated to conditions for the purposes of the study, but the experimenter went to them and the study was conducted in the children’s natural environment.

3. Study A
   Natural experiment, because it uses a naturally occurring IV (school’s choice of reading scheme).

   Study B
   Field experiment, the decision to place children in one or other maths programme (the IV) has been made by the experimenter.
Study C
This could be seen to be a field experiment (because experimenter goes to the children, and ‘natural’ school environment) or it could be, arguably, a lab experiment (because the task is contrived).

Study D
Lab experiment, because the IV (feminine or neutral ads) is manipulated and it sounds like it is not done in a natural environment.

Study E
Natural experiment, the IV (amount of TV watched) is naturally occurring.

4. Study A
a. Strength: Take advantage of existing conditions, in this case it would not be ethical to dictate the reading scheme used by a particular school so the researcher would just use schools that have elected to use the schemes already.
b. Weakness: You cannot be certain that improved reading is due to one of the reading programmes because improvement might be due to other factors in one school.
c. Internal validity relates to lack of control of extraneous variables such as the other differences between the schools. This could be dealt with by identifying other key variables (such as soci-economic status) and selecting schools matched on such variables.

Study B
a. Strength: Field experiments mean that participants may not be aware of being studied, so in this case the pupils’ lack of awareness means their performance on the test is less likely to be affected by expectations about what the researcher is trying to investigate.
b. Weakness: In field experiments there is less control over extraneous variables. In this case the two classes may differ in other ways (e.g. one class may have more able students) and this might explain the difference in their later performance rather than the maths programme.
c. Internal validity relates to lack of control of extraneous variables, as described above. This could be dealt with by assessing the classes’ maths ability (and other abilities) before the programme was introduced and taking this into account when comparing performance at the end of the study.

Study C
a. Strength: In this situation the controlled conditions of a lab means that extraneous variables, such as other distractions, could be controlled.
b. Weakness: The controlled environment might mean that everyday performance is not really tested by using word lists which are not the same as everyday uses of computers/books.
c. External validity relates to the low generalisability because the task involved word lists to test the value of computers. This could be dealt with by using a more realistic task.

Study D
a. Strength: Having the control over an IV and minimising extraneous variables means that we can be more confident about whether gender-oriented tasks have a causal influence on stereotypes.
b. Weakness: In a lab experiment participants are more likely to be affected by demand characteristics and, in this case, the use of gender-stereotyped questions may alert participants to the aims of the study.
c. Internal validity relates to demand characteristics. This could be dealt with by having lots of different questions so participants wouldn’t especially notice the gender-stereotyped ones.
**Study E**

a. Strength: In order to study aggression in a lab/field experiment the researcher would have to deliberately expose people to a situation which will make them more aggressive, which raises ethical issues. So it is better to make use of naturally occurring exposure.

b. Weakness: Causal conclusions cannot be reached because the IV (TV exposure) hasn’t been deliberately varied. It might be that more aggressive people choose to watch more TV as a calming measure.

c. Internal validity relates to lack of control – in this study things that haven’t been controlled include levels of aggression prior to TV exposure. This could be dealt with by doing a study in a situation like St Helena where people haven’t been exposed to TV before and testing their aggression levels first.

---

**Page 43 No. 2.6**

1. | The environment | The task or means of assessing the dependent variable | Whether participants were aware they were being studied | Ecological validity |
--- | --- | --- | --- | --- |
A Piliavin *et al.* (1969) | Natural (subway) | Natural – victim collapses | No | High |
B Veitch and Griffitt (1976) | Natural? Just a waiting room | Fairly natural – asked to rate a stranger | No | Reasonably high – may be alerted by being asked to rate the stranger |
C Bandura *et al.* (1961) | It was a lab but may have felt natural for the children | Natural – playing with toys | No | Reasonably high – may have felt uncomfortable because of all the different activities. Means results may not generalise well because of possible demand characteristics |
E Roethlisberger and Dickson (1939) | Natural – in their own factory | Natural – ordinary working day | Yes – aware their performance was being studied | Low because they were behaving differently as they knew they were being watched |
F Peterson and Peterson (1959) | Natural – tested in their teaching room | Not natural – memorising trigrams | Yes | Low because task doesn’t generalise well to other memory tasks |
2. e.g. Aim: to study context effects on memory.
   a. IV: Learning and recall in same environment (in lab both times) versus learning and recall in different environments (learn in lab – recall in classroom).
   b. DV: Performance on test of learned materials (general knowledge quiz).
   c. Independent groups design.
   d. (1) Participants should be randomly allocated to groups to control participant variables, (2) the two contexts should be similar in terms of distractions (noise, lighting, people present should all be the same), (3) ensure test isn’t too hard or too easy because that could create floor/ceiling effects and make it hard to analyse results.
   e. Ecological validity may be reduced by demand characteristics because students alerted to hypothesis by different testing conditions.

3. a. IV: learning and recall in same environment (in classroom both times) versus learning and recall in different environments (learn in classroom, recall in exam room).
   b.–e. Answers would be similar to b–e in (2).

4. To do this as a natural experiment you would need to find a situation where some students are tested in their classroom whereas others are tested in a different room from their classroom.
   a. IV: Learning and recall in same environment (in classroom both times) versus learning and recall in different environments (learn in classroom, recall in another place).
   b.–c. Answers would be the same as b.–c. (2).
   d. The environments might well be different in ways that would affect the DV.
   e. This would reduce ecological validity because internal validity reduced.

5. a. Likely to be an online questionnaire because people might not wish to be identified. Could be a field (naturalistic) observation. Or questionnaire filled in in a lab environment.
   b. Field experiment – test helpfulness in a natural environment by e.g. dropping something to see if it is picked up.
   c. Field experiment – observing student behaviour in class (observation used to assess DV [misbehaviour]).
   d. Could be a laboratory or field or online experiment – test reaction times in one of those locations. Reaction time testing doesn’t need fancy equipment.
   e. Questionnaire again could be conducted in a laboratory or field or online experiment.
   f. Likely to be field – where someone has watched a film with or without friends and ask them to rate the film they just saw.

Page 45 No. 2.7

1. a. This is a matter of personal opinion. It could be argued that the study was acceptable because participants were debriefed afterwards and showed no signs of lasting harm. It also could be argued that the benefits (understanding obedience) outweighed the costs to participants, so the deception is acceptable.
   On the other hand it could be argued that the participants were exposed to unacceptable levels of stress, and that the benefits have been exaggerated (more recent research suggests that Milgram’s explanations lack validity).
   b. It could be argued that he did obtain informed consent because participants were aware of what they were being asked to do, the key feature of why informed consent is necessary. They didn’t know about the deception, so in that way they weren’t informed.
   c. He did tell the participants they were free to leave and would be paid, so they had the right to withdraw. However, the nature of the prods that were used made it, in practice, very difficult to withdraw – for example, the participants were told ‘they must continue’.
The importance of this study in understanding obedience behaviour and stimulating a wealth of subsequent research could be used to argue that the distress was acceptable, especially as many of the participants seemed to have been glad to have participated.

2. **Study A**  
   a. Ethical issue: Lack of fully informed consent.  
   b. The participants were not given fully informed consent because they did not know about the humiliation of being arrested at home.

**Study B**  
 a. Ethical issue: Lack of informed consent.  
 b. Participants were not informed of the true aims – however, this is a relatively mild omission as the task did not involve any harm. It might, however, lead participants to be less trusting of psychologists in future.

**Study C**  
 a. Ethical issue: Invasion of privacy.  
 b. Individuals would regard a urinal as a place where they should not be observed – though it could be argued that men are observed by other men when in a urinal so such observation is similar to what people do in everyday life.

3. (Notice that the question does say *experimental* research), e.g. Loftus and Palmer’s study of the effect of leading questions on eyewitness testimony – participants were deceived insofar as the true nature of the study was withheld. Another popular experiment is Bandura et al.’s study of aggression in children – participants were exposed to aggressive behaviour to see if they would learn to be aggressive, which is a form of psychological harm especially as children were the participants.

4. This is one for you – ideally you would consider several pieces of research. For each study you should present arguments as to why the study was and was not acceptable. For example, in Milgram’s study you can argue that the lack of right to withdraw was unacceptable, however the benefits of the research in terms of its importance in understanding obedience outweigh the costs. Furthermore, you should explain why the lack of right to withdraw is unacceptable in this study and why understanding obedience is important.

Page 47 No. 2.8

1. **Study A** (Zimbardo et al.)  
   Participants could have been informed about the arrests or the arrests could have been eliminated. Follow ethical guidelines and debrief participants carefully.

   **Study B** (Craik and Lockhart)  
   The deception could be approved by an ethical committee.

   **Study C** (Middlemist et al.)  
   You could ask a group of similar individuals whether the observations would be ethically acceptable.

2. **Study A**  
   a. IV: Amount of sleep (5 hours or normal amount of sleep). DV: Cognitive abilities tested using a memory test.
b. e.g. Test memory using a word list.
c. e.g. Whether they sleep at all during the day.
d. Low because population validity restricted by using only students, and also only assessing cognitive ability with a memory test.
   High because ‘setting’ is natural – sleep in their own beds and at night time, as usual.
e. Field experiment because participants studied in natural environment though they are aware of being studied (therefore not a lab experiment), IV manipulated by experimenter (therefore not a natural experiment).
f. Issue 1: Distress caused by sleep deprivation.
   Issue 2: Lack of informed consent if students don’t know possible harm caused by sleep deprivation.
g. Issue 1: Ensure that students not doing anything important during the study which might be affected by sleep deprivation.
   Issue 2: Comprehensive briefing, outlining possible distress and harm.
h. Issue 1: Students may not be able to anticipate what events crop up during the period of the study.
   Issue 2: Even if you tell people that it is distressing they may not fully understand the consequences.

Study B

a. IV: Feedback from audience (smiling or disinterested). DV: Effect on speaker.
b. e.g. Count how often the speaker smiles.
c. e.g. Experience at public speaking.
d. Low because only looking at the effect of one kind of task (public speaking) so may not generalise to other situations.
   High because public speaking is a realistic task (high mundane realism/representativeness).
e. Lab experiment because participants not studied in natural environment (therefore not a field experiment, also participants aware that they are being studied).
f. Issue 1: Distress caused by speaking in public.
   Issue 2: Deception.
g. Issue 1: Ensure that participants don’t mind speaking in public.
   Issue 2: Debriefing.
h. Issue 1: Can’t accurately predict risks, other aspects of the study may be distressing.
   Issue 2: Doesn’t turn the clock back, any lowered self-esteem may be permanent.

Study C

a. IV: Time of day when test is taken (am or pm). DV: Performance on test.
b. Do a maths test, the score is the measure.
c. Demand characteristics e.g. because the pm group might suspect something because they knew the am group had done ‘something’. It might make them try harder (John Henry effect).
d. Low because only assessing performance on a maths test, which may not generalise to other situations.
   High because the children are doing something which has high mundane realism.
e. Lab experiment because it is a contrived task, participants aware that they are being studied. It might be a natural experiment if teacher used test scores from occasions when the students had done tests.
f. Issue 1: Confidentiality (test score).
   Issue 2: Lack of informed consent (students don’t know purpose).
g. Issue 1: Names shouldn’t be recorded.
   Issue 2: Debriefing.
h. Issue 1: This may protect anonymity but not confidentiality.
   Issue 2: Can’t turn the clock back, may have long-term effects.
1. a. IV: Adult present or not. DV: Children’s play.
b. Children play differently when an adult is present as compared with no adult present.
c. Divide playgroup into two sets. Children play together in groups of three. One set plays with an adult present, the other set plays on their own. The experimental groups are the ones with the adult, the control groups are those playing on their own.
d. This is a field experiment (if it is conducted in the children’s own environment). It is not a natural experiment because the experimenter controls what groups children are assigned to – but could be designed as a natural experiment where the experimenter simply records play when an adult happens to be present and when there is no adult.
e. Strength of field experiment: The children are likely to behave more like they would in everyday life than if they were playing in a contrived lab environment. Weakness of a field experiment: Lack of control over extraneous variables. In this situation such problem variables would be those that would affect their play, e.g. the toys that happen to be available to played with and what other activities are going on nearby. (If such variables were controlled, then the environment would become increasingly less ‘natural’ and more like a contrived lab environment.)
f. The experimental groups are the ones with the adult, the control groups are those playing on their own.
g. Informed consent – children can’t be told they are being observed otherwise this might affect their behaviour (but informed consent would be obtained from parents). Psychological harm – the children might be embarrassed by the experience. They should be told afterwards that they had been observed and may be shown videos of what was observed.
h. Study has representativeness because it aims to find out about children playing and the playgroup setting is a natural environment for this.
i. Difficult to generalise beyond this group of children and the particular playgroup because different playgroups may differ e.g. in their use of adult supervisors (so adults may be more or less familiar). The lack of control reduces internal validity and therefore reduces generalisability.
j. Take one group of children and watch them playing on their own and then playing with an adult present, and vice versa (for counterbalancing).
k. The experimental condition is when there is an adult, the control condition is the when they are on their own.
l. Strength: Controls for individual differences (some children more playful than others). Weakness: Order effects e.g. by the time the children get to the second condition they might be bored with playing.

2. a. IV: Drug, placebo or control. DV: Assessment of improvement.
b. People who receive the drug show greater improvement than those receiving the placebo or no drug. (There could be a second hypothesis: People who receive the placebo show greater improvement than those receiving the no drug.)
c. Recruit participants with clinical depression to take part in a drug trial. Explain that they will be randomly allocated to one of three conditions and assessed before the trial starts and again after a period of two months.
d. Not a natural experiment because of the control over allocation to groups, but could be seen as a lab or a field experiment. Probably the latter because lots of extraneous variables uncontrolled.
e. Strength: Studying people in the course of their everyday lives, so more generalisable. It would be difficult to conduct this as a lab experiment where people would have to have all aspects of their lives controlled. Perhaps the key strength relates to the fact that this is an experiment, so we can draw conclusions about causal relationships.
Weakness: Despite random allocation it is possible that the groups of participants will differ in terms of key participant variables, such as reaction to a placebo (some people may be more susceptible to placebos).

f. The experimental group are those taking the real drug. The placebo condition acts as a control for the physiological effects of the drug and the no drug condition acts as a control for the value of belief in making people improve.

g. Psychological harm – some participants may not receive an effective treatment because of the group they were allocated to.

h. The population used would represent ‘people suffering from depression’. However, some people may refuse to take part so that, for example, the study may not involve those suffering from serious depression.

i. Such representativeness would make the findings generalisable – except if certain groups of people had declined to take part or there were differences between groups studied.

j. Over a period of three months expose participants to each of the three conditions and compare improvement in each of the three conditions.

k. The experimental condition would be taking the drug, the two control conditions would be the placebo and the no drug.

l. Strength: Would eliminate participant variables such as susceptibility to placebo or levels of depression.

Weakness: Order effects would be a problem, for example if you had the drug treatment first this might lead to an improvement which might continue even when you stop having the drug. Counterbalancing could be used.

Chapter 3 Non-experimental techniques

Page 57 No. 3.1

1. e.g. Bandura et al. (Transmission of aggression), the DV was assessed by observing children playing in a room with toys. Rosenhan (Sane in insane places), the behaviour of staff and other patients in the mental hospitals was observed.

2. • Science is empirical e.g. Loftus and Palmer demonstrated that leading questions do alter what people remember.
   • Science is objective e.g. Bandura et al. systematically recorded observations by using specific categories (physical responses, verbal aggression, non-verbal aggression, mallet aggression etc.), making their observations objective.
   • Science is controlled e.g. Milgram’s study involved a very tight set of parameters (e.g. experimenter only said certain prods).
   • Scientific research can be replicated e.g. all the studies above can be repeated exactly because the procedural details are recorded.
   • Scientists construct theories e.g. Bandura constructed social learning theory and Milgram developed the agentic shift explanation for obedience.

3. e.g. ‘People are willing to obey legitimate authority when they are required to behave in an anti-social manner.’ (A statement of what he believed to be true.)

Page 59 No. 3.2

1. Question 1
   a. Closed.
   b. Quantitative data.
   c. It isn’t clear what ‘Do you diet?’ means.
d. Operationalise the term ‘diet’ e.g. ‘Do you count your calories each day?’ or ‘Do you try to avoid fatty foods?’
e. It would be easy to analyse.

**Question 2**

a. More closed than open!
b. Quantitative (range of answers is going to be limited and not related to thoughts or feelings).
c. Not very informative question. Also it’s vague, and a leading question.
d. ‘The idea of dieting is a bad idea.’ Do you strongly agree, moderately agree, agree, disagree, moderately disagree, strongly disagree?
e. Easy to answer/analyse.

**Question 3**

a. Open.
b. Qualitative.
c. Difficult to analyse.
d. Change to quantitative question by giving range of possible answers. Could leave a space for ‘other answers’ to leave it as an open question but would still be easier to analyse.
e. Means that unexpected information can be collected.

2. a. ‘Do you believe in ghosts? YES NO’
b. ‘What is your opinion of people who believe in ghosts?’
c. ‘What kind of paranormal phenomena do you believe in, for example ghosts?’
d. People want to be seen in a good light so if they thought it was foolish to believe in paranormal phenomena and they were a believer they might not answer the questions truthfully.
e. Once you have designed the questionnaire you can give it to lots of people, which means you will collect a lot of data about paranormal phenomena.
f. People may not be honest about their views about paranormal phenomena (but possibly more so than in an interview).

3. Qualitative data concerns thoughts and feelings that cannot be easily counted. Quantitative data can be about thoughts and feelings but in a form where you can count it up. When people provide qualitative data their answers are not restricted.

**Page 61 No. 3.3**

1. In a structured interview the questions are predetermined. In an unstructured interview you may start with one or two predetermined questions but subsequent questions are created in response to answers given by the respondent, i.e. there is no predetermined structure.

2. Interviews are conducted face to face and require greater skill, whereas a questionnaire is a set of structured questions in written form. A questionnaire can be handed out face to face but the respondent fills it out themselves.

3. Kohlberg’s questionnaire was structured as all the questions were decided in advance.

4. Gilligan’s questionnaire was semi-structured as some questions were predetermined but most were developed in response to answers.
5. **Kohlberg**
Closed question: Should Heinz steal the drug?
Open question: Why or why not?

**Gilligan**
Closed question: Have you ever been in a situation of moral conflict where you had to make a decision but weren’t sure what was the right thing to do?
Open question: Could you describe the situation?

6. a. Because you could ask more pertinent questions depending on the answers the respondent gives.
b. People may be less truthful in an interview when talking about sensitive issues.

7. They lead people to give certain answers rather than others and therefore bias the data collected.

8. The effect an interviewer has on the responses given by the interviewee.

9. It means respondents don’t give honest answers, which threatens the validity of the data collected.

10. By phrasing questions so that people feel they can answer honestly.

11. The validity is threatened because people have not truly expressed their answers but instead given answers that put them in a good light.

12. Confidentiality (because people would expect their answers to be kept confidential) and informed consent (to be informed of confidentiality and how the answers will be used).

**Page 63 No. 3.4**

1. a. Interviewer bias: The way the interviewer asks the questions may encourage respondents to give certain answers, answers which do not truly represent their views/experiences. Deal with this by using a structured set of questions.
   Population validity: The sample may be biased – just students from a posh school so they do homework all the time. Deal with this by using more than one school.
b. Social desirability bias: The students may well overestimate how much homework they do to ‘look better’. Deal with this by encouraging students to be honest by telling them that the average number of hours is quite low. You could also ensure answers are anonymous.
   Recall: Students may forget to record the time each day and then they make up a figure. Deal with this by using a smart phone app to make recording easier.

2. a. There is an IV (getting drug or placebo) and a DV (assessing improvement). It is a field experiment because all conditions are not controlled.
b. Participants show greatest improvement when given drug therapy rather than a placebo (directional hypothesis). OR Participants show a difference in behaviour when given drug therapy rather than a placebo (non-directional hypothesis).
c. Using a questionnaire with a symptom checklist. You could also ask some open-ended questions.
d. Issue 1: Patients may not report their symptoms accurately.
   Issue 2: Selection of participants; if only certain age groups are chosen the findings would not have representativeness.
e. Issue 1: Could ask relatives to answer a questionnaire as well to confirm symptoms.
   Issue 2: Use a quota or stratified sampling method.
3. a. It seems to test mathematical and verbal abilities – so it depends whether you feel these are key components of intelligence.
   b. Mathematical and verbal ability.
   c. Concurrent validity: Select another test of intelligence and see whether this test correlates well. Predictive validity: Correlate the test results with something like exam performance – e.g. intelligence test performance should predict exam success.

4. a. Some of the scenarios may not relate to an individual's experience so you just guess what you might do. The test might therefore be measuring how aggressive you think you are rather than actually how aggressive you are.
   b. To check whether people are being honest or whether they are giving socially acceptable answers.
   c. ‘If someone is rude to you, does it make you feel angry’ (assuming that everyone feels angry in such a situation, so they should answer yes.)
   d. Construct validity: Select aspects of being aggressive and then devise scenarios for each of the elements.

Page 65 No. 3.5

1. a. Using test–retest: Give participants the test and then repeat this a few weeks later in order to compare the two scores.
   b. Low reliability (low correlation between scores) means that people didn’t give similar answers each time, which suggests they were not expressing something enduring about themselves. Their answers were just temporary feelings.
   c. The questions should be examined to decide if some of them lack construct validity i.e. do not relate to the construct of self-esteem.
   d. Could use concurrent validity and compare the score from this scale with an existing scale for self-esteem. A high positive correlation would demonstrate the validity of the new scale.
   e. Validity can also be improved by removing items that do not relate to the construct.

2. a. It seems to be measuring dieting beliefs as claimed. This could be determined through concurrent validity (comparing with other scales on, e.g., locus of control).
   b. Yes, low reliability affects validity. Low reliability means that individuals are not giving consistent answers from one occasion to another, which means that they are not expressing something enduring about themselves. This means that their scores are meaningless i.e. lack validity.

Page 67 No. 3.6

1. In an interview people may be more willing to elaborate their answers. Also the interviewer can follow up one answer with some further questions related to the answer which would help elaboration.

2. a. Respondents may feel more able to answer honestly on a sensitive topic such as drug taking in a questionnaire rather than face to face with an interviewer (or on a telephone).
   b. An unstructured interview would give the interviewer the opportunity to develop questions on the basis of the answers given. This allows respondents to elaborate their answers and also allows the interviewer to go in unanticipated directions.
   c. Structured interviews are good for inter-interviewer reliability because each interviewer does the same thing. It might also reduce interviewer bias, for example an interviewer who was anti-drugs might develop some negative questions if it was an unstructured interview.
3. a. The headteacher could develop a questionnaire to assess how much time students spend on Facebook and what kind of things the students do on Facebook. This would enable him to develop various scores for quantity and type of Facebook use and correlate this with the students' exam performance.

b. Exam performance could be assessed by giving a numerical value to each GCSE grade achieved and adding up the score and working out a mean value.

c. Strength: Could give the questionnaire out to all students and collect the views confidentially. Weakness: Students may not be honest in their answers (social desirability bias) and, for example, underestimate their Facebook usage.

d. Students may underestimate their usage because they can’t remember how much time they have spent during, for example, the previous week. This would make their answers inaccurate and unreliable. This could be dealt with by asking students to keep a diary and record every time they use Facebook. They might have a phone app to click when they stop and start.

e. Social desirability bias would affect validity because the students would not be honest in their answers. This could be dealt with by collecting data anonymously.

f. Confidentiality is an ethical issue, which could be dealt with by collecting data anonymously. Informed consent is an issue – the students should be told how the information will be used and what kind of information will be collected.

4. a. Strength 1: Makes it easier for someone to record their data on the spot. Strength 2: People may feel more willing to provide confidential information in this way rather than in an interview.

b. Weakness 1: It means the sample is biased because only people who own a smart phone can take part, and this might be more middle class. Weakness 2: People may feel that their answers can be linked to them as the phone owner so the data would not be anonymous or confidential.

c. It’s more like a questionnaire than an interview because, even though the questions are read by someone, there is no presence of an interviewer or sense that someone is listening to you or interacting with you. There is not clear answer.

Page 69 No. 3.7

1. a. The magnitude is .39 and the sign is positive.
   b. The scattergram should show dots that are clustered fairly closely together going from bottom left to top right. The scattergrams below show +.39 correlation.
c. The significance table shows that a value equal to or .380 would be significant with 20 participants; .39 is slightly better and therefore significant.

3. a. It means that as age increases we would expect liking for spicy food to decrease. Perhaps older people find spicy food harder to digest.
   b. Strength: If you don’t find a significant correlation between age and liking spicy food then you can rule out a causal link (note contextualisation of answer).
   Weakness: Can’t demonstrate a causal relationship between age and liking spicy foods.
   c. The questionnaire (rating spicy foods) might not be reliable i.e. people have a different result if asked again two weeks later. To deal with this problem check the reliability (test–retest) and if it is low, you need to rewrite the questionnaire perhaps with some more specific questions about particular foods.
   d. Informed consent – make sure participants understand what is involved and how their data is going to be used.

4. a.
b. There is a strong positive correlation between ice cream sales and aggressive crimes.
c. Hot weather.
d. e.g.
   IV: Give some participants a large bowl of ice cream and other participants no ice cream (independent groups design).
   DV: See if they behave more aggressively when playing a game.
e. One issue might be the way ‘aggression’ has been operationalised by just looking at criminal behaviour. It might be better to have some other more everyday measure of aggression.

Page 71 No. 3.8

1. a. People who have a difficult task to do show more signs of stress than people who have a boring task.
b. Your summary might contain a note of what behaviours were more common in Person A than Person B, or a general note about behaviours common to both conditions.
c. e.g. The fact that the person knows they are being observed, or whether the person does find the task difficult/boring.
d. Observe people through a one-way mirror, ask people to assess task afterwards in terms of difficulty/boringness.
e. It’s the technique because there is an IV and DV and therefore it is an experiment.
f. Observations are naturalistic.
g. e.g. Too much data to collect, not sure what behaviours to note down.

2. Strength: They are in their natural environment (the nursery school) so their behaviour is normal.
   Weakness: If they are aware of being watched the children may alter their behaviour.

3. Strength: Could control extraneous variables better.
   Weakness: Would be a contrived environment and therefore children may not behave naturally, might respond differently to the novelty.

4. a. IV: Seeing the aggressive model or not.
    DV: Aggressiveness when playing with the toys.
b. IV: The model played aggressively with the Bobo doll.
    DV: Aggressiveness determined by scoring behaviour e.g. imitation of physical aggression, imitation of verbal aggression.
c. A specific set of toys were provided.
d. Strength: As with all participants it removes distractions (extraneous variables), children may be more easily distracted.
   Weakness: Children may not react well to a contrived situation and their behaviour would be quite different than in everyday life.
e. They had a list of behavioural categories which organised the observations.
f. Strength: It helps observers record relevant data. The children might do lots of other things which were not relevant.
   Weakness: The set of behavioural categories might exclude some behaviours that could have been relevant in the final analysis.

5. a. Sit in a corner of a library and watch students while they are studying, noting anything they do.
b. Do the same as above but control the environment in some way e.g. have students in a room with a few tables and books but try to restrict distractions that might affect their behaviour such as the presence of other students.
c. e.g. Reading book, staring into space, talking to friends, making notes, looking for a book, getting ready to work.
d. e.g. Students spend more time staring into space than reading their books. Students spend a significant amount of time talking to friends.
e. Pretend to be another student working in the library (participant observation). One-way mirror is a possibility.
f. Invasion of privacy (library is a public place therefore not invading privacy, though people may still feel you shouldn’t observe them when they don’t realise they are being observed). Informed consent.
g. It’s a method.

Page 73 No. 3.9

1. In each case the answer could arguably be event or time sampling, but some example answers are given below.
a. Time sampling. Every 10 seconds, record the behaviour of a target child.
b. Event sampling. Note down each time a child makes a particular sound or sounds.
c. Event sampling. Every time someone uses the crossing, make a note of the details.
d. Event sampling. Note down every time someone drops litter and the details, e.g. age, gender, what was dropped.
e. Follow a dog owner and use time sampling. Every 30 seconds, write down what they are doing.

2. a. e.g. (1) Test the students’ performance as a means of assessing teacher success. (2) Attention paid to teacher in class.
b. Tell students the observer is a student teacher observing the teacher’s behaviour (might still be a bit obtrusive), or use one-way mirror.
c. Use a behaviour checklist with items such as ‘teacher smiles’, ‘teacher gets angry’.
d. Strength: Observe real-world behaviour, teaching style in a natural context. Weakness: May be hard to do it without students/teachers being aware of being studied, which would affect the validity of the findings.
e. For example, obtain teacher’s informed consent (advising teacher of purpose of study), ensure no psychological harm (unsuccessful teachers may feel discouraged by the label so make sure that all teachers receive some positive feedback), deal with invasion of privacy (tell teachers when they will be observed and by whom).

3. a. A calm environment (playing soft music etc.) in a dentist’s waiting room makes patients more relaxed than when waiting in a ‘normal’ waiting room (hard chairs etc.).
b. Independent groups.
c. Strength: No order effects, if patients were in relaxing environment followed by ‘normal’ one the first environment would have an effect and also might alert them to the aims of the study. Weakness: Participant variables (patients in different waiting rooms may be more relaxed naturally), or some dentists may be more relaxing so patients anticipate what it will be like.
d. Strength: It makes the study easy because you just use the patients who happen to arrive for a dentist’s appointment. Weakness: The participants in one group might be more relaxed naturally.
e. Behaviours that indicate anxiety such as tapping feet, blinking eyes, pacing around, or behaviours that indicate little anxiety such as smiling, sitting with legs uncrossed, reading a magazine.
f. Could be either time or event (using behavioural categories). Time might be more suitable because the receptionist has other things to get on with and therefore could not watch all the time. With time sampling the receptionist might just look every 5 minutes. For event sampling the receptionist would have to be observing all the time.
g. Form would have a list of behavioural categories listed down the left and a second column for the receptionist to tick when she observed a behaviour. Behaviours might include: patient biting lip, patient smiling etc. (any signs of high or low anxiety).

Page 75 No. 3.10

1. a. It would be a good idea to see if the behaviours you have decided to observe are the only ones, or whether there are others, or to see if your checklist/method of collecting data/sampling method works.
b. Get more than two observers to watch a video of behaviours and score them. Then calculate the inter-observer reliability. Do this by correlating the observations made by each observer (e.g. the number of smiles recorded by each observer, the number of times friend touched etc.). If the correlation is less than .80 then retrain the observers.
c. (1) People behave differently in different public places, e.g. if they were in a bar, so it depends where the observations were made. (2) Depends how close the friends are, friends who are closer will behave differently from more distant acquaintances.
d. (1) Ensure that all observations are made in the same kind of place and where behaviour not affected by other factors e.g. alcohol. (2) Give the friends a questionnaire afterwards to see how close they are.
e. e.g.

<table>
<thead>
<tr>
<th>Behaviour</th>
<th>Tick each time for each person</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Person A</td>
</tr>
<tr>
<td>Smiling</td>
<td></td>
</tr>
<tr>
<td>Touching friend on arm</td>
<td></td>
</tr>
<tr>
<td>Leaning towards friend</td>
<td></td>
</tr>
</tbody>
</table>

f. Obtain informed consent afterwards, and offer right to withdraw data retrospectively.
g. People may feel they have to say yes because you have already done the study.

2. a. Yes, for both of them the inter-observer correlation is more than .80 and therefore it is close.
b. Strength: The behavioural checklist is tried and tested (demonstrated as reliable and valid).
Weakness: Some categories need more explanation, some behaviours may be missing from list.

3. a. • Work out a checklist of behaviours to record.
• Decide on a method of sampling the behaviours – choose event sampling.
• Decide who to observe, use volunteer sample.
• Decide where to do the observations, invite participants to be interviewed and tell some of them they have to lie. Neither interviewer nor observer would know if they had been told to lie or not.
• Consider ethical issues e.g. informed consent.
• Decide how to make the observations – on paper, use a video camera, or an audio tape.
b. This study is an experiment because there is an IV (the two conditions – participants lying or not) and a DV (judgement of whether participant was lying or not). It is a lab experiment because it is conducted in an artificial, contrived setting.
c. Naturalistic observation means that people do not change their behaviour because they are just behaving as normally (though if they knew they were being observed they might), a lab experiment allows you to control the IV and other extraneous variables.
   b. Calculate inter-observer reliability and train observers if low.
   To calculate inter-observer reliability, collect scores from two or more observers, e.g. how many times they observed a particular event, and then do a correlation. To train observers, watch a video of behaviours and discuss operational definitions.
   c. Event sampling – each time a girl passes by note the length of time the man spent watching the girl, also tick items of clothing worn by the girl.
   d. e.g. The checklist didn’t provide observers with all kinds of clothing in order to determine whether the person is casually or smartly dressed.
   e. It’s an experiment because there is an IV (girls dressed casually or smartly) and a DV (time men spend watching). It is a natural experiment if the IV is not controlled by the researchers. The IV could be controlled (use same girl dressed in two different outfits) which would make it a field or even lab experiment (depends on how much control on the environment).

Page 77 No. 3.11

1. A content analysis involves making observations of behaviour even though the observations are indirect (i.e. the behaviours have been previously recorded).

2. The observers’ expectations might lead them to ignore certain information that is inconsistent with their expectations.

3. a. Item 1: Girls dressed in male clothing or vice versa.
   Item 2: Boys doing male/female tasks.
   Item 3: Old people shown doing activities that a young person wouldn’t do.
   b. Item 1: List of male and female clothing e.g. skirts, jeans.
   Item 2: List of male and female tasks e.g. washing up, repairing a car.
   Item 3: List of old and young activities e.g. gardening, listening to iPod.
   c. Count up instances in each category.
   d. Might overlook certain occurrences, deal with it by having more than one observer and compare outcomes. Check inter-observer reliability, if it is low then train observers to be more reliable.
   e. Categories might be unclear so observers record observations in wrong categories (e.g. is a kilt a skirt?). Deal with this by piloting the behavioural categories to see what is missing or ambiguous.
   f. In each category record some examples of what was observed.

4. a. e.g. Political comments, reference to racial groups, betrayal, love.
   b. Analyse a variety of songs from the 1950s and 1970s so you have a score in each category e.g. political comments. Consider how different the scores are and list those categories where there is a large difference.

5. a. Produce a checklist with e.g. number of buttons on Santa’s jacket, height of image, number of lines used to draw hat.
   b. Produce data in each category and add up. The higher the score, the greater the complexity.

Page 79 No. 3.12

1. Strengths: Addiction is more difficult to study on a large scale because of the unique characteristics of individual addicts and the lack of suitable individuals. Doing a case study permits special insights into the causes and experiences of addiction, overlooked by statistical studies.
Weaknesses: The uniqueness of each case means you may get a biased view of what addiction is like. Case studies often involve retrospective recall so the addict and their family/friends may misrepresent what actually happened.

2. a. The rioters.
   b. Observations of their behaviour, interviews, personality tests.
   c. Could ask rioter to describe what they did, to identify the kind of shops and people they targeted, to identify people they did not attack, to consider in retrospect what they should have done. The data is qualitative and to analyse it you would need to develop behavioural categories e.g. of the type of people attacked and not attacked, and then consider patterns in this data.
   d. Confidentiality – deal with it by recording responses anonymously.

3. a. You could interview some patients about their experiences during recovery e.g. how many visitors they had, the nature of their relationships with medical professionals. You could assess personality and measure physiological systems e.g. immune system activity.
   b. Interview the patients and use a personality test to assess personality.
   c. Strength: It’s a more holistic approach because you can study all the different factors that may be important, e.g. social support, physiological and psychological hardiness, past experiences etc.
      Weakness: Such analyses are very time consuming and therefore expensive.
   d. Any retrospective recall could be unreliable because of selective recall. You can deal with this by interviewing other family members to confirm what is recalled.
   e. The hospital might select one factor, e.g. social support, and compare recovery rates of patients with high and low social support. This would be a natural experiment as the IV (high/low social support) varies naturally.
   f. Strength: It would provide evidence of suggested causality. A natural experiment cannot demonstrate causality but if no relationship is found it shows that social support is not related to recovery.
      Weakness: Important intervening variables may be overlooked, e.g. educated people may have more social support, and education level may be the causal factor.

4. a. Strength 1: Case studies provide rich in-depth data. In the case study of Clive Wearing we find out what impact this has had on his everyday life and the kinds of things he does remember (e.g. playing piano) and what he doesn’t remember.
      Strength 2: Can be used to investigate rare behaviour. In the case of HM the specific effects of the destruction of an area of his brain could be studied.
   b. Weakness 1: Each case has unique characteristics so it may be difficult to make generalisations. Even though we knew what area of HM’s brain was damaged we don’t know how good his memory was before the operation so subsequent behaviour might be different if he had had an especially good or poor memory.
      Weakness 2: Ethical issues such as confidentiality and informed consent. Patients such as HM cannot give informed consent because they don’t remember anything and therefore they can’t be informed. The intensive study involved is an invasion of their privacy and they may have little choice about being studied.

5. Arguments against using real names: Such names are discovered anyway so there is little point trying to hide them. In cases where relatives can give consent it is acceptable.
   Arguments against using real names: Patients such as HM or Genie are not in a position to give informed consent but is it right that someone else consents on their behalf? You could argue that actual names are not relevant to understanding the case.
6. Information about Little Hans’s early childhood was recorded over a number of years, including extensive conversations and drawings. Little Albert was involved in a brief investigation to see whether he would acquire fear when conditioned. This was studied with a number of different stimuli but the study did not involve rich details of the experience. (More recently, however, psychologists have endeavoured to find more detail about this case.)

7. In situations where a behaviour is rare it is not possible to study a larger group of individuals so it is better to focus in depth on a few individuals.

8. Such programmes or articles do not use rigorous methods. Psychologists, for example, are trained in how to conduct interviews objectively and how to analyse qualitative data. Media reports are likely to be biased by the reporters’ subjective views.

Page 81 No. 3.13

1. a. Weakness 1: The data was collected from people in many different countries by native researchers. There may have been differences in the way the interviewers asked the questions. This could be dealt with by using the same researchers (but then they would be foreigners and might not have understood local expressions).
   b. Weakness 2: People may not have taken the task seriously and just written down any answer. This could be dealt with by telling people why the research is so important.
   c. Research collaborators were unaware of the hypothesis and therefore this would reduce interviewer bias.

2. a. The IV was age which varies naturally (arguably this is actually a quasi-experiment rather than a natural experiment because age has not been changed).
   b. It is cross-sectional because the same person is not being followed over time; instead the effects of age are studied by looking at different people at different ages, a cross-section of the population at a single moment in time.
   c. Strength: It is quicker to do it this way rather than waiting until the children aged 5 become 6, 7 and then 8 years old. The study can be done all at once.
   d. Weakness: Participant variables may act as extraneous variables e.g. some of the children may be more intelligent and therefore more able to conserve.
   e. As a longitudinal study you would start with a group of 5 year olds and then test them every year for the next three years.

3. a. Age.
   b. Cross-sectional.
   c. The older people grew up at a time when diets were less nutritious so individuals in this cohort are smaller in size and their intelligence may also have been affected. Therefore differences between the two groups may be due to this rather than to one group being older.

4. a. The IV is being adopted or returned to your natural home (both of which are ‘naturally occurring’ i.e. they have not been specifically manipulated by the researcher).
   b. Longitudinal.
   c. This cohort of children started life in the 1960s when approaches to child rearing were different to what they are now. Children today might not be affected in the same way by early privation.
   d. Strength: The alternative would be to identify a group of children now who had been adopted but you wouldn’t know what they were like when adopted. There may have been differences in the groups from this start, so the longitudinal method means you can observe the children before the IV is varied.
Weakness: Studies like this take a long time and therefore money. Attrition can be a problem (and in fact was a problem because the researchers wanted to test the children a further time but could not contact enough of the participants).

Page 83 No. 3.14

1. a. Strength: This technique enables high-quality images that show the structures of the brain and can provide detail sufficient to compare the brains of people with schizophrenia to people without schizophrenia.
   Weakness: Using this technique exposes people to radiation which is not desirable.

   b. Strength: A detailed picture of the brain can be provided without the risk of radiation. This is appropriate for a study of memory because participants have no personal gain from this study.
   Weakness: MRI scans are unpleasant (uncomfortable and lengthy) which may be something people are not willing to submit to.

   c. Strength: As Raine et al. were looking at aggressive behaviour rather than structural detail PET scans would have the advantage of showing the brain in action.
   Weakness: The downside is less precision than an MRI scan, which might be important if the researchers wanted to know about relative size of parts of the brain in relation to the effects of aggression.

2. a. An effect size of 5% means that, for people with mild depression + taking antidepressants meant 5 out of 100 people did better than those not taking antidepressants.

   b. The findings suggest that antidepressants don’t offer that much benefit for people with mild depression but the effect is stronger the more severe the depression.

3. a. The findings mean that the cognitive interview produced better recall than standard interviewing. Specifically, people interviewed with the cognitive method remembered 34% more correct information than those interviewed with the standard interview.

   b. Strength: Effect size gives a measure of the size of the difference instead of just providing the mean scores for standard and cognitive interviews. This is important to gauge just the importance of one method. Effect size comes from using meta-analysis, which means the conclusion is not just based on one study which would be less reliable/valid.

   Weakness: The studies used in this meta-analysis may not be comparable e.g. the methods used in either type of interview could be different – one type of cognitive interview may differ from another type of cognitive interview. This makes the conclusions meaningless.

Page 85 No. 3.15

1. NHA are good models because they have much physiology in common with humans. They are not good models because their behaviour is not influenced by social and cognitive factors in the same way that human behaviour is.

2. This is a personal opinion – sentience is difficult to use because the view is now that all animals have sentience.

3. Speciesism refers to the idea that the reason for discriminating is based solely on being a member of a particular species rather than any other criteria, and thus is a prejudice.

4. This is a matter of personal opinion. I would take a utilitarian view – that using NHA provides important benefits for humans and therefore is justifiable if well regulated.
5. a. The Animals Act (1986) specifies that research with animals must only be conducted at licensed premises by licensed researchers on licensed projects.
   b. Guidelines are produced by organisations such as the BPS. These suggest how researchers should treat animals, they should consider how the animals are housed to minimise distress and should reduce the number that are used.
   c. The 3Rs suggests that the way forward is to reduce the number of animals used, refine the procedures that are used so fewer animals are needed, and replace animals with other methods such as computer simulation.
   d. This is a personal opinion – I think it has helped as research using animals has reduced. However, Kilkenny’s report suggests that many researchers are flouting the rules and not being straightforward in reporting what they are doing.

6. It doesn’t make it acceptable but does mean there are no choices if we want to continue to use drugs for mental disorders.

**Chapter 4 Quantitative and qualitative methods**

**Page 93 No. 4.1**

1. a. Quantitative: Viewing figures, ratings, number of episodes, number of main characters. 
   Qualitative: Genre (e.g. comedy, drama, soap), kind of audience (children, young adults etc.).
   b. Quantitative: Population, density, area (sq. kilometres), height above sea level, latitude/longitude.
   Qualitative: Main industries, main political party, region, photographs.
   c. Quantitative: e.g. Olympics: number of events, audience figures, number of GB medals.
   Qualitative: Most popular sports, opinions about opening ceremony, descriptions of main buildings.

2. a. Strengths: Easier to compare, could be represented on a graph.
   Weaknesses: Loses out on important details by presenting numerical data.
   b. Strengths: Gives richer, more meaningful information than just statistics/numbers.
   Numbers are abstract representations.
   Weaknesses: Can’t compare as easily, can’t see differences at a glance (as you would on a graph), difficult to make systematic analyses.

3. a. Rosenhan (Sane in insane places). Quantitative data included the length of hospitalisation for the pseudopatients, the number of patients judged to be pseudopatients, number of psychiatrists who stopped to talk to pseudopatients.
   b. Strengths: (1) Means it is possible to calculate statistics such as the mean, which makes it easier to make comparisons and draw conclusions, (2) quantitative data may be easier at a glance to understand trends in data (e.g. looking at a graph).
   Weaknesses: (1) Fails to represent the experience of the individuals, (2) some statistics can be easily distorted by outliers.
   c. Pseudopatients kept a diary describing their experience on the ward.
   d. Strengths: (1) Gives useful insights into topic being studied (life on the ward and why patients continued to be judged as abnormal, (2) represents the real experience of people.
   Weaknesses: (1) Can’t make comparisons between pseudopatients at a glance, (2) may distort our perception because a limited number of people were studied and they may be unique.
4. a. Loftus and Palmer DV: Speed estimates given for each of the verbs (e.g. participants given the word ‘smashed’ estimated a speed of 40.8 compared to those given the word ‘contacted’ who estimated 31.8).
   
   b. You can’t actually measure anything qualitatively! In Bandura’s Bobo study the DV (aggression levels in the children) was assessed using observations of their behaviour. This was then used to produce mean aggression scores (quantitative).

5. a. IV: Male or female. DV: Moral orientation (care or justice or a mixture of the two).
   
   b. Participants were interviewed and their answers analysed qualitatively. Each element was judged as care or justice.
   
   c. The number of care or justice decisions were counted up to work out whether, on balance, an individual was more care- or justice-oriented.
   
   d. Women use more care judgements than justice judgements when commenting on moral decisions, whereas men use more justice judgements than care judgements.

   
   b. Children’s drawings.
   
   c. You can’t count up a drawing without identifying separate elements. It isn’t easily reduced to numbers. For example, use of shading can’t be easily counted.
   
   d. The researchers could have counted the number of lines, the size of the drawings, the individual elements in the drawings.

Page 95 No. 4.2

1. Karlsson et al.
   
   a. Perspectives on euthanasia.
   
   b. In dying cancer patients, especially related to autonomy i.e. whether patients felt in control/independent.

2. Lebowitz et al.
   
   a. OCD children.
   
   b. Looking at parents’ views of coercive and disruptive behaviours.

3. Inman et al.
   
   a. Intercultural marriages.
   
   b. The experience of being in such a marriage for Asian Indian and white couples.

4. Helweg-Larsen et al.
   
   a. Attitudes about smoking.
   
   b. Looking at perceptions of risk, looking at attitudes in different cultures.

5. Löfgren-Mårtenson and Månsson
   
   a. Pornography.
   
   b. The thoughts of young men and women on pornography, gender and sexuality, their viewing habits.

6. Ownsworth et al.
   
   a. Experiences of having a brain tumour.
   
   b. The personal and social processes in adjusting.

2. a. Look at dreams in young children, or at dreams in students before exams.
   
   b. Graffiti in toilets, or relating to sexual orientation.
   
   c. Tattoos on women, or related to science.
   
   d. Drawings of parents, or by children in different cultures.
   
   e. How smoking started, what people enjoy about smoking.

   
   Helweg-Larsen et al.: In-depth interviews.
   
   Löfgren-Mårtenson and Månsson: Interviews and focus groups.
   
   Ownsworth et al.: In-depth semi-structured interviews.
b. Inman *et al.*: Ten highly educated couples.
Helweg-Larsen *et al.*: 15 smokers from Denmark and 15 from USA.
Löfgren-Mårtenson and Månsson: 36 women and 37 men aged 14–20 from Sweden (22 interviewed individually and 51 in 7 focus groups).
Ownsworth *et al.*: 18 participants, 9 had benign tumours and 8 had malignant tumours, also 15 family caregivers.

c. As appropriate for the studies you looked up.

4. e.g. Furniture, PowerPoint slide shows, buildings, games.

5. a. Culture (smoking-lenient or smoking-prohibitive).
b. Representative country selected: Denmark or USA.
c. Danes minimised the risks more and rejected moralised opinions, they didn’t change how much they smoked.
d. e.g. The way people dealt with the risks depended on whether their culture was smoking-lenient or smoking-prohibitive.

**Page 97 No. 4.3**

1. You can identify various categories and then count up instances that would be placed in each category.

2. Qualitative data can be analysed using bottom-up methods by starting from the data collected and identifying meaningful units within the data. Each unit is coded and may be given more than one code. These units can then be combined into larger categories or themes. A final report would use these categories to describe the data collected and give examples of items in each category, such as quotes from people interviewed. Top-down analysis starts with preexisting categories and uses these to group items in the data collected. The final report can again present examples of items in each category. (101 words)

**Page 99 No. 4.4**

1. Example: taking exams.
   a. Students taking exams, parents of students taking exams, people marking their answers, teachers’ expectations about individual students’ performance.
   b. Interview the different people, analyse their exam answers.
   c. When analysing interview scripts, get several people to do this and compare their analysis.
   d. Compare qualitative analyses with other research methods e.g. compare data collected from students with physiological assessments of students before and during exams to see if experience of stress is similar to measured stress levels. Could also show analysis to participants and ask them if they feel it represents what they feel/experience.
   e. Use a variety of different methods, some more controlled than others – so some studies might be rather artificial but high in internal validity whereas others are less artificial but low in internal validity. The aim is to see if the results all point to the same conclusion, therefore validating each other.

2. Subjectivity is considering experience from the perspective of the person experiencing it, which is bound to be biased by that person’s perspective. Objectivity aims to be unbiased and in some way represent everyone’s perspective.
3. Scientific research should not be affected by a researcher’s expectations or any other beliefs because, for example, the outcome is likely to be that a researcher’s expectations will affect what data they collect. Scientists are trying to see the world as it is rather than the world as one individual sees it.

4. Qualitative research argues that there is no single reality and what matters is the different individual perspectives, so by its very nature the qualitative approach is subjective. This subjectivity is acknowledged as part of the research process.

5. In quantitative research, investigators aim to minimise subjectivity, but this may just be an illusion because it is not possible to remove bias and investigator and participant effects. So quantitative research masquerades as being objective when it isn’t.

6. It is a means of acknowledging the effect of a researcher’s values and thoughts, and recognising the effects these have on data collected.

7. Triangulation is the use of a number of different research methods to investigate a topic and use each of the findings from these studies to produce a combined conclusion. For example, you could investigate the experience of divorce through interviews but also observe interactions between family members and conduct experiments measuring attachment relationships between family members and also the effects on other relationships.

8. e.g. Obedience research includes Milgram’s studies, the research by Hofling et al. in a more natural environment and also historical examples from the Second World War. Together these help us understand human obedience.

**Page 101 No. 4.5**

1. **Graph 1**
   a. Histogram.
   b. (1) IQ peaks in late adulthood, (2) there is a decline in IQ in old age.

2. **Graph 2**
   a. Bar chart.
   b. (1) Women generally are considerably more likely than men to seek resources from a prospective partner, (2) men are not very interested in a woman’s resources.

3. **Graph 3**
   a. Scattergram.
   b. (1) There is a positive correlation between age and drug usage when driving, (2) young people (under 30) do not use drugs while driving.

4. **Graph 4**
   a. Histogram.
   b. (1) Racial violence is low when economic conditions are poor or good, (2) racial violence increases when economic conditions improve but only up to a point when it decreases again.

5. **Graph 5**
   a. Scattergram.
   b. (1) There is no correlation between final test score and time spent doing the test, (2) it appears that spending more time on a test does not lead to a better test score.

6. **Graph 6**
   a. Line graph.
   b. (1) When children are very young, parents believe they have a high IQ, (2) parents’ estimates drop off sharply through childhood.
Graph 7
a. Scattergram.
b. (1) There is a negative correlation between IQ and religion, (2) people who have high IQs are less likely to believe that religion is important.

Graph 8
a. Bar chart.
b. (1) Combined therapies are most effective in reducing anxiety, (2) placebo is at least as good as psychological or drug therapies when they are used individually.

2. a. Minutes of sleep in one night and minutes spent doing exercise on the previous day.
b. Strength: More convenient to get participants to do it themselves, presumably in their own homes.
   Weakness: Participants were asked to make the record themselves and might have been inaccurate, for example they would start the record when they went to bed but may not have gone to sleep for a while. It was initially measured in hours and then represented in minutes, would be better to have been measured in minutes to gain more detail.
c. Strength: Again this is a convenient way to collect data by getting participants to do it themselves.
   Weakness: Participants may not be clear about what counts as exercise, so their records would not be comparable.
d. Could wear a pedometer to measure walking activity during the day. One strength of this is that it is a simple quantitative measure. One weakness is that we are only measuring one kind of exercise.
e. Participants could describe what their sleep had been like e.g. Did they fall asleep straight away? Did they sleep well? Participants could describe the kind of activities as well as the amount e.g. vigorous versus fairly relaxed activity, and also how typical it was.

3. a. DV: Breathing rate.
b. Strength: Can be easily measured by placing hand on a person’s chest.
   Weakness: Does it really represent awareness?
c. Could ask people to rate their awareness, could measure brain waves. Evaluation – rating is a subjective measurement and may not be accurate, brain waves is an objective measure which is a strength.
d. IV: No caffeine and then caffeine.
e. Give participants cup of exactly 400 ml coffee so caffeine amounts are the same.

Chapter 5 Inferential statistics

Page 109 No. 5.1
1. a. Certain = 100% 1.00
   b. Likely = 75% 0.75
   c. Even chance = 50% 0.50
   d. Unlikely = 25% 0.25
   e. Impossible = 0% 0.0

2. a. There is no difference between men and women in terms of their knowledge of psychology.
   b. Women are better than men at psychology.
3. A null hypothesis is a statement of no relationship or difference, there is nothing going on. The alternative possibility is that there is a relationship, there is something going on. This is the alternative hypothesis.

4. a. There is no difference between the memory scores.  
b. Directional.

5. There is no relationship between amount of sleep at night and amount of exercise taken during the day.


7. The first states that the probability level is less than or equal to 0.05 (or 5%) whereas the second says that the probability level is exactly 0.05.

8. The chance of obtaining a particular set of results was less than 5% if the null hypothesis were true.

9. If the investigation is considering an important issue such as the use of a drug we would want to be more certain, or when replicating an existing study.

**Page 111 No. 5.2**

1. The numerical value at which significance is shown, the probability (e.g. 0.05) at which the null hypothesis has been rejected or which has been agreed as the level to determine if the null hypothesis can be rejected.

2. The observed or calculated value.

3. \( \rho \).


5. (1) Number of participants in each condition, (2) whether hypothesis is directional or non-directional, (3) significance level selected, (4) whether the critical value has to be more or less than the observed value for significance to be shown.

6. It is used for a directional hypothesis.

7. a. Ordinal.  
b. Interval.  
c. Nominal.  
d. Interval.  
e. Ordinal.  
f. Interval.

8. Nominal: Take the different names of different stables and present frequency data for each stable.  
Ordinal: For each stable say how many horses had come first, second, etc.  
Interval: Give the time each horse took to run 100 metres.

9. a. 48 is less than 53, so significance is shown.  
b. 0.01 (1%) is not in the table for a two-tailed test but 0.02 is, so using that critical value of 65, 75 is not less than 65 so significance is not shown.
c. 5% is not shown but at 10% (0.10) the critical value is 67 so the result is significant at the 10% level. At 2% (0.02) the critical value is 58, so 60 is not less than that, so the result is not significant at the 2% level.

10. a. IV: Alcohol or no alcohol. DV: Reaction time.
   b. There is no difference in reaction times when people have imbibed alcohol or haven’t.
   c. There is a difference in reaction times when people have imbibed alcohol or haven’t.
   d. Two-tailed.
   e. Interval (would be measured in seconds).
   f. Nominal – just group people as very fast, fast, medium, etc. Or you could rank participants and then it would be ordinal.
   g. One group of participants may happen to have faster reaction times.
   h. To determine whether any difference found between the mean reaction times of both groups is larger than we would expect by chance (i.e. if the null hypothesis were true).
   i. 5% is the usual level because it is not too lenient or too stringent.

Page 113 No. 5.3

1. This is a matter of personal opinion. I think a Type 1 error is worse – putting an innocent man in prison – but it would depend on the seriousness of the crime.

2. The Type 1 error would be being diagnosed with cancer when you are healthy.

<table>
<thead>
<tr>
<th>Truth</th>
<th>Has cancer</th>
<th>Does not have cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosed with cancer</td>
<td>True positive</td>
<td>False positive</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Type 1 error</td>
</tr>
<tr>
<td>Diagnosed with no cancer</td>
<td>False negative</td>
<td>True negative</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Type 2 error</td>
</tr>
</tbody>
</table>

3. They both involve accepting a mistaken hypothesis. A Type 1 error occurs when a null hypothesis that is true is mistakenly rejected whereas the Type 2 error is when a null hypothesis that is false is mistakenly accepted.

4. The likelihood of a Type 1 error is 5%, a 5% chance that a null hypothesis that was true has been erroneously rejected.

5. a. Alternative: Reaction times are slower after drinking coffee than before.
   Null: There is no difference in reaction times before and after drinking coffee.
   Procedure: Test reaction time, then give people some coffee and test reaction time again.
   Inferential test: Wilcoxon because this requires a test of difference for repeated measures.
   b. Alternative: Reaction time is correlated with age.
   Null: There is no relationship between reaction time and age.
   Procedure: Test reaction time in people of different ages.
   Inferential test: Spearman, because a test of correlation is needed.
   c. Alternative: Nurses have higher stress levels than doctors.
   Null: There is no difference in the stress levels of doctors and nurses.
   Procedure: Give doctors and nurses a test that assesses how stressed they are, e.g. could use galvanic skin response which sees how active their autonomic nervous system is.
   Inferential test: Mann-Whitney because this requires a test of difference for independent groups.
d. Alternative: Revision technique 1 produces better test scores than revision technique 2.
   Null: There is no difference in test scores between the two revision techniques.
   Procedure: Teach students the revision technique they are to use, give them some material to learn and then test them, comparing their test results.
   Inferential test: Wilcoxon because this requires a test of difference for repeated measures (matched pairs counts as a repeated measure because the participants are related through the matching).

e. Alternative: There is a difference between men and women in terms of whether they own a pet.
   Null: There is no difference between men and women in terms of whether they own a pet.
   Procedure: Ask men and women to tell you if they own a pet or not.
   Inferential test: Chi-squared because this is nominal data (answer is yes or no).

6. a. Students obey male teachers more than female teachers.
   b. Student follows direct verbal order from teacher, student doesn’t follow direct order, student listens to teacher when teacher is talking, student doesn’t listen.
   c. As this is observational you could use event sampling, focusing on one student at a time.
   d. Count up instances of obedience and lack of obedience, calculate percentage of obedient behaviour out of all behaviour. Calculate mean percentages for male and female teachers which can be shown in a bar chart. Data can be analysed with Mann-Whitney as a test of difference between male and female teachers (independent groups).

7. a. Open: What are your reasons for smoking?
   Closed: How many cigarettes do you smoke a day on average?
   b. Open: Group answers into categories and count how many in each category, or just give some of the most common answers.
   Closed: Can show in a bar chart, or give the mean answer.
   c. Strength: People may be more willing to answer a questionnaire truthfully. Also would be quicker to give out to hundreds of participants.
   Weakness: Participants may be less willing to elaborate their answers to open questions because they have to write it out.
   d. For the closed question you might correlate number of cigarettes with age, use Spearman’s test. Or you might relate some of the answers to why people smoke to the number of cigarettes smoked. This is categorical data (the reasons for smoking) and therefore you use chi-squared.

Page 115 No. 5.4

1. a. The ratio between the index/ring finger, and numeracy skills.
   b. Testosterone (male hormone).
   c. Informed consent – make sure participants understand what will be assessed and how their data will be used.
   Confidentiality – numeracy scores should be confidential.
   d. Problems – People may give consent without truly understanding.
   e. Scattergram:
The scattergram suggests a moderate negative correlation. (In fact the calculation shows it is actually quite a strong one as it is significant.)

2. a. Personality test score is correlated with athletic performance (measured by bench press).
b. There is no relationship between personality test score and athletic performance.
c. Spearman’s, because a test of correlation is required and the data are ordinal or better.
d. A two-tailed test is required because the hypothesis is non-directional. The table only gives 10% and 2% values for two-tailed tests. N = 10, the values are .564 and .648 respectively. The observed value must be greater than these values for significance to be shown.
e. A positive correlation of .32.
f. This value is not greater than either critical value and is therefore not significant.
g. For a one-tailed test this would still not be significant.
h. Strength: If significance is not shown this indicates that you can exclude a causal relationship.
Weakness: If significance is shown this shows a relationship but does not show cause and effect relationship, intervening variables may be involved.
i. Men do better at bench press than women so this would act as a participant variable and could act as an extraneous variable if not controlled.
j. Could use a random sample of all the students in a school. Number every pupil in the school and use a random number table to select the required sample.
k. Validity could be demonstrated by comparing scores with scores on another personality test (concurrent validity). Reliability could be demonstrated by testing participants twice (test-retest) to see if there is a good correlation.

Page 117 No. 5.5
1. a. Contingency table:

<table>
<thead>
<tr>
<th></th>
<th>Old</th>
<th>Young</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sleep more than 8 hours</td>
<td>11</td>
<td>11</td>
<td>22</td>
</tr>
<tr>
<td>Sleep less than 8 hours</td>
<td>25</td>
<td>33</td>
<td>58</td>
</tr>
<tr>
<td>Totals</td>
<td>36</td>
<td>44</td>
<td>80</td>
</tr>
</tbody>
</table>

b. Alternative non-directional hypothesis: There is an association between age and number of hours slept. Null hypothesis: There is no association between age and number of hours slept.
c. For a one-tailed test, and $df = 1$ the critical value would be 2.71. The observed value of 3.02 is greater and therefore significant.
However, for a two-tailed test, the critical value is 3.84. The observed value 3.02 is less than this and therefore not significant.
d. In part c it is clear that, for a one-tailed test, the result is significant and therefore the null hypothesis is rejected and alternative hypothesis accepted. For a two-tailed test the result is not significant and therefore we accept the null hypothesis.

**Page 119 No. 5.6**

1. **a.** Alternative: People do better on a memory test in the morning than when tested in the afternoon.
   Null: There is no difference between memory test performance when tested in the morning or afternoon.

   **b.** A test of difference is required (difference between morning and afternoon tests), repeated measures design (participants take part in both conditions), data is ordinal or better (test scores are interval data).

   **c.** 30.

   **d.** Volunteer sampling.
   Strength: Can access a wide variety of participants not just those who happen to be there, and also get participants who are keen.
   Weakness: Volunteers may be untypical of the wider population because of their motivation to be good participants i.e. they may be more likely to try to please the experimenter and search for cues about the hypothesis which would affect their performance.

   **e.** If it was the same memory test the students would do better the second time because of practice.

   **f.** Give the two tests to a different group of participants and see if they do better on one. It would be important to counterbalance this so some participants did Test A first and some did Test B first. If they all did Test A first then they may do better on Test B because of a practice effect.

   **g.** Match pairs of participants in terms of a prior memory test. Then place one member of each pair in the morning group and the other member of the pair in the afternoon group. Compare partners’ scores.

   **h.** Could use ABBA – give participants four tests over four days: morning afternoon afternoon morning.

**Page 121 No. 5.7**

1. | High bridge group | Low bridge group |
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>7.9</td>
</tr>
<tr>
<td>Median</td>
<td>8</td>
</tr>
<tr>
<td>Range</td>
<td>5–10</td>
</tr>
</tbody>
</table>

Graph should show mean and/or median – not individual scores.

2. **a.** IV: Time spent running on the spot (2 minutes or 15 seconds). DV: Rating woman’s attractiveness.

   **b.** Alternative: Men who are more aroused (run for 2 minutes) rate women as more attractive than do men who are less aroused (run for 15 seconds).
   Null: There is no difference in the attractiveness ratings given by the men in the high or low arousal group.

   **c.** Directional.
d. For a directional hypothesis (one-tailed test) the critical value of $U$ is 33, this is less than the observed value of 40 and therefore is significant. So you would reject the null hypothesis. For a non-directional hypothesis (two-tailed test) the critical value of $U$ is 28, this is less than the observed value of 40 and therefore is significant. So you would reject the null hypothesis.

e. It is likely to be an opportunity sample of just interviewing the first men who cross the bridge. Allocation to groups is not an issue as the researcher would either be standing on the low or high bridge – though this might be problematic because only men not afraid of heights would select the high bridge (and be less aroused). So you might approach people in the park and ask if they would meet the interviewer on the high or low bridge. Allocate first person to high bridge and second to low bridge etc., or draw conditions out of a hat to randomise.

f. This would reduce extraneous participant variables in the woman’s appearance and behaviour and in the way the female experimenter asked the questions, so each participant saw exactly the same thing.

3. a. If memory was being assessed, you would test memory of all participants and use this for matching. Could also match on GCSE exam performance as this is likely to be related to memory. Then place one member of each pair in the morning group and one in the afternoon group. Compare partners’ scores.

b. The participants in each group are related so a repeated measures design is used and the data is at least ordinal, therefore you could use Wilcoxon. (In fact the data – test scores – could be regarded as interval so you could also use a related $t$ test (discussed on page 126).

c. Using a Wilcoxon test, for 20 participants and a two-tailed test and significance level of 10% the critical value is 60. Which means the results would be significant at 10% but not 5% so they would accept the null hypothesis.

Page 123 No. 5.8

1. a. Problem 1: Some people might not actually know the famous names or their gender.
   Problem 2: Participants didn’t really listen to the names.

b. Problem 1: Afterwards check which names the participants knew and exclude the participant if they didn’t know all the famous names, or thought one of the non-famous names was someone they knew (therefore the name would be ‘more available’).
   Problem 2: Ask participants to write names down as they are read out or tick on a list.

c. Looking for a test of difference, repeated measures (both items of data come from the same person) and the data are nominal (categories).

d. Two possibilities shown below – should be fully labelled.

![Bar Chart]

- Judged male majority
- Judged female majority

e. The value of 2 (or less) would be significant at the 5% level.

2. a. Alternative: People select similar cards to the sender more often than chance (chance level = 5 out of 25).
   Null: There is no difference between performance and chance level.

b. Directional.

c. Put an advertisement up asking for participants.
d. Random (using all people in your school and selecting names from registers using random number table). This method is unbiased whereas volunteer sampling produces people who are more motivated and may try harder to please the investigator.

e. Looking for a test of difference and data is nominal.

f. The observed value of 4 would be significant (the critical value is 4 and the observed value is equal to this) and therefore you would reject the null hypothesis.

g. If $S$ is 4 and there were 16 participants, then 12 of the signs would be plus and 4 would be minus.

3. There are 5 minus signs (the less frequent sign), so $S = 5$, with 12 participants and a one-tailed hypothesis at the 5% level the critical value is 2. The observed value is greater and so the null hypothesis must be accepted (as it was on page 118).

**Page 125 No. 5.9**

1. a. Difference, repeated measures, ordinal (i.e. non-parametric) = Wilcoxon.
   b. Difference, matched pairs = related measures, interval, normal etc. (i.e. parametric) = related $t$ test.
   c. Difference, independent groups, interval, normal etc. (i.e. parametric) = independent $t$ test.
   d. Correlation, ordinal data (i.e. non-parametric) = Spearman’s.
   e. Difference, independent groups, nominal = sign test.
   f. Association, nominal data = chi-squared.
   g. Correlation, ordinal data (i.e. non-parametric) = Spearman’s.

2. **Study 1**
   a. Scores on personality tests = plastic interval scale, also normal etc. so parametric.
   b. Correlation between two variables (two test scores) = Pearson’s.
   c. Two-tailed.
   d. 5% as this is a good compromise between too lenient or too stringent, which would be likely to result in Type 1 or 2 errors (respectively). The study does not concern something of major importance so 5% is OK. (We want a stringent level for something of major importance so would use 1% but then risk making a Type 1 error – rejecting a hypothesis that was in fact true. Too lenient e.g. 10% means we might make a Type 2 error and accept a hypothesis which was in fact false.)
   e. For 20 participants, two-tailed 10% would be .360 and 2% would be .423.
   f. The observed value must be greater so it would be significant at the 2% level.
   g. Scattergram would be from bottom left to top right with reasonable closeness, as below (shows $+.45$).

   ![Scattergram](image)

**Study 2**

a. Data is interval, normal etc. so parametric.

b. Related data and correlation, so Pearson’s.

c. Two-tailed test (though since a negative correlation was expected a one-tailed test might have been more suitable).

d. See above, 5%.

e. For 15 participants, two-tailed at 10% would be .412 and at 2% would be .482.

f. The observed value must be greater so it would be significant at the 2% level. (Ignore the negative sign which just confirms that it is a negative correlation.)
g. Scattergram would be dots from top left to bottom right, reasonably closely together to show −.521, as below (shows −.54).

Page 127 No. 5.10

1. a. The expectation stated does not suggest a direction, so a non-directional hypothesis.
   b. The data (recalling word lists) is interval, normal etc., so a parametric test of difference with repeated measures (same participant tested twice), so related t test.
   c. For two-tailed test, at 10% the critical value is 1.782. For significance at 10% to be shown the observed value must be equal to or greater than this value. At 2% the value is 2.179.

2. a. It will now be directional.
   b. The data is interval and also we would expect memory to be normally distributed.
   c. For an independent t test (independent groups) \( df = 10 + 10 – 2 = 18 \).
   d. For a one-tailed test at 5%, the critical value for \( df = 20 \) would be 1.725. The value of 1.931 is greater than this so it would be significant.

3. a. The expectation stated means a directional hypothesis.
   b. A test of difference (looking at difference between morning and afternoon), interval, normal etc. data so parametric, related data so related t test.
   c. One-tailed test, 5%, \( df = 20 – 2 = 18 \), means a critical value 1.734. Ignoring the minus sign the observed value is greater than the critical value.
   d. You would reject the null hypothesis.

Chapter 6 Designing and reporting your own research

Page 135 No. 6.1

1. e.g.
   • Give name of organisation, researcher and address in case participants wish to contact you later.
   • Details of procedures so participants are in a position to give informed consent, include duration and how data will be used e.g: The aims of this study are to investigate verbal processing. You will be given a list of 30 words and asked simple questions about them. This should take about 10 minutes in total. Your performance will not be used in any way to judge you and data will be kept confidential. At the end you will be fully debriefed about this research study and the results will be shared with you.
   • Details of pay: You will be paid £5 for participating.
   • Right to withdraw: At any time you should feel free to ask questions and if you feel uncomfortable you are free to leave and will still be paid for participating.
   • Confidentiality: You will not be asked to give your name and all answers will be held confidentially.
   • Psychological harm: You should not feel distressed during this task.

2. • Give name of organisation, researcher and address in case participants wish to contact you later.
Details of procedures so participants are in a position to give informed consent, include duration and details of the placebo procedure, as well as how data will be used e.g: The aims of this study are to investigate the effectiveness of a particular drug. One of the reasons that people get better when given a drug is because they believe it will make them better. In order to separate the actual effects of the drug from this belief in research we use a placebo drug as a control (i.e. one that has no physiological effect). In this study you will be given a small pill to take once a day for 30 days. You may be in the ‘real’ drug group or the placebo group. You will not be told this until the end of the study. At the beginning and the end of the study we will conduct a health check and ask you to fill in a short questionnaire. At the end of the study we will tell you which group you were in. You will be fully debriefed about this research study and the results will be shared with you.

Details of pay: Participation is voluntary and there is no payment.

Right to withdraw: At any time during the 30 days you should contact us if you have concerns about your health or just wish to ask further questions. If you feel uncomfortable you are free to leave the study and your health will still be checked.

Confidentiality: You will not be asked to give your name and all information will be held confidentially.

Physical harm: The drug you are being given has been through many trials and is regarded as safe. There are no known or anticipated side effects.

3. IV = exercise, can be operationalised by requiring participants to take part in one of two conditions: One hour walking briskly each day or a control group who take no exercise.

DV = happiness, can be operationalised by using a questionnaire to assess happiness, the DV is the score on the questionnaire.

Extraneous variables: general patterns of exercise before the study (people who already exercise a lot might find the lack of exercise makes them unhappy and people who don’t exercise might feel that having to do exercise makes them unhappy), other activities during the period of the study, general health.

Recruit 100 participants by advertising for people who take no exercise as part of their daily routine. Randomly allocate participants to the exercise or no exercise group. Brief each group about what is required – they are not to do any kind of exercise except for the exercise group who do one hour a day of vigorous walking. The duration of the study will be 30 days. At the beginning of the study and at the end of the study participants should be given the happiness questionnaire. Participants should be given information about the study in advance and, based on this information, asked to provide informed consent.

4. A Are people more helpful after watching a pro-social film?

a. Null: There is no difference in people’s helpfulness before and after watching a pro-social film.

b. Provide details of: Operationalisation of IV (length of film and details of film) and DV (include how you would test helpfulness – might need two tests for before and after and then have to check they are equivalent), extraneous variables to be controlled (e.g. factors that might affect willingness to think helpfully – the room might be too cold), method of sampling: how many and what kind of participants to recruit, the sequence of events (test helpfulness, show pro-social film, retest helpfulness), time interval between tests and seeing the film, information to be given to participants for informed consent.

c. Strength of independent groups would be no need to have two tests.

d. Validity of helpfulness questionnaire. Compare with an existing test of helpfulness to check concurrent validity, or ask friends how helpful a person is to see if the test is measuring what it intends to measure.

e. Future research could consider the effect of anti-social films on helpfulness, or could look at the behaviour of certain individuals in the film – is a film more effective if there are men or women in the lead role?
B Do people remember more in the morning or the afternoon?
   a. Null: There is no difference in recall in the morning or the afternoon.
   b. Provide details of: Operationalisation of IV (specify time of day for morning and afternoon) and DV (includes how you would test recall – might need two tests for morning and afternoon and then have to check they are equivalent), extraneous variables to be controlled (e.g. distractions that might reduce recall), method of sampling: how many and what kind of participants to recruit, the sequence of events (could use ABBA design: morning afternoon afternoon morning on four successive days), information to be given to participants for informed consent.
   c. Strength of independent groups would be no need to have two tests.
   d. Validity of test of recall. Validate by comparing with measures that should be correlated e.g. GCSE exam performance (predictive validity). Compare with other tests of memory (concurrent validity).
   e. Future research could consider whether other cognitive tasks such as concentration are different in the morning/afternoon.

C Do relaxation exercises reduce stress?
   a. There is no difference in stress levels before and after doing a relaxation exercise.
   b. Provide details of: Operationalisation of IV (relaxation exercise) and DV (measurement of stress levels, can use galvanic skin response to assess ANS activity or stress questionnaire, but that would need to be in two versions in order to be used twice), extraneous variables to be controlled (e.g. experience using relaxation techniques), method of sampling: how many and what kind of participants to recruit, the sequence of events, information to be given to participants for informed consent.
   c. Strength of independent groups would be no effect from being assessed twice.
   d. Validity of measuring stress – a physiological measurement is not the same as the experience of stress but it is objective. A questionnaire would need to be in two forms.
   e. Future research could consider other activities that might reduce stress such as vigorous exercise or watching a happy film.

Page 137 No. 6.2
1. • Kind of questions e.g. closed (How many minutes per day do you spend speaking with your mother/father? Rate your closeness to your different family members) and open (What do you like best about your family? Describe your relationship with your mother/father/other family members).
   • Sample: Select 50 × 12 year olds and 50 × 18 years olds. To make the sampling representation it would be best to use a quota system – within each age group recruit children from different socio-economic groups and equal numbers of boys and girls. Include lie scale questions to check on the honesty of participants (e.g. ‘Do you think you behave well all the time?’ Anyone who answers ‘yes’ is lying!).
   • Procedure: Select participants and brief them about what will be required, including informed consent. Ask them to complete the questionnaire in a quiet classroom, on their own. Instructions on the questionnaire say ‘This is an important questionnaire, aiming to help us to understand family relationships. Take your time and try to answer all questions truthfully and fully. The information will be used to help families understand their children better’.

2. • Method of investigation: Naturalistic observation – good to observe people in their usual environment. Controlled observations – using a behaviour checklist to enable unbiased recording of behaviour.
   • Materials/apparatus: Behaviour checklist (would contain categories such as ‘shouts at players on pitch’, ‘shouts at individuals nearby’, ‘shakes fist/arm at individuals nearby’), video recordings of a variety of fans at football matches.
• Procedure: Recordings made at different venues involving different football team supporters (to ensure representativeness). Observers trained in use of the behaviour checklist and then three observers record the behaviour of individual people at football matches (use three observers to ensure reliability). Record information relating to 50 different individuals to get representativeness. For each person observe their behaviour throughout the match. This could be done using, for example, time sampling where every 30 seconds you record what participants are doing.

3. • Hypothesis: TV programmes shown between 8 and 9pm contain less sex and violence than programmes shown between 9 and 10pm.
• Categories to use: Kissing, talking about sex, hitting someone, using a gun.
• Sampling method: Select 20 programmes between 8 and 9pm and 20 programmes between 9 and 10pm. Observe recordings using time sampling – once every minute observers tick behaviours being displayed.
• Procedures: Use three observers (to enhance reliability) and record sample of programmes. Observations between the three observers should be compared to see if there is close agreement.

4. **A Attitudes to eating**
   a. What do people think and feel about their eating behaviour?
   b. Design interview: Some questions might be closed, e.g. Tick the items in this list that you wouldn’t eat. Some questions may be open e.g. Would you say you had healthy eating habits? Could you explain your answer? etc. Train interviewers in asking follow-up questions to collect as much information as possible about what people think and feel. Advertise for participants (volunteer sample). Brief participants explaining the purpose of the study and what will be entailed (an hour interview), advise them on their right to withdraw and confidentiality. Participants interviewed in a comfortable room with no distractions and one interviewer. Interview recorded on video to help analysis.
   c. Participants may not be willing to reveal personal information such as issues about restricting their food intake if they have an eating disorder. Therefore answers may lack validity.
   d. Strength: Makes it easier to analyse the answers people give e.g. about the kind of food they eat. Weakness: Restricts the kind of answers people give because the researchers have decided in advance on the range of answers that may be given. People may have quite different attitudes about eating.
   e. Reliability could be improved by training interviewers so they learn how to develop questions on the spot. Interviewer could discuss what they do so they all use similar strategies. This would improve reliability because interviewers will be more likely to be consistent in what they do – theoretically if they repeated an interview with the same participant they should get the same result.
   f. Opportunity sampling might be more appropriate because it is just easier to take participants who are there rather than waiting for volunteers.

**B Experience of stress in daily life**
   a. What kind of things stress people in their everyday lives?
   b. Design interview: Some questions might be closed, e.g. Which of the following experiences do you find stressful? Some questions may be open e.g. Think of a recent occasion when you felt stressed, describe how you felt and what you did. What do you think stresses you most? etc. Train interviewers in asking follow-up questions to collect as much information as possible about what people think and feel. Advertise for participants (volunteer sample). Brief participants, explaining the purpose of the study and what will be entailed (an hour interview), advise them on their right to withdraw.
and confidentiality. Participants interviewed in a comfortable room with no distractions and one interviewer. Interview recorded on video to help analysis.

c. Participants may not be willing to reveal personal information such as issues about stressful experiences because they feel embarrassed. Therefore answers may lack validity.

d. Strength: Makes it easier to analyse the answers people give e.g. about the things that stress them. Weakness: Restricts the kind of answers people give because the researchers have decided in advance on the range of answers that may be given. People may have quite different attitudes about stress.

e. Reliability could be improved by training interviewers so they learn how to develop questions on the spot. Interviewer could discuss what they do so they all use similar strategies. This would improve reliability because interviewers will be more likely to be consistent in what they do – theoretically if they repeated an interview with the same participant they should get the same result.

f. Opportunity sampling might be more appropriate because it is just easier to take participants who are there rather than waiting for volunteers.

C Successful revision techniques

a. What revision techniques are most successful for students?

b. Design interview: Some questions might be closed, e.g. Which of the following techniques do you use? Some questions may be open e.g. Describe what you usually do when revising. List the techniques that you have used. etc. Train interviewers in asking follow-up questions to collect as much information as possible about what people think and feel. Advertise for participants (volunteer sample). Brief participants, explaining the purpose of the study and what will be entailed (an hour interview), advise them on their right to withdraw and confidentiality. Participants interviewed in a comfortable room with no distractions and one interviewer. Interview recorded on video to help analysis.

c. Participants may not be honest about their revision techniques because they don’t use any, so they would then describe what they think they should be doing. Therefore answers may lack validity.

d. Strength: Makes it easier to analyse the answers people give e.g. about the techniques they use. Weakness: Restricts the kind of answers people give because the researchers have decided in advance on the range of answers that may be given. People may use unusual techniques that are not included.

e. Reliability could be improved by training interviewers so they learn how to develop questions on the spot. Interviewer could discuss what they do so they all use similar strategies. This would improve reliability because interviewers will be more likely to be consistent in what they do – theoretically if they repeated an interview with the same participant they should get the same result.

f. Opportunity sampling might be more appropriate because it is just easier to take participants who are there rather than waiting for volunteers.

Page 139 No. 6.3

1. Your abstract should cover aims, methods, findings and conclusions. For example:

   The matching hypothesis states that people are attracted to members of the opposite sex who are similar in terms of physical attractiveness rather than seeking the most physically attractive mate. This study aims to test this hypothesis by selecting a set of photographs of married couples and asking participants to rate the attractiveness of each of the partners (females rate male photos and males rate female photos). Forty-eight participants from the sixth form at our school took part (24 girls and 24 boys) and were asked to rate the physical attractiveness on a scale of 1 to 10 (10 = highly attractive). The correlation was not significant (p = 0.05,
critical value = 0.65, observed value = 0.44, null hypothesis accepted). This suggests that, when looking for a partner, people do not try to match their own physical attractiveness, they may be influenced by a variety of other factors. (145 words)

2. Your introduction should be divided clearly into paragraphs. Paragraph 1 should be an outline of the general topic area. Paragraph 2 should narrow down to the specific area of study, further paragraphs should discuss specific and relevant studies. The final paragraph should provide a bridge to the specific hypothesis you are testing. For example:

Relationships start with interpersonal attraction. Psychologists have proposed various explanations for interpersonal attraction. One view is that we seek partners who are physically attractive, possibly because this is evidence of their good reproductive potential (evolutionary theory). Features that are considered physically attractive, such as a good complexion and white teeth, suggest the possessor is healthy and has good genes. Mating with such an individual will help maximise your own reproductive success.

An alternative view is the matching hypothesis which suggests that we actually seek a partner whose physical attractiveness matches our own physical attractiveness. This is likely because even though we find physically attractive people most attractive we go for a compromise when selecting a potential partner in order to avoid rejection. We don’t select a partner who is much less physically attractive because this would limit our reproductive success. We can do better and maximise our reproductive success.

Walster et al. (1966) tested this hypothesis in a study called the ‘Computer Dance experiment’. About 400 students were invited to a fresher’s week dance and told they would be paired with a similar partner (in fact they were paired randomly and judges rated each student in terms of physical attractiveness). At the end of the dance students were all given questionnaires including a question about whether they would like to see their partner again. Walster et al. found that students were most likely to want to see a physically attractive partner again rather than one who was more of a match.

However, this study was criticised because it didn’t relate to real-life relationships very well. When Walster and Walster (1966) repeated the study they did find support for the matching hypothesis probably because this time the participants spent time together beforehand and were given a choice of whom to partner. This time they did prefer someone who matched their own perceived physical attractiveness. This makes sense as matching is likely to occur if you are seeking a relationship rather than rating someone who you’ve been paired with. In the latter case there was no opportunity for selection so matching would not have taken place.

Further support for the matching hypothesis has been found in studies of real-life couples. Silverman (1971) conducted an observational study of couples in public places (such as bars) and rated the couples on a 5 point scale and found high similarity between members of a couple. The observers also noted that the more similar the attractiveness, the happier the couple were rated in terms of the degree of physical intimacy (e.g. holding hands).

Murstein (1972) asked couples who were engaged or going steady to rate their own and their partners’ attractiveness on a 5-point scale. Independent judges also rated the participants’ attractiveness. The similarity ratings for couples were compared with ratings made of randomly paired couples. Murstein found that real couples were significantly more similar than the randomly paired couples.

All of these studies support the matching hypothesis in a variety of different settings and in relation to couples in varying stages of a relationship.
The aim of this study is to replicate Murstein’s research, adapting the original design. Instead of actually asking participants to rate their own attractiveness, this study will use photographs of strangers and require participants to rate the attractiveness of the people in the photographs. We would expect couples to be similar (matching) in attractiveness because they have chosen each other. They should go for someone of a similar level of attractiveness rather than someone who is much more or much less physically attractive.

Statement of alternative hypothesis: There is a positive correlation in the attractiveness ratings given to the partners in a long-standing romantic relationship.

3. a. It is an observational study/content analysis (observing what people say and analysing the content).
   
b. Strength: It has high ecological validity because it is based on what people actually say without realising they are being observed.
   
   Weakness: Observers may be biased in their categorisations of newspaper content e.g. what one observer sees as a negative reference to homosexuality is not rated negatively by another observer.
   
c. The quantitative data would be numerical whereas the qualitative data would be in words expressing aspects that cannot be reduced to numbers.
   
d. The final data will be quantitative as the students are counting instances.
   
e. The methods section should include: details of what newspapers were collected and their dates; the method of sampling these newspapers should be discussed (e.g. front page articles only examined or all pages examined). Examples should be recorded of what counts as a positive, negative or neutral e.g. statements that enhance perceptions (would make you feel better about the person or event).

4. a. | Heart rate after 15 minutes of isolation | Heart rate after 15 minutes of viewing pictures of other sheep |
    |----------------|--------------------------------------------------|
    | 102.5 beats per min. | 78.3 beats per min. |
   
b. Bar chart should have a title and vertical axis labeled ‘beats per minute’, horizontal axis labeled ‘15 minutes of isolation’ and ‘15 minutes of viewing pictures of other sheep’.
   
c. Wilcoxon (or related t test) because data is interval (beats per minute of sheep), repeated measures (each sheep tested twice), looking at difference.
   
d. You would use a 5% significance level as this offers a good compromise between Type 1 and Type 2 errors. As this is not a study concerning ‘life and death’ issues a more stringent 1% level is not necessary.
   
e. The hypothesis is non-directional, so two-tailed.
   
f. Look up critical value in the table of significance (N = 4, two-tailed, 5% level). If the observed value is equal to or less than the critical value the null hypothesis can be rejected.
   
g. The bar chart suggests that there is a large difference between the two groups and therefore suggests that isolation causes a different heart rate than lack of isolation (seeing other sheep).
   
h. Implies that isolation leads to stress whereas no isolation (even just seeing pictures of other sheep) reduces stress/anxiety.

   
   Researcher’s name: E. M. Abernethy
Page 141 No. 6.4

1. Peer review refers to the process by which academic equals (experts in a particular field) present critical information about research that has been conducted in order to judge the validity and value of a piece of research.

2. Experts in the field are able to judge if the research is objective, has been well controlled and the conclusions drawn are based on the data. These all contribute to the validity of the research.

3. e.g.
   - It is not always possible to find an expert so the person who is used may not be able to judge the research accurately.
   - Anonymity may enable people to present overly critical research because they don’t wish to see certain research published. So anonymity might not provide valid peer review.
   - Reviews tend to be poor for straightforward replications and also studies with negative results (because reviewers don’t think such results are valuable). This results in a publishing bias.

4. a. Yes.
   b. Listeners send in comments and the radio station monitors its own broadcasts.
   c. Yes, in order to ensure high quality broadcasting and making sure people don’t say things that are wrong or harmful.

5. You could look at:

Chapter 7 Mathematical skills

Page 147 No. 7.1

1. a. Volunteer (self-selecting).
   b. Random.
   c. Opportunity.
   d. Opportunity and event sampling.

2. a. Mean, median or mode.
   b. Range, standard deviation or variance.
   c. 5%
   d. Tail is longest going to right.
   e. Tail is longest going to left.
   f. Should be gaps between bars and often data is categories (nominal data). Frequency on one axis.
g. Wilcoxon or related t test.
h. Spearman’s or Pearson’s.
i. +.56 or −.37
j. List of people’s favourite football teams or categories of height (small, medium and tall).
k. Measuring height using metres.
l. Data collected by Milgram in his obedience studies.
m. Research using census data or crime statistics.
n. Measurements of height.
o. People describing the qualities they like in a partner.

Page 149 No. 7.2

1. a. 8
   b. 9.75
   c. 24
   d. 29.25
   e. 30
   f. 18

2. a. 57
   b. 33
   c. 102
   d. 4

3. a. 57.3
   b. 32.7
   c. 101.6
   d. 4.1

4. a. 57.26
   b. 32.71
   c. 101.63
   d. 4.10

5. a. \( \frac{3}{5} \)
   b. \( \frac{3}{4} \)

6. Having to come in to work early
   Fraction: \( \frac{5}{50} \)
   Degrees in a pie chart: \( \frac{5}{50} \times 360 = 36 \)

   Too much noise
   Fraction: \( \frac{12}{50} \)
   Degrees in a pie chart: \( \frac{12}{50} \times 360 = 86.4 \)

   Too much work
   Fraction: \( \frac{23}{50} \)
   Degrees in a pie chart: \( \frac{23}{50} \times 360 = 165.6 \)
Not enough appreciation

Fraction: \( \frac{8}{50} \)
Degrees in a pie chart: \( \frac{8}{50} \times 360 = 57.6 \)

Other

Fraction: \( \frac{2}{50} \)
Degrees in a pie chart: \( \frac{2}{50} \times 360 = 14.4 \)

7. a. 0.375
b. 0.5
c. 0.75
d. 0.66667
e. 0.95

8. 4

9. a. \( \frac{13}{100} \)
b. \( \frac{53}{200} \)
c. \( \frac{3}{4} \)
d. \( \frac{39}{50} \)

10. a. 13%
b. 26.5%
c. 75%
d. 78%

11. a. 5
b. 42
c. 3.9

12. 10 students texting = 25%, 25 students answering questions = 62.5%, 5 students out of their seats = 12.5%

13. a. About \( \frac{40}{60} = 67\% \)
b. About 100 to 365 = 265

Part-to-whole ratio = 15:50 (3:10)

15. Calculating \( \frac{3}{10} \) of 60 = 18 men.

Page 151 No. 7.3

1. a. \( 3.45 \times 104 \)
b. \( 7.63 \times 1011 \)
c. \( 4.5 \times 10^{-7} \)
2. a. 23,000,000  
   b. 0.023  
   c. 562,000  
   d. 0.0000562  

3. a. 20,000,000  
   b. 0.02  
   c. 600,000  
   d. 0.00006  

4. a. Two significant figure = 0.036  
    Two decimal places = 0.04  
   b. Two significant figures = 440.0  
    Two decimal places = 435.42  
   c. Two significant figures = 5.2  
    Two decimal places = 5.21  
   d. Two significant figures = 3,600  
    Two decimal places = 3,593.00  

5. | Answers to question | Convert to the full value | Work out angle for pie chart |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly agree</td>
<td>2.1 × 105</td>
<td>210,000</td>
</tr>
<tr>
<td>Agree</td>
<td>1.18 × 106</td>
<td>1,180,000</td>
</tr>
<tr>
<td>Disagree</td>
<td>4.5 × 104</td>
<td>45,000</td>
</tr>
<tr>
<td>Strongly disagree</td>
<td>6.5 × 104</td>
<td>65,000</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td><strong>1,500,000</strong></td>
</tr>
</tbody>
</table>

6. Approximately 4,000 participants and mean 80,000 = mean score 20  

7. 3kg  

8. $8.14 \times 10^{-6}$ % (note that US billions have been used here, 1 billion = $1,000,000,000$)  

9. | Numbers | 22 | 19 | 10 | 23 | 18 | 10 | 12 | 15 | 17 | 19 | 10 | 18 | 15 | 22 | 20 |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Rank</td>
<td>13.5</td>
<td>10.5</td>
<td>2</td>
<td>15</td>
<td>8.5</td>
<td>2</td>
<td>4</td>
<td>5.5</td>
<td>7</td>
<td>10.5</td>
<td>2</td>
<td>8.5</td>
<td>5.5</td>
<td>13.5</td>
<td>12</td>
</tr>
</tbody>
</table>

**Page 153 No. 7.4**  

1. a. Round up or down = $20+10+10+10++10+20+30+30+10+20 = \frac{170}{10}$ values = 17  
   b. About 50 divided by 7 = about 7  

2. a. 16.50  
   b. 7.29  

3. a. Sum of  
   b. Number of data items  

4. a. Since the mean is 18, and excluding more extreme values the SD looks like 8  
   b. Using same logic SD looks like 3
5. Calculation using formula in textbook (sometimes \( n-1 \) is used in the formula):
   a. 7.80
   b. 3.19

6. a. \[ 1 - \frac{(6 \times 954)}{21 \times (21 \times 21 - 1)} \]
   b. 0.381
   c. For 20 participants the observed (calculated) value must be greater than or equal to .380 for significance for a one-tailed test at 5% level. For 25 participants this value is even lower so .381 is significant for 21 participants.
   d. Therefore we can reject the null hypothesis and accept the alternative hypothesis at \( p \leq 0.05 \)

7. a. Select the smaller value \( R1 = 23 \)
   Substitute values: \( U1 = 23 - \frac{6(6 + 1)}{2} \)
   b. \( U1 = 2 \)
   c. The critical value is 8. The observed (calculated) value must be equal to or less than the critical value for a one-tailed test at 5% level, therefore the result is significant.
   d. Therefore we can reject the null hypothesis and accept the alternative hypothesis at \( p \leq 0.05 \)