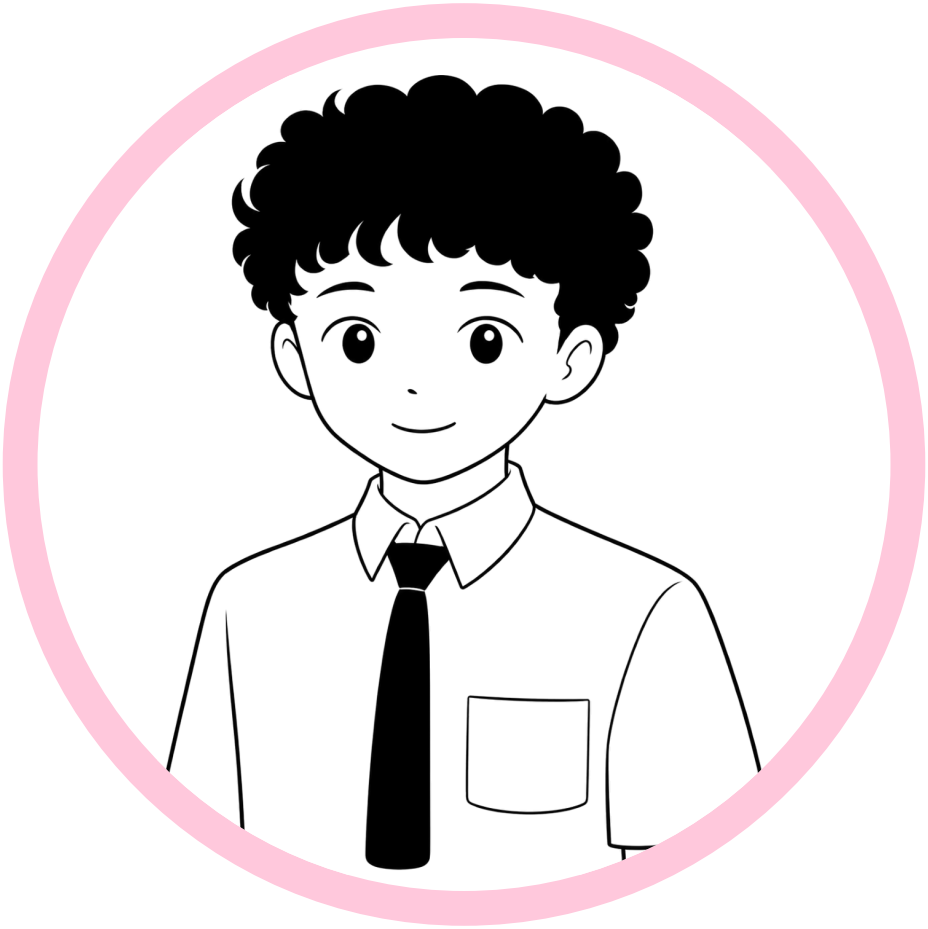
AQA A-LEVEL PSYCHOLOGY · 7182

PSYCHOLOGY



MARK SCHEME

3.2.3.3 INFERENTIAL TESTING

SCORE	PERCENTAGE
/470	%

STEMUP

These questions belong to AQA, not to us.

The examination questions, mark schemes and examiner reports in this pack were written and published by AQA. They are past assessment materials, reproduced here so that students can revise from them.

What we do not own

STEMUP is not affiliated with AQA and does not own, and makes no claim to own, any of the question, mark scheme or examiner report material in this pack. AQA has not endorsed, approved or been involved in the production of it. AQA and the qualification titles used here belong to their respective owners and appear for identification only.

Where to find the originals

AQA publishes its past papers, mark schemes and examiner reports free of charge. Always check the current specification at aqa.org.uk before you revise, because content and wording can change between series.

If you hold the rights to any material in this pack and would like it amended or removed, contact admin@stemup.co.uk and we will act on it.

Q1.

(a) **[AO1 = 1]**

1 mark for D

1

(b) **[AO3 = 3]**

3 marks for a clear and coherent discussion of why a covert observation of children might be more beneficial than an overt observation using the detail given below.

2 marks for a discussion of why a covert observation of children might be more beneficial than an overt observation.

1 mark for a muddled or limited discussion, or for an answer with no comparison.

Possible content:

- compared to an overt observation if the researcher is covert then the children are completely unaware that they are being monitored and this increases the likelihood of the behaviour that is produced being spontaneous and natural
- the researcher can also be more certain that covertly collected data are more likely to be valid measurements of behaviour and / or that these behaviours might therefore indicate possible cause and effect relationships.

Credit other valid points including discussion of ethical issues.

3

(c) **[AO2 = 4]**

Level	Mark	Description
2	3-4	There is clear knowledge of why the sign test would be appropriate in this study, with some accurate detail. The answer is generally coherent with effective use of appropriate terminology.
1	1-2	There is limited or partial knowledge of why the sign test would be appropriate with some detail. The answer lacks coherence and use of appropriate terminology.
	0	No relevant content.

Possible content:

- the researcher is looking for a difference in helping behaviour and doing a sign test is one way in which the analysed data would show such a difference
- the study focused on a single group of children who were tested under both conditions – the sign test can only be used with one group of participants / repeated measures design
- using a sign test will allow the researcher to decide whether differences in helping behaviour are due to chance factors or a ‘real’ effect

STEMUP.CO.UK

- this study produces quantitative/numerical data and the sign test is one way of analysing such data
- there are rules about when a particular test can be used and in this case the design of the study meets the rules for using a sign test.

NOTE: reference to level of measurement is not expected, but can be credited, e.g.: this study produces data in the form of categories / behavioural categories / frequency counts (nominal level of measurement).

4

[8]

Q2.

(a) **[AO2 = 1]**

1 mark for:
A Cognitive

1

(b) **[AO2 = 4]**

1 mark for each of the following:

- a title that includes both co-variables and reference to correlation / relationship
- appropriately labelled X axis
- appropriately labelled Y axis
- accurately plotted points

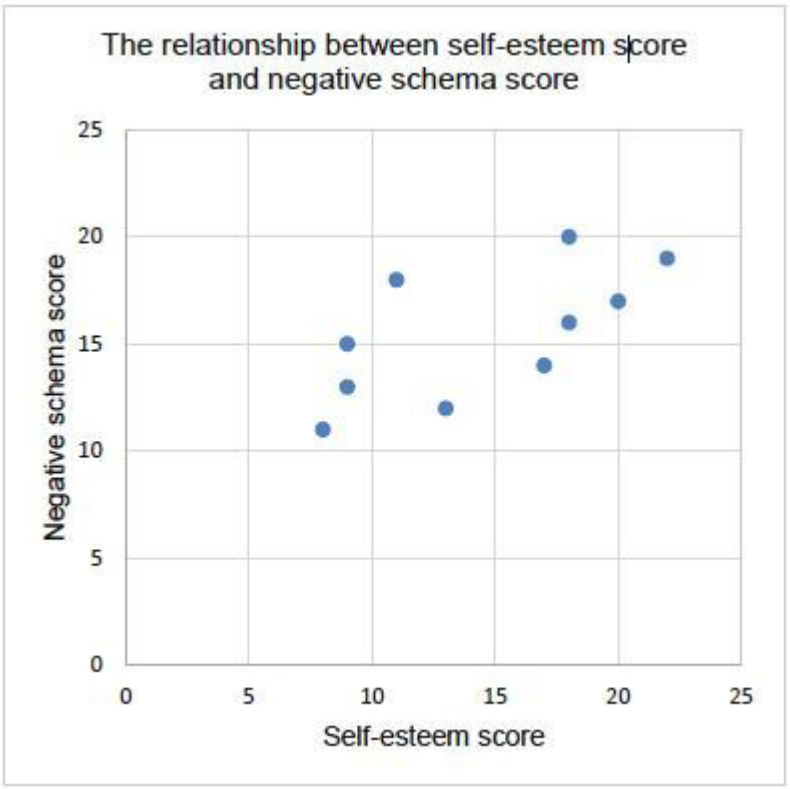
Note: co-variables should be operationalised ('score') either in the title or on the axes.

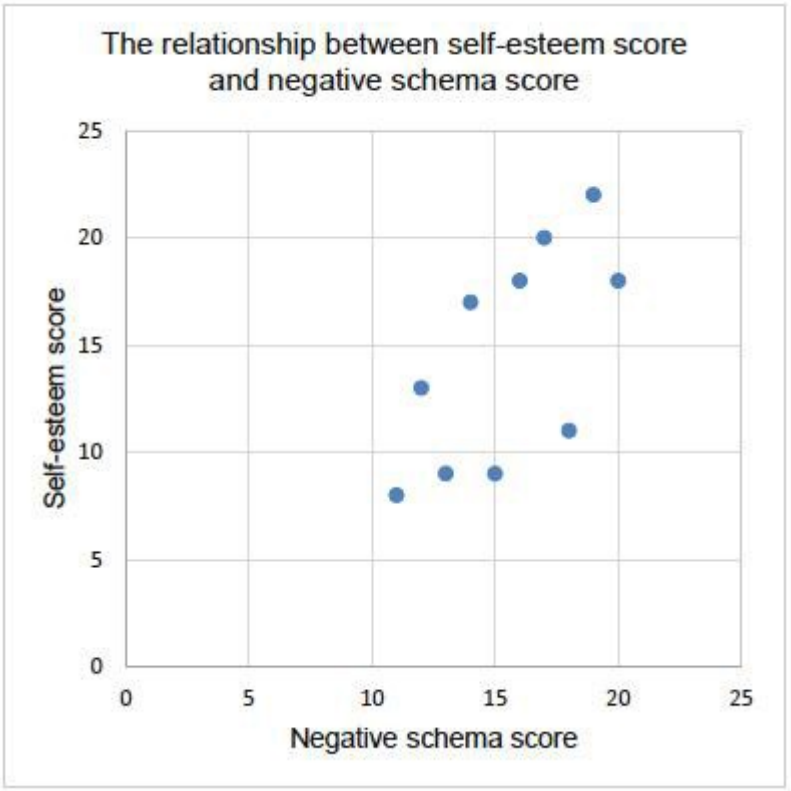
Accept: line of best fit.

Do not accept: points connected by lines (e.g. frequency polygon); this is an inappropriate graphical display, so maximum 1 mark for Title (bullet 1 above).

If maximum 1 for Title, Title does not need to include 'score'. Must include both co-variables and reference to correlation / relationship.

STEMUP.CO.UK





4

(c) **[AO2 = 2]**

1 mark for stating that the level of measurement is ordinal data.

PLUS

STEMUP.CO.UK

1 mark for an explanation:

Possible content:

- the co-variables (self-esteem and negative schema) have been measured using an 'unsafe' nonstandardised scale / the researcher constructed the scales herself
- the co-variables cannot be measured objectively / mathematically / may not be 'real things'
- because units of measurement are not of equal size / of unknown size

Accept alternative wording.

2

(d) **[AO2 = 1]**

+0.70.

1

[8]

Q3.

(a) **[AO2 = 6]**

Level	Mark	Description
3	5-6	Consent form is well detailed and practical, showing sound understanding of the requirements of a good consent form. All three elements are appropriately represented. The answer is clear and coherent. Specialist terminology is used effectively. Minor detail and / or explanation sometimes lacking.
2	3-4	Consent form is mostly detailed and practical, showing some understanding of the requirements of a good consent form. The answer is mostly clear and well organised. Specialist terminology is mostly used effectively.
1	1-2	Consent form only includes basic details of what participants can expect to happen in the experiment or how they will be protected. Knowledge of the requirements of a good consent form is limited. The whole answer lacks clarity, has many inaccuracies and is poorly organised. Generic consent forms without link to the stem.
	0	No relevant content.

Possible content:

Outline of what the experiment entails:

- an explanation of the general purpose of the research
- they would need to watch an assigned horror film every night before going to bed for 7 nights and would need to watch an assigned romantic comedy every

STEMUP.CO.UK

night before going to bed for 7 nights
- a daily requirement to truthfully respond to a text message asking whether they had experienced a nightmare
- the two-week duration of the experiment.

Ethical guidelines:

- no pressure to consent
- they can withdraw at any time
- they can withdraw their data from the experiment
- their data will be kept confidential and anonymous
- they should feel free to ask the researcher any questions at any time
- they will receive a full debrief at the end of the programme.

Format and style of consent form:

- will require the participant's agreement
- could be written as a form that participants need to sign
- could include space for the participant to sign the consent form
- could include a space for the participant to write the date
- could include a space for the participant to print their name.

Credit any other relevant information.

6

(b) **[AO2 = 3]**

3 marks for an appropriate, clear and coherent directional operationalised hypothesis.

2 marks for a statement with both conditions of the IV and a DV that lacks clarity and coherence or has only one variable operationalised.

1 mark for a muddled statement with both conditions of the IV and DV present or where neither variable is operationalised.

0 marks for expressions of aim / questions / correlational / null / non-directional hypotheses or statements with only one condition of the IV present OR if the IV is incorrectly described as group 1 and group 2.

Possible content:

- participants will report more nightmares after watching a horror film before bedtime than after watching a romantic comedy film before bedtime. Accept alternative wording
- participants will report fewer nightmares after watching a horror film before bedtime than after watching a romantic comedy film before bedtime. Accept alternative wording.

3

(c) **[AO2 = 2]**

2 marks for a clear explanation that is linked to the experiment

1 mark for a limited or muddled explanation.

Possible explanations:

STEMUP.CO.UK

- necessary to avoid the effects of individual differences in frequency of nightmares
- film viewing habits, gender, hours of sleep, personality, etc, can have a big impact on the number of nightmares recalled.

Credit other relevant explanations.

2

(d) **[AO1 = 1 AO2 = 1]**

1 mark for each of the following points:

- order effects are likely to occur
- if you saw the horror films in the first week you may still be thinking about them / have higher levels of fear in the second week, resulting in more nightmares.

Accept alternative wording.

2

(e) **[AO2 = 3]**

1 mark for each of the following points (up to 3 marks):

- all 50 participants' names / numbers are put into a hat / container / computer
- a name is drawn from the container or a random name is generated by the computer and is assigned to the first group
- a second name is selected as before but this time goes in to the second group; this process continues until there are 25 in each group.

OR

1 mark for each of the following points (up to 3 marks):

- use 50 slips of paper, label 25 'horror' and 25 'romantic comedy'
- put / mix the slips of paper in a hat / container
- each participant draws a group from the container

Note: Credit alternative methods of randomly splitting the sample

3

(f) **[AO2 = 2 AO3 = 2 maths]**

Mean:

1 mark – participants who watch horror films before going to bed report more nightmares then those who watch romantic comedies before bed. Accept alternative wording.

Plus

1 mark – mean number of nightmares reported is greater when horror films are watched than when romantic comedies are watched. Accept alternative wording.

Standard deviation:

1 mark – there is greater dispersion / variation in scores / spread of scores in the horror film condition than in the romantic comedy condition. Accept alternative wording.

STEMUP.CO.UK

Plus

1 mark – standard deviation is greater when horror films are watched before going to bed than when romantic comedies are watched before going to bed. Accept alternative wording.

Note - **0 marks** for just stating the data from the table.

Note – Justifications are not creditworthy in isolation

4

(g) **[AO2 = 2 maths]**

2 marks for a clear and appropriate explanation in the context of this experiment.

1 mark for a limited or muddled explanation.

This means that the difference in the number of nightmares reported after watching horror films compared to romantic comedies is significant at 0.05 level. This means there is less than 5% (1 in 20) likelihood (probability) that the difference was due to chance / due to something other than the IV.

Accept any other valid answer.

2

(h) **[AO2 = 1 AO3 = 2]**

1 mark for a modification to the design of the experiment that could improve validity.

Plus

2 marks for clear and coherent explanation for how the suggested modification might improve validity of this study.

1 mark for a limited/muddled explanation for how the suggested modification might improve validity of this study.

Possible content:

- include more than one question in the text message to the students. This would make the aim of the experiment less obvious to guess which would in turn reduce demand characteristics and improve the validity of the experiment
- make the conditions less obvious. Rather than having one film the students could be asked to watch an episode from a TV series. The episodes watched in one condition would be those which contained any scary concepts / themes whereas episodes watched in the other condition would not contain any scary themes / concepts
- guarantee anonymity so people will give honest answers and not feel embarrassed
- use a broader sample, not just students. Students may be more or less inclined to watch horror films anyway
- use a different sampling technique to avoid a self-selected sample and thus avoid volunteer bias as that may make them more susceptible to demand characteristics.

Accept alternative answers.

Note: Independent groups design would not be an appropriate modification in this

STEMUP.CO.UK

study; however candidates can make the case for a matched pairs design, which could be credit worthy.

3

[25]

Q4.

(a) **[AO2 = 2]**

1 mark for identifying the type of data as quantitative data.

Plus

1 mark for either of the explanations below:

- the data is numerical.
- the number of participants who completed the questionnaire in each condition.

OR

1 mark for explaining the type of data as primary data.

Plus

1 mark for either of the explanations below:

- the data is collected first hand (directly) from the participants.
- data is collected for the purpose of the investigation.

OR

1 mark for explaining the type of data as nominal/categorical data.

Plus

1 mark for either of the explanations below:

- the data is presented in categories/is discrete.
- the number of participants who did and did not complete the questionnaire in each condition.

If students identify more than one type of data, take the first type as the basis for their answer. If the data is not identified or identified incorrectly, no credit can be given for an explanation.

(b) **[AO2 = 3]**

1 mark for the likely outcome: more participants in condition 1 will complete the questionnaire than in condition 2/fewer participants in condition 2 will complete the questionnaire than in condition 1.

Plus

2 marks for clear and coherent explanation
1 mark for explanations that are limited or muddled or for explanations not linked to outcome
0 marks for an incorrect or irrelevant explanation

STEMUP.CO.UK

Possible content:

- accept reference to normative social influence – participants will follow the majority to avoid rejection/fit in
- the participants in condition 2 experienced social support/disobedient role models and so were more likely to defy the researchers
- accept reference to diffusion of responsibility/increased confidence to defy orders in condition 2
- accept explanation based on relevant studies, eg Asch variations.

Credit other relevant social influence explanations.

(c) **[AO3 = 4]**

1 mark for random allocation of participants to each condition.

Plus

3 marks for explanation of how this could be conducted in this experiment
1 mark for each bullet

- each participant is assigned a number or identified by name
- the numbers/names are placed into a random number generator/hat/lottery method
- the first participant drawn is assigned to condition 1, the second to condition 2, etc. OR the first 15 participants are assigned to condition 1 and the next 15 are assigned to condition 2.

OR

1 mark for using a matched pairs design.

Plus

3 marks for explanation of how this could be conducted in this experiment

1 mark for each bullet

- participants should be matched on a variable that is relevant to the experiment
- this could be ascertained through the use of a pre-test e.g. completing a confidence questionnaire
- participants from each matched pair are allocated to different conditions.

Credit other plausible ways of addressing the question. To gain any credit answers based on repeated measures should be appropriately detailed e.g. parallel versions of the questionnaire, time lapse etc.

(d) **[AO2 = 2]**

1 mark for the design is unrelated/independent groups/independent data

1 mark for the researcher is looking for a difference (between two conditions/sets of data) **or** an association/relationship (between two variables).

No credit for answers referring to correlations.

(e) **[AO2 = 2]**

STEMUP.CO.UK

1 mark for stating that the value of chi squared is significant (at the 5% level).

Plus

1 mark for explanation:

the calculated/observed value (3.97) is more than/exceeds the critical/table value of 3.84/at 5%.

Q5.

(a) **[AO2 = 1]**

1 mark

A

(b) **[AO2 = 2]**

2 marks for identification of the operationalised dependent variable: running time in seconds/time in seconds taken to run.
1 mark for dependent variable not operationalised: time in seconds; running time.
0 marks for incorrect reference to speed or to the IV.

(c) **[AO2 = 2]**

1 mark for explaining **either** you need to have continuous data **or** scores for each participant in order to draw a histogram.

Plus

1 mark for identifying that the data represents two separate conditions (with music/without music). Accept categorical/nominal.

Note: credit can be given for two separate conditions if the student explains clearly why this would make a histogram “inappropriate”.

(d) **[AO3 = 3]**

1 mark for each bullet point:
- a bar chart
- appropriate X-axis label eg ‘With or without music’; ‘Auditory Condition’, etc.
- appropriate Y-axis label eg ‘Mean/Average (400m) time in seconds’.

Note: these are independently awarded marks, eg candidates can achieve 2 marks for correctly labelled axes despite an incorrect graph type

Note: axis labels are interchangeable.

(e) **[AO2 = 2 AO3 = 2]**

Mean:

1 mark for interpreting what the mean times suggest about the effect of music on the participants’ 400m performance – participants run faster with music (take less time to run 400 metres) or participants run more slowly without music (take more time to run 400 metres). Accept alternative wording.

STEMUP.CO.UK

Plus

1 mark for an accurate justification about the difference in the mean scores in each condition – mean time is greater in condition A than condition B (or mean time is lower in condition B than condition A).

Standard deviation:

1 mark for an accurate comment about what the standard deviations suggest about the spread of scores in each condition – performance is more consistent in condition A than condition B (or performance is less consistent in condition B than condition A). Accept alternative wording.

Plus

1 mark for a justification about the difference between the standard deviations in each condition – standard deviation is smaller in condition A than in condition B (or standard deviation is greater in condition B than condition A).

Note: **0 marks** for just stating the data from the table, eg the mean time with music is 117 whereas it is 123 without music.

(f) **[AO2 = 4]**

Marks are for calculations and/or numerical answer – no need to show unit (%)

4 marks for the correct answer given to three significant figures: 4.88 (even if no correct workings are shown).
3 marks for correct answer **not** given to three significant figures eg 4.878 or 4.9.
2 marks if incorrect answer is provided even if all working is correct.
1 mark if incorrect answer and workings are partially correct eg one or two of the correct steps.
0 marks if the incorrect answer is given to three significant figures.

Correct workings:
$123 - 117 = 6$
$6 \div 123 = 0.048780$
$0.048780 \times 100 = 4.878$
Answer = 4.88

Note: alternative methods used for working may be creditworthy.

(g) **[AO2 = 5]**

1 mark for each bullet point:

A t-test is an appropriate choice of test for this data because
- a difference between the two sets of data is predicted
- the data is of interval/ratio level
- a repeated measures design has been used.

Credit other appropriate reasons e.g. reasons related to possible normal distribution, power of the test.

Plus

1 mark for each bullet point:

STEMUP.CO.UK

- the result is not significant (at the 5% level)
- because the calculated value of t (1.4377) is less than the critical/table value of t, which is 1.833 (at 0.05, for a directional hypothesis where df is 9).

(h) **[AO1 = 3]**

2 marks for a clear and coherent definition of a Type II error.

A Type II error would occur where a real difference in the data is overlooked as it is wrongly accepted as being not significant, accepting the null hypothesis in error (a false negative).

1 mark for a limited or incomplete definition of a Type II error.

Plus

1 mark for a reason for why the 5% level of significance is used in psychological research.

The 5% level is used as it strikes a balance between the risk of making the Type I and II errors (or similar).

(i) **[AO2 = 2 AO3 = 1]**

1 mark for identification of **one** appropriate extraneous variable eg type/rhythm of music/equipment/environmental variables.

Note – participant variables only acceptable if feasibly could have changed in a week eg illness/hydration level/injury/etc.

Plus

1 mark for an explanation of why it is important to control their chosen variable in the context of this study ie the impact on the DV.

Plus

1 mark for how the stated variable could be controlled.

(j) **[AO1 = 6]**

Level	Marks	Description
3	5 – 6	Knowledge of the process and purposes of peer review is clear and generally accurate. Specialist terminology is used appropriately.
2	3 – 4	Some knowledge of the process and/or purposes of peer review but there may be some omissions/lack of clarity. There is some appropriate use of specialist terminology.
1	1 – 2	Knowledge of the process and/or purposes of peer review is present but there may be serious omissions and/or inaccuracy. Specialist terminology is either missing or inappropriately used.
	0	No relevant content.

STEMUP.CO.UK

Possible content:

Process
- other psychologists check the research report before deciding whether it could be published
- independent scrutiny by other psychologists working in a similar field
- work is considered in terms of its validity, significance and originality
- assessment of the appropriateness of the methods and designs used
- reviewer can accept the manuscript as it is, accept with revisions, suggest the author makes revisions and re-submits or reject without the possibility of re-submission
- editor makes the final decision whether to accept or reject the research report based on the reviewers' comments/recommendations
- research proposals are submitted to panel and assessed for merit.

Purposes
- to ensure quality and relevance of research, eg methodology, data analysis etc
- to ensure accuracy of findings
- to evaluate proposed designs (in terms of aims, quality and value of the research) for research funding.

Credit other relevant material.

Q6.

(a) **[AO2 = 1]**

1 mark C

(b) **[AO2 = 2]**

1 mark 4/5^{ths}

1 mark for workings 16 divided by 4 = 4 and 20 divided by 4 = 5 (so 4/5^{ths})

(c) **[AO3 = 2]**

2 marks for a clear and coherent conclusion, plus relevant explanation based on the data.

1 mark for a vague/muddled conclusion.

Conclusion and explanation:

The training course appears to have a beneficial effect on teacher confidence as the majority of them (16 out of 20) say their confidence has improved.

(d) **[AO2 = 2]**

2 marks for a clearly operationalised dependent variable: the DV is whether the teachers thought their confidence in managing difficult behaviour was better, worse or the same after the course.

1 mark for a dependent variable that is not fully operationalised eg teachers' confidence/how they felt after doing the course.

(e) **[AO2 = 3]**

STEMUP.CO.UK

1 mark repeated measures design

Plus

2 marks for a clear and coherent explanation of why this design is appropriate in this case

1 mark for a vague or muddled explanation of why this design is appropriate in this case

Content:

It is important to ask the same participants to consider their level of confidence before and after taking part in the training in order to see whether their confidence has changed. It would not make sense to ask one group of people before training and another group of people after training because there may be individual differences in their levels of confidence anyway.

(f) **[AO2 = 3]**

1 mark calculated value of S = 2

Plus two marks for any two of the following points:

1 mark this is determined by converting the differences or outcomes to signs + or −

1 mark then taking the numerical value for/number of participants with the least common/frequent sign

1 mark any nil differences are ignored

(g) **[AO1 = 2]**

2 marks for a clear and coherent explanation

1 mark for a vague/muddled explanation

Content:

researchers use statistical tests to determine the likelihood that the effect/difference/relationship they have found has occurred due to chance.

(h) **[AO3 = 2]**

1 mark for each valid behavioural category suggested.

For credit the behaviour should be an **observable** behaviour that is disruptive or disobedient and would be likely to occur in a primary school classroom eg throwing something, shouting, banging on the desk.

(i) **[AO3 = 3]**

Award marks for a suitable record sheet/tally chart in table form.

1 mark for each bullet point addressed

- Table with spaces for tallies/recordings of instances

- Separate spaces for first and last 10 minutes
- Headed correctly with the six category spaces (may include the two used in answer to part (h) but names of categories not essential here)

No marks for drawing a bar chart or graph.

(j) **[AO3 = 4]**

Level	Marks	Description
2	3 – 4	A relevant problem is clearly identified. The explanation of how the observation would be improved by addressing this problem is appropriate and effective. The answer is generally coherent with effective use of appropriate terminology.
1	1 – 2	A relevant problem is identified. There is partial/limited explanation of how the observation would be improved by addressing this problem. The answer lacks coherence and use of appropriate terminology.
	0	No relevant content.

Relevant problems:

- observer effect (as observation is overt) means pupils would behave differently because they are aware that they are being observed – could be addressed by carrying out a covert observation
- observer bias/lack of objectivity because the researcher is working alone – could be addressed by working as part of a pair for inter-observer reliability
- limiting observations to first and last 10 minutes means the data may not be a valid representation of disruptive behaviour in lessons. Need to carry out observations at other times during the lesson too.

Q7.

(a) **[AO2 = 2 AO3 = 4]**

Level	Marks	Description
3	5 – 6	Conclusions in respect of both means and standard deviations are presented with clarity. Understanding of the relevance of each statistic is demonstrated. Justifications for each make good use of the values given.
2	3 – 4	Conclusions and justification in respect of both means and standard deviations are relevant, but there is some lack of clarity in both. Or, one is done well and justified appropriately (most usually this will be the mean).
1	1 – 2	One conclusion is drawn or two are partially correct. Any

STEMUP.CO.UK

		justification is limited. The answer lacks clarity.
	0	No relevant content.

Means

- Conclusion: when people believe they are presenting to a large audience they are less fluent in their spoken communication than when they believe the audience is small (or vice versa).
- Justification / Application: this is supported by the difference in the mean fluency scores which show more verbal mistakes (on average 6 more mistakes) when the audience is believed to be large (or vice versa).

Standard deviations

- Conclusion: performances of participants in Condition A where audience is believed to be small are less varied / dispersed / spread out than in Condition B where audience is believed to be large (or vice versa).
- Justification / Application: lower SD in Condition A suggests that individual performances in Condition A were more similar to each other and / or all quite close to the mean of 11.1.

(b) **[AO3 = 3]**

1 mark – this would be an improvement because the SD is a measure of dispersion that was less easily distorted by a single extreme score.

Plus

1 mark – one that takes account of the distance of all the verbal error scores from the mean.

Plus

1 mark – not just the distance between the highest verbal error score and the lowest verbal error score.

(c) **[AO2 = 4]**

1 mark for naming the t-test for independent / unrelated groups or a Mann-Whitney test.

Plus

Up to 3 marks for explanation for unrelated t-test. Credit relevant points as follows:

- can assume interval data because verbal errors can be assumed to be of equal size (ie one verbal error is equivalent to any other verbal error)
- the experimental design is independent groups
- the psychologist is looking for a difference between the two conditions.

OR

Up to 3 marks for explanation for Mann-Whitney test. Credit relevant points as follows:

STEMUP.CO.UK

- data should be treated as ordinal. Cannot assume interval data because verbal errors cannot be assumed to be of equal size (ie one verbal error is not equivalent to any other verbal error)
- the experimental design is independent groups
- the psychologist is looking for a difference between the two conditions
- SDs are quite different.

(d) **[AO1 = 2]**

2 marks for a clear and appropriate definition as follows:

This means that there is a less than 5% likelihood that this difference would occur if there is no real difference between the conditions **OR** the researchers would have a 95% confidence level.

1 mark for a less clear answer which shows some understanding, eg this means the researcher can conclude that the difference was not due to chance.

Accept any other valid answer.

(e) **[AO2 = 2]**

2 marks for a clear and detailed explanation applied to this study.

1 mark for a partial or muddled explanation or one that is only loosely applied to the study.

Credit answers based on any type of validity. Most answers will refer to either face or concurrent as follows:

- asking other people if verbal errors are a good measure of verbal fluency (face validity)
- giving participants an alternative / established verbal fluency test and checking to see that the two sets of data are positively correlated (concurrent validity).

Q8.

(a) **[AO2 = 4]**

2 marks for identifying two factors that are relevant for use of the sign test: nominal/categorical data; test of difference; related design/repeated measures.

Plus

Up to 2 marks for application of these to the investigation described:
- Nominal data as patients are assigned to one of three categories – ‘improved’, ‘deteriorated’ or ‘neither’.
- Testing for difference in the number of absences in the year following and prior to treatment.
- Repeated measures as the same patients' work records are compared before and after treatment.

(b) **[AO2 = 2]**

1 mark for identifying the correct value of s as 5

Plus

STEMUP.CO.UK

1 mark for explanation/calculation of how this was arrived at:

- The most commonly occurring sign is + (12) and the least frequently occurring sign is – (5). The 0s are disregarded.
- The total for the least frequently occurring sign is the value of s = 5

(c) **[AO2 = 2]**

1 mark for stating that the value of s (5) is not significant at the 0.05 level.

Plus

1 mark for explanation:

- The critical value is 4. As the calculated value is higher than/exceeds the critical value, the result is significant not at the 0.05 level.

Accept alternative wording

(d) **[AO3 = 3]**

Marks may be awarded for a single point that is expanded/elaborated or more than one point briefly stated.

1 mark only if there is no reference to the investigation described.

Possible points:

- Primary data are obtained ‘first-hand’ from the participants themselves so are likely to lead to greater insight: e.g. into the patients' experience of treatment, whether they found it beneficial, negative, etc.
- Secondary data, such as time off work, may not be a valid measure of improvement in symptoms of depression. Primary data are more authentic and provide more than a surface understanding: e.g. participants may have taken time off work for reasons not related to their depression.
- The content of the data is more likely to match the researcher’s needs and objectives because questions, assessment tools, etc. can be specifically tailored: e.g. an interview may produce more valid data than a list of absences.

(e) **[AO1 = 3 AO2 = 2]**

Level	Marks	Description
3	4 – 5	Knowledge of the implications of psychological research for the economy is clear. Application to the investigation described is effective. The answer is coherent with effective use of terminology.
2	2 – 3	Some knowledge of the implications of psychological research for the economy is present but there is a lack of detail/clarity. Application to the investigation described is limited or absent. Terminology is used appropriately on occasion.
1	1	An implication of psychological research for the economy is briefly stated.
	0	No relevant content.

STEMUP.CO.UK

AO1 – possible content:

- Psychological research may lead to improvements in psychological health/treatment programmes which may mean that people manage their health better and take less time off work.
- Absence from work costs the economy an estimated 15 billion a year annually and much of this absence is due to 'mild' mental illness: e.g. stress, anxiety.
- Psychological research may lead to better ways of managing people whilst they are at work to improve productivity: e.g. research into motivation and workplace stress.
- 'Cutting-edge' scientific research may encourage investment from overseas companies into this country.

Credit other relevant points/implications, including examples not linked to psychopathology.

AO2 – application

- If research (such as the investigation described) suggests that depressives are better able to manage their condition following CBT and return to work, then it may benefit the economy to make treatment more widely available, improve funding, etc.
- Psychological research such as this plays an important role in sustaining a healthy workforce and reducing absenteeism.

Credit other relevant application points.

Q9.

(a) **[AO2 = 2]**

1 mark – the hypothesis should be directional

Plus

1 mark – because there is past research indicating the likely direction of the effect (or similar)

(b) **[AO1 = 2 AO3 = 2]**

2 marks for a clear and coherent explanation of operationalisation

1 mark for a brief or muddled explanation of operationalisation

Content: operationalisation involves clearly specifying/defining observable behaviours that represent the more general construct under investigation/to enable the behaviour under investigation to be measured

Plus

1 mark each for two observable behaviours that could represent 'riding a bike with care'.

Examples: use of cycle lanes/tracks, passing pedestrians at a distance of at least 1 metre, using bicycle bell.

STEMUP.CO.UK

Credit any relevant observable behaviour.

(c) **[AO2 = 3]**

1 mark for knowledge of the term face validity – where a behaviour appears at first sight (on the face of it) to represent what is being measured

Plus

2 marks for clear and coherent application of the concept of face validity to the context

1 mark for brief or muddled application of the concept of face validity to the context

Application: Having a dog on a lead appears at first glance to be measuring considerate behaviour because if a dog is on a lead it is less able/likely to upset other people by coming close, frightening, chasing, biting, growling etc.

Credit other relevant applications.

(d) **[AO1 = 4]**

1 mark for each of two types of validity identified

Plus

1 mark each for a brief outline of each type of validity identified

Content:

- Concurrent – where performance on one measure correlates highly with performance on another measure of the same variable
- Ecological – where a measure of a behaviour accurately reflects the way in which the behaviour would occur in normal circumstances
- Temporal – where findings from research that took place at a certain point in time accurately reflect the way that behaviour would occur at a different point in time

Credit also other types of validity eg criterion, content, construct, population, predictive.

(e) **[AO2 = 1]**

1 mark – C Event sampling

(f) **[AO2 = 3]**

1 mark for each of the following:

- The student pair should discuss and agree beforehand their interpretation of the behavioural categories
- Each student should then observe the same people/space/target at the same time but record/tally independently
- Their respective recordings/tallies should be correlated using an appropriate

statistical test ascertain the level of agreement

(g) **[AO2 = 3]**

1 mark for the correct ratio: 3:2

Plus

2 marks for full workings: 23 + 12 + 19 = 54 and 10 + 17 + 9 = 36

54:36 both divisible by 9 (or 18)

54 ÷ 9 = 6 and 36 ÷ 9 = 4

6:4 can be simplified to 3:2

1 mark for partial workings eg first 2 of the above stages

(h) **[AO2 = 3]**

1 mark for each of the following:

- Data is categorical/nominal/frequency
- The students are looking for a difference or an association between two variables
- Design is independent/unrelated or categories are exclusive (observations cannot appear in more than one cell)

(i) **[AO2 = 2]**

1 mark for the correct answer: df = 2

Plus

1 mark for correctly substituting values into the formula as follows:

(2 − 1) x (3 − 1) = 1 x 2 = 2

(j) **[AO2 = 3]**

1 mark for each of the following:

- Yes, the result is significant
- Because the calculated value of Chi-square is more than the critical table value at 0.05, either 4.60 or 5.99, depending on whether student uses one or two-tailed values
- One-tailed is consistent with the hypothesis and Q13, but there is an argument that the Chi-square test should always be two-tailed so either can be credited
- Where df equals 2

(k) **[AO2 = 4]**

Level	Marks	Description

STEMUP.CO.UK

2	3 – 4	The paragraph is clear and coherent, showing sound understanding of the concept of levels of significance and effective application to the context. There is effective use of terminology.
1	1 – 2	The paragraph shows some understanding of the concept of levels of significance and/or some relevant application to the context. The answer lacks clarity and coherence. Terminology is either absent or inappropriately used.
	0	No relevant content.

Possible content:

- Explanation of levels of significance as an indication of the measure of the influence/effect of chance/random factors on the findings.
- With the present results there is a 95% confidence in accepting the research hypothesis/confidence that any difference/effect is due to the variables under investigation, in this case the location of the public spaces.
- There is a 5% possibility that the same frequencies would occur if there was no real difference between the two towns.
- The calculated value in this case well exceeds the critical value at 0.05 but does not meet the more stringent level of significance of 0.01.
- Possibility of type one error.

Credit other relevant material.

(l) **[AO3 = 3]**

1 mark for explaining that content analysis is suitable because the students are analysing recordings which are a form of media.

Plus

2 marks for a clear, coherent account of how the content could be analysed

1 mark for a brief or muddled account of how the content could be analysed

Content:

Students could identify specific ideas/concepts that occur in the recordings

They could then set up a system of categories and tally the ideas/concepts

(m) **[AO3 = 1]**

1 mark for any relevant inconsiderate behaviour eg leaving rubbish, leaving dirty mugs/plates, playing music loudly, throwing books, shouting

Q10.

STEMUP.CO.UK

(a) **[AO2 = 2]**

1 mark – because it is collected from participants directly/first hand

Plus

1 mark – rather than gathered from reports of previous research

(b) **[AO3 = 3]**

3 marks for clear statement of a strength coherent elaboration

2 marks for statement of a strength some elaboration

1 mark for vague/muddled strength

0 marks for no relevant material

Possible strengths

- Data gathered is more likely to be focused on the purpose of the research
- Less peripheral/redundant information
- Control issues – first hand data can be controlled whereas secondary data may have been gathered under differing conditions

Credit other relevant strengths.

(c) **[AO3 = 4]**

1 mark Mann Whitney test

Plus

3 marks for justification – **1 mark** for **each** of the following:

- Testing for a difference
- Independent design – different participants in each condition
- Data assumed to be ordinal ie not fixed intervals (also credit data is assumed to be non-parametric)

Q11.

(a) **[AO2 = 2]**

1 mark – because it is collected from participants directly/first-hand

Plus

1 mark – rather than gathered from reports of previous research

(b) **[AO3 = 3]**

3 marks for clear statement of a strength plus coherent elaboration

STEMUP.CO.UK

2 marks for statement of a strength plus some elaboration

1 mark for vague/muddled strength

0 marks for no relevant material

Possible strengths

- Data gathered is more likely to be focused on the purpose of the research
- Less peripheral/redundant information
- Control issues – first-hand data can be controlled whereas secondary data may have been gathered under differing conditions

(c) **[AO3 = 4]**

1 mark Mann Whitney test

Plus

3 marks for justification – **1 mark** for **each** of the following:

- Testing for a difference
- Independent design – different participants in each condition
- Data assumed to be ordinal ie not fixed intervals (also credit data is assumed to be non-parametric)

Q12.

(a) **[AO2 = 2]**

1 mark – because it is collected from participants directly/first-hand

Plus

1 mark – rather than gathered from reports of previous research

(b) **[AO3 = 3]**

3 marks for clear statement of a strength coherent elaboration

2 marks for statement of a strength plus some elaboration

1 mark for vague/muddled strength

0 marks for no relevant material

Possible strengths:

- Data gathered is more likely to be focused on the purpose of the research
- Less peripheral/redundant information
- Control issues – first hand data can be controlled whereas secondary data may have been gathered under differing conditions

Credit other relevant strengths.

(c) **[AO3 = 4]**

1 mark Mann Whitney test

Plus

3 marks for justification – **1 mark** for each of the following:

- Testing for a difference
- Independent design – different participants in each condition
- Data assumed to be ordinal ie not fixed intervals (also credit data is assumed to be non-parametric)

Q13.

(a) **[AO3 = 2]**

1 mark for a brief explanation

Plus

1 mark for elaboration

Possible content:

Control for individual differences so that the researcher can be more certain that the effect is not due to characteristics such as gender, personality etc.

Credit other relevant advantages.

(b) **[AO2 = 2]**

1 mark - Value of 'S' is 2.

Plus

1 mark for explanation - 'S' is the frequency of the least common difference. There are 8 positive differences, 2 negative differences and 2 ties.

(c) **[AO2 = 2]**

1 mark for identifying the correct value of N (N = 10, total number of differences)

Plus

1 mark for stating that for an N of 10, a value of 'S' = 2 is not significant at the $p \leq .05$ level.

Q14.

(a) **[AO3 = 2]**

1 mark for a brief explanation

STEMUP.CO.UK

Plus

1 mark for elaboration

Possible content:

Control for individual differences so that the researcher can be more certain that the effect is not due to characteristics such as gender, personality etc.

Credit other relevant advantages.

(b) **[AO2 = 2]**

1 mark - Value of 'S' is 2.

Plus

1 mark for explanation - 'S' is the frequency of the least common difference. There are 8 positive differences, 2 negative differences and 2 ties.

(c) **[AO2 = 2]**

1 mark for identifying the correct value of N (N = 10, total number of differences)

Plus

1 mark for stating that for an N of 10, a value of 'S' = 2 is not significant at the p ≤ .05 level.

Q15.

(a) **[AO3 = 2]**

1 mark for a brief explanation

Plus

1 mark for elaboration

Possible content:

Control for individual differences so that the researcher can be more certain that the effect is not due to characteristics such as gender, personality etc.

Credit other relevant advantages.

(b) **[AO2 = 2]**

1 mark - Value of 'S' is 2.

Plus

1 mark for explanation - 'S' is the frequency of the least common difference. There are 8 positive differences, 2 negative differences and 2 ties.

(c) **[AO2 = 2]**

1 mark for identifying the correct value of N (N = 10, total number of differences)

STEMUP.CO.UK

Plus

1 mark for stating that for an N of 10, a value of 'S' = 2 is not significant at the p ≤ .05 level.

Q16.

Please note that the AOs for the new AQA Specification (Sept 2015 onwards) have changed. Under the new Specification the following system of AOs applies:

- AO1 knowledge and understanding
- AO2 application (of psychological knowledge)
- AO3 evaluation, analysis, interpretation.

Although the essential content for this mark scheme remains the same, mark schemes for the new AQA Specification (Sept 2015 onwards) take a different format as follows:

- A single set of numbered levels (formerly bands) to cover all skills
- Content appears as a bulleted list
- No IDA expectation in A Level essays, however, credit for references to issues, debates and approaches where relevant.

(a) **AO1 = 2**

Content analysis is a technique for analysing qualitative data of various kinds. Data can be placed into categories and counted (quantitative) or can be analysed in themes (qualitative).

Award 1 mark for a brief statement and a further mark for elaboration.

(b) **AO3 = 4**

- The psychologist could have begun by watching some of the film clips of driver behaviour.
- This would enable the psychologist to identify potential categories which emerged from the data of the different types of distractions seen in the film.
- Such categories / themes might include: passenger distractions, gadget distractions, etc.
- The psychologists would then have watched the films again and counted the number of examples which fell into each category to provide quantitative data.

Credit variations in so far as they explain the process.

Note: maximum 1 mark if no engagement with the stem.

AO3 Mark bands
4 marks Effective Effective explanation of the processes involved in content analysis referring to some or all of the above points.
2 – 3 marks Reasonable

STEMUP.CO.UK

Reasonable accurate coverage of the processes involved.
1 mark Basic Basic identification of the processes involved in content analysis (‘watching the films and counting’).
0 marks No creditworthy material.

(c) **AO3 = 3**

1 mark for identification of an appropriate way of assessing reliability in this investigation. By far the most likely answers here are inter-rater reliability or test-retest reliability.

2 marks for some explanation / elaboration: ‘the two psychologists could carry out content analysis of the films separately and compare their answers’ or ‘they could re-code the films at a later date and compare the two sets of data’.

3 marks for an accurate and clear explanation which refers to deriving the categories and checking the data. ‘The two psychologists could watch the films separately and devise a set of categories. They could compare these and use categories they both agreed on. They could carry out content analysis of the films separately and compare their answers looking for agreement’.

(d) **AO3 = 3**

Candidates can cover one reason explained in detail here or several reasons in less detail.

A repeated measures design was chosen in this experiment:

- to remove the effects of individual differences in reaction times which would occur if an independent groups design was used
- to avoid the potential difficulties involved in matching participants
- to reduce the number of participants required for the experiment.

(e) **AO3 = 3**

This is a repeated measures design and is counter-balanced hence points about order effects and individual differences will not gain credit.

There are a range of potential extraneous variables here including:

- the nature and content of the conversation with the psychologist on the hands-free phone
- interaction between the sex of the psychologist and sex of participant which could influence the type of conversation
- the number of hazards in the computer-based test, hence difficulty of the tests
- the presence of the hands-free headset could have produced distraction.

STEMUP.CO.UK

Award 1 mark for basic identification of a confounding variable and a further 2 marks for elaboration of how this could have affected the dependent variable.

Example: The chat with the psychologist was not controlled (1 mark) so the difficulty or number of questions could have varied (2 marks). This would influence the DV as more or less attention would be required (3 marks).

(f) **AO3 = 4**

There are several potential ethical issues here. Candidates can focus on one in detail or several in less detail.

- Protection of participants from harm whilst studying the effects of a hands-free phone on driving. Two key issues here are the use of a computer-based test with no risk attached and of an experienced sample of police drivers.
- Informed consent: Participants should be given full information about the nature of both tasks before deciding whether or not to participate.
- Debriefing: A full debriefing should take place at the end of the experiment. This should provide feedback on performance and allow participants to ask questions if they wish to.
- Freedom to withdraw: Participants should be made aware of their freedom to withdraw before and during the experiment. They should be made aware of their right to withdraw their data after the experiment.
- Confidentiality: Individuals should not be identified, but should retain anonymity (use of numbers or initials instead of names).

Lists of ethical issues with no elaboration 1 mark.

AO3 Mark bands
4 marks Sound An appropriate ethical issue is identified and explained in detail. Material is accurate – or several issues are identified and discussed accurately in less detail.
2 – 3 marks Reasonable One or more appropriate ethical issues are identified and discussed. The answer is generally accurate.
1 mark Basic Basic identification of an ethical issue (e.g. ‘right to withdraw’) or very brief answers which lack detail.
0 marks No creditworthy material.

(g) **AO3 = 5**

The standardised instructions should include the following information:

a. You will take part in a simulated driving test which will last for three minutes.

STEMUP.CO.UK

b. Your task will be to identify potential hazards on the road ahead.

c. When you see a hazard, you should press the mouse button as quickly as possible.

d. Whilst you are doing the test, I will chat to you on a mobile phone and I would like you to reply using the hands-free mobile phone headset.

e. Do you have any questions?

For full marks, the instructions should adopt an appropriate formal tone. Instructions which are not suitable to be read out should be awarded a maximum mark of 2.

AO3 Marks bands	**Standardised instructions**
5 marks Effective The standardised instructions provide accurate detail of the procedure in a clear and concise form and participants' understanding is checked.	
4 – 3 marks Reasonable The standardised instructions provide sufficient detail of the procedure in a reasonably clear form.	
2 marks Basic The standardised instructions provide some details of the procedure though these may not be clear.	
1 mark Rudimentary The standardised instructions provide few details of the procedure and may be muddled and or inaccurate. Omissions in the instructions compromise the procedure.	
0 marks No creditworthy material is presented.	

(h) **AO3 = 3**

Students are required to identify an appropriate test and are asked to justify their choice.

Award 1 mark for identification of the Wilcoxon (signed ranks) test. Candidates could receive credit for Sign test or related t test. Note that reasons / justification must be correct for the test supplied.

If an incorrect test is identified **no marks** can be awarded.

Award 1 mark for basic statement of a reason, and a further mark for elaboration, within the context of the experiment or a further reason.

e.g. for Wilcoxon test:

- A repeated measures design was used (1 mark) as drivers take part in both the hands-free phone and non-phone (silent) conditions (1 mark).

- A repeated measures design was used (1 mark) and the data can be

STEMUP.CO.UK

treated as ordinal (1 mark).

Test of difference cannot gain credit.

(i) **AO3 = 2**

Students are told that the difference in reaction times was significant at the $p \leq 0.01$ level.

Award 1 mark for a basic understanding of this ('the result is highly significant') and a further mark for elaboration e.g. identifying that the probability of a Type 1 error here is less than 1 / 100.

(j) **AO3 = 3**

Replication is an important tool in the scientific method. It allows scientists to check findings and ensure that they are robust. In this study, replication is important, as the original sample is small (30 people) and specific (experienced police drivers). For this reason, replication on a larger sample will be used to check if findings apply outside this specific group.

Award 1 mark for a general answer on the importance of replication to check findings.

Q17.

Please note that the AOs for the new AQA Specification (Sept 2015 onwards) have changed. Under the new Specification the following system of AOs applies:

- AO1 knowledge and understanding
- AO2 application (of psychological knowledge)
- AO3 evaluation, analysis, interpretation.

Although the essential content for this mark scheme remains the same, mark schemes for the new AQA Specification (Sept 2015 onwards) take a different format as follows:

- A single set of numbered levels (formerly bands) to cover all skills
- Content appears as a bulleted list
- No IDA expectation in A Level essays, however, credit for references to issues, debates and approaches where relevant.

(a) **[AO3 = 1]**

One mark for answers either:

- referring to the strength and the direction of the relationship – a positive correlation between the number of hours spent reading fiction and the empathy test score.
  or:
- describing the relationship – the more hours spent reading fiction, the greater the empathy test score.

No credit for just stating type of correlation eg strong positive.

(b) **[AO3 = 2]**

One mark for naming a test: Spearman's rank order correlation / rho or Pearson's product moment correlation.

STEMUP.CO.UK

One mark for justification. For Spearman's rank order correlation accept: not all data is interval – data collected for empathy test score most likely treated at ordinal level of measurement due to self-report.
For Pearson accept: Pearson's product moment correlation is a robust test, even if not all data can be treated as truly interval.

Just stating ordinal / interval no credit. Accept ordinal or interval providing this is justified with reference to at least one variable.

Unlikely but allow for an informed argument made for treating both sets of data at interval level.

(c) **[AO3 = 2]**

1 mark for a knowledge of a way (not just naming a type of validity) and 2nd mark for explaining how this would be implemented in this case. Most likely answers will address face validity or concurrent validity, but accept any other way such as construct validity, content validity, criterion validity and predictive validity.

For full marks, the answer must refer to either the empathy questionnaire or empathy test items. The 'way' need not be named or defined.

(d) **[AO3 = 2]**

One mark for the identifying a methodological limitation of the study.
Likely answers: size / composition of sample / one school only; for test of empathy – no evidence of testing reliability; parental involvement in 'time spent reading questionnaire'; self-report measures; correlation study.
One mark for a brief explanation.
Suggested explanations might cover: limits to generalisation; confidence in a test and its findings rests on it being deemed reliable; social desirability of parental responses and consequent bias; honesty of reporting / memory recall; cause and effect issues in correlation studies.
Accept any other plausible answers.

(e) **[AO3=3]**

Up to three marks for a discussion of reasons for correlation studies rather than experiments when investigating behaviour.
Likely answers: unethical / impossible to manipulate these variables (reading and empathy in children) to investigate cause and effect; impractical to sometimes do an experiment; may discover a link between two existing variables which might suggest future research ideas; interested in relationships **rather than** a causal explanation.
Accept comparison with the experimental approach.
For full marks, the answer must be coherent and applied to this study.
Maximum of two marks for general answers not applied to this study.

(f) **[AO3 = 8]**

Up to 8 marks for answers demonstrating an ability to design an experiment effectively. Answers should refer to:

- clearly identified independent and dependent variables and at least one extraneous variable identified and control suggested;
- the experimental design – independent groups, repeated measures or

STEMUP.CO.UK

matched pairs;
- detail of sample;
- materials required for carrying out the research, eg task for assessing levels of recall, timing device if needed;
- sufficient procedural details to carry out a replication (might include standard instructions, ethics, etc.)

Note: standardised instructions and ethical issues are not required for full marks.

Mark bands

8 – 7 marks	**Very good answers** All 5 points well addressed and some sound justification. Answer shows sound knowledge and understanding and an ability to design an appropriate experiment. The proposal is coherent and feasible, and includes details of all the essential elements of the chosen design. Information allows for clear understanding of the proposed design. There may be some minor omission(s) at the bottom of the band.
6 – 5 marks	**Good answers** 3 or 4 points well addressed and some justification. The design shows knowledge and understanding and some ability to design an appropriate experiment. The proposal is feasible but may lack the clarity and coherence of the top band. There may be some inaccuracies and omissions.
4 – 3 marks	**Average to weak answers** At least 3 points are addressed and attempt at justification. The answer shows some knowledge and understanding but detail of the proposal may lack clarity. There are inaccuracies and omissions.
2 – 1 marks	**Poor answers** 1-2 points are addressed. There must be some relevant material. The experimental method may not be obvious. There may be substantial confusion, inaccuracy and / or irrelevance.
0 marks	**No relevant content**

Q18.

Please note that the AOs for the new AQA Specification (Sept 2015 onwards) have changed. Under the new Specification the following system of AOs applies:

- AO1 knowledge and understanding
- AO2 application (of psychological knowledge)
- AO3 evaluation, analysis, interpretation.

Although the essential content for this mark scheme remains the same, mark schemes for the new AQA Specification (Sept 2015 onwards) take a different format as follows:

- A single set of numbered levels (formerly bands) to cover all skills
- Content appears as a bulleted list

STEMUP.CO.UK

- No IDA expectation in A Level essays, however, credit for references to issues, debates and approaches where relevant.

(a) **AO2 / AO3 = 2**

Award 2 marks for an appropriate non-directional hypothesis which is operationalised. 'There is a relationship between happiness scores on a questionnaire and intelligence test scores'.

Award 1 mark for a non-directional hypothesis which is not fully operationalised or lacks clarity ('there is a relationship between happiness and intelligence').
Award no marks for a null or directional hypothesis, or one that predicts a difference / link / association / connection.

(b) **AO2/AO3 = 4**

An interview is the most likely answer. An interview would be a more appropriate method than a questionnaire as it enables questions to be clarified and responses to be probed, thus overcoming the main disadvantages of questionnaires.

Students could also make a case for the analysis of diaries/written materials as a way of collecting data about happiness. These would generally overcome the problems of social desirability and demand characteristics inherent in questionnaires. Students could also make a case for the use of observation.

Award one mark for identifying an appropriate method. Award up to three further marks for an explanation of why this method would be better than a questionnaire.

(c) **AO2/AO3 = 2**

Award 1 mark each for any two of the following reasons:

- Study is looking for a correlation (relationship)
- Suitable for pairs of scores
- The data type obtained is ordinal, at least ordinal or interval level
- Linear relationship between scores.

(d) **AO2/AO3 = 3**

Students should state that the obtained value of + 0.42 exceeds the critical value for a twotailed test (.362) for N = 30. The results are therefore statistically significant ($p \leq 0.05$) Award 2 marks for a student who supplies two pieces of information. Award 1 mark for a student who states that the results are significant but does not provide an explanation OR the student who states results are significant but uses incorrect values from the table. Award 0 marks for students who argue that results are not significant.

(e) **AO2/AO3 = 4**

This question requires students to interpret a further correlation co-efficient (this time demonstrating a non-significant negative correlation) and put both findings together. For full marks, answers should cover the two key bullet

points below:

- At age 11, there is a significant positive correlation between happiness and intelligence, demonstrating that more intelligent children tend to be happier

- At age 16, the correlation is not statistically significant.

Students may also make the point that there may be a weak tendency for more intelligent teenagers to be less happy at 16 years of age, although this is not statistically significant. Students may also refer to the contradiction in the results or provide an overall conclusion.

AO2 / AO3 Mark bands
4 marks Effective Effective analysis and understanding. The answer includes the findings of the two studies which are expressed clearly and fluently with appropriate reference to intelligence and happiness. Effective use of statistical terminology.
3 marks Reasonable Reasonable analysis and understanding. The answer is generally focussed and includes reference to both of the key findings which are reasonably clear. There is reasonable use of statistical terminology.
2 marks Basic Basic, superficial understanding. The answer is sometimes focussed OR covers only one of the key conclusions. Expression of ideas lacks clarity. Limited use of statistical terminology.
1 mark Rudimentary Rudimentary with very limited understanding. The answer is weak, muddled and may be mainly irrelevant. Deficiency in expression of ideas results in confusion and ambiguity. The answer lacks structure, often merely a series of unconnected assertions.
0 marks No creditworthy material is presented.

Q19.

(a) **AO2 / AO3 = 3**

A suitable non-directional hypothesis would be ‘There is a correlation (relationship) between pupils’ scores on a test of mathematical ability and pupils’ scores on a test of musical ability’.

3 marks for a fully operationalised non-directional hypothesis.
2 marks for non-directional hypothesis that identifies both variables but does not operationalise them.
1 mark for non-directional hypothesis where the variables are not identified.
No marks for a null or directional hypothesis or one referring to association or difference.

(b) **AO2 / AO3 = 3**

STEMUP.CO.UK

The main issue is that the teacher has made up her own test:

- This involved subjective judgement on the part of the teacher who rates the students' musical ability. Her judgement may not reflect real differences in musical ability and is likely to differ from other people's judgement and / or any absolute criteria for tunefulness.
- Lack of reliability in rating musical ability would compromise the validity of the measure.
- As the students can choose the song they will sing, the rating of ability could reflect the teacher liking / dislike of the song rather than the student's ability.
- The rating may be invalid as the students selected songs which varied in difficulty so the tunefulness reflected the difficulty of the song not the students' ability.
- Operationalising musical ability as tuneful singing is a very narrow measure. Someone can have musical ability such as playing an instrument which would not be reflected by this measure.

1 mark for identifying an appropriate reason.
2 further marks for elaboration, explanation of why it is a problem, how it might affect the result or for further reason(s).
Note that 3 marks can be awarded for one reason elaborated or more than one reason in less detail.

(c) **AO2 / AO3 = 3**

In the case of the maths test candidates could refer to split half or test retest as methods of checking reliability. They could also refer to checking the reliability of scoring by using two separate markers for the test and comparing the scores. Credit any other appropriate suggestion.

1 mark for identifying an appropriate method or a brief explanation eg 'repeat the maths test'.
2 further marks for appropriate elaboration.

(d) **AO2 / AO3 = 2**

The teacher chose to use a random sample because it would probably be more representative of the whole GCSE group than if she had used an opportunity or volunteer sample. Candidates could also say that she had ready access to her target population making it convenient for her to select a random sample.

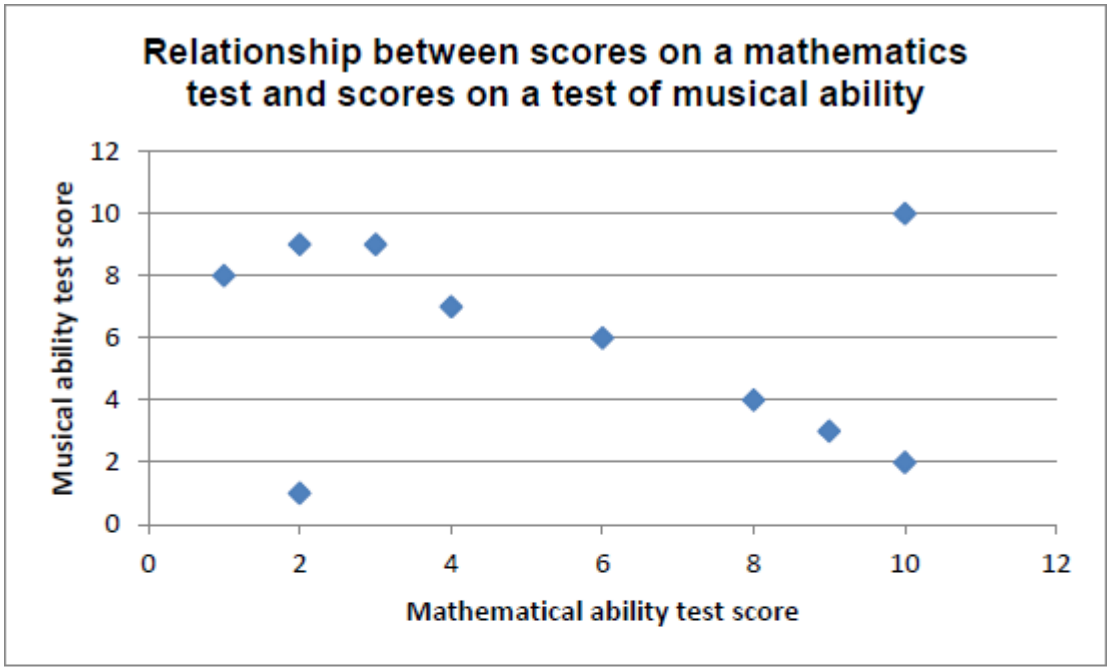
No credit for definition of a random sample.
1 mark for a brief or muddled reason (it is not biased).
2 marks for a reason that clearly points to an advantage of random sampling. This could be achieved through a comparison with another method (it is less likely to be biased than a volunteer sample).

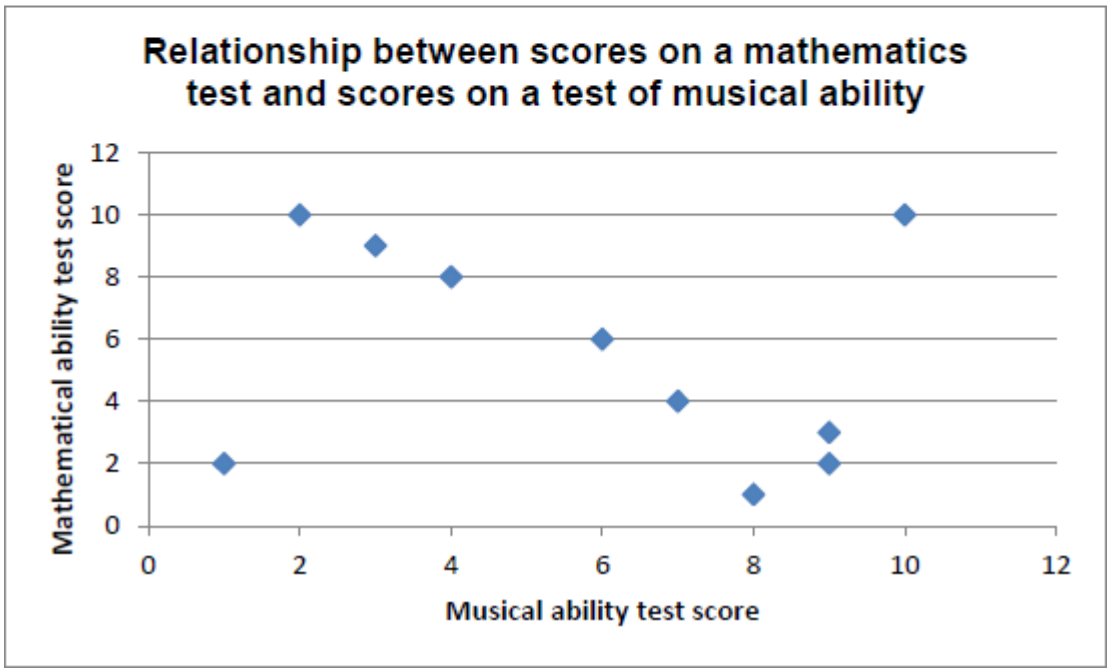
(e) **AO2 / AO3 = 3**

Credit should only be awarded for scattergraphs. Other graphs gain 0 marks.

1 mark for appropriately plotted scores.
1 mark for an appropriate title.
1 mark for correctly labelled axes.

STEMUP.CO.UK





(f) **AO2 / AO3 = 3**

Up to 3 marks for a discussion of the relationship between mathematical and musical ability. Likely points include:

- The graph seems to show a negative correlation between mathematical and musical ability.
- This means that high scorers in mathematical ability tend to achieve low scores on musical ability and vice versa.
- The presence of two strong outliers, means that the actual correlation is very weak and closer to zero.
- Comment on the small sample size which limits the conclusions that could be drawn.

STEMUP.CO.UK

- Credit can be achieved for plausible interpretations of the strength of the correlation which are justified (ie looks moderate to strong or the outliers make it weak in practice) or those based on rough calculations (around -0.2).

1 mark for a very brief answer eg negative correlation or zero correlation.
2 further marks for elaboration/discussion this could be focused on one point in detail or several points in less detail.

(g) **AO2 / AO3 = 10**

In this question, candidates are asked to design a study to test if there is a difference between left-handed and right-handed students in musical ability.

Design – 1 mark

- Award 1 mark for identification of an appropriate design (independent measures or matched pairs).

Sampling – 2 marks

- Award 1 mark for explaining an appropriate sampling method and 1 further mark for justifying why this method would be appropriate. As left-handed people are less common in the population than right-handed people this needs to be addressed in the sampling method.

Procedure and assessment of musical ability – 4 marks

Award 1 mark for procedure, 1 mark for assessing musical ability and two further marks for elaboration of either or both of these.

- Description of the procedure eg each participant will be given a standardised musical ability test, participants should be tested within a controlled environment, with minimal noise or distraction.

- Students are required to suggest a plausible alternative method of assessing musical ability to the one in the stem (eg singing a short, novel phrase played on the piano). Further credit could be given for stating that the test should be identical for all students or for explaining how it will be assessed.

Debrief – 3 marks

- Award up to 3 marks for writing a debrief. This could include the aim of the study, thanking participants for taking part, asking if they have any questions, relevant ethical considerations.

- If this is not suitable to be read out to participants, maximum 1 mark.

(h) **AO2 / AO3 = 3**

Award 1 mark for a clear table appropriate for the study described in (h).

Musical ability scores:

Participant number	Left handed	Right handed
1		

STEMUP.CO.UK

2		
3		

Award 1 mark for the identification of an appropriate statistical test for the proposed design.
Award 1 mark for one correct justification eg a test of difference, at least ordinal level data.

Q20.

Please note that the AOs for the new AQA Specification (Sept 2015 onwards) have changed. Under the new Specification the following system of AOs applies:

- AO1 knowledge and understanding
- AO2 application (of psychological knowledge)
- AO3 evaluation, analysis, interpretation.

(a) **[AO3 = 2]**

One mark for an explanation.
Inform researcher of spread of scores.
One mark for link to the study.
The programme did not seem to affect people in each condition differentially as spread of scores in each condition is similar / large.
The offenders' anger behaviour showed a wide variation both before and after the programme.
Accept spread is greater in the after condition than before.

(b) **[AO3 = 2]**

One mark for statement that the result is significant.
One mark for rationale: the calculated value of *T* is 22 and is less than the critical value of 25 (at the 0.05 level of significance).
Can accept not significant at 0.02 level.

(c) **[AO3 = 3]**

Maximum of 3 marks can be obtained from: one mark for each reason or two marks for each reason with explanation.

- Reason – ordinal level of measurement / non-parametric
  Explanation – self reports / estimated scores of anger; data might not be normally distributed
- Reason – design of the study is related / repeated measure
  Explanation – same people before and after
- Reason – research involving differences between the 2 sets of scores
  Explanation – anger scores before and after

(d) **[AO3 = 4]**

Up to two marks for each reason and explanation. Likely points: as an aid to memory; a qualitative measure to supplement the quantitative data collected; to check the validity of the questionnaire; part of the therapeutic process /

STEMUP.CO.UK

increased self-awareness.
Accept other valid reasons.
One mark for an appropriate reason and one mark for an explanation of the reason.

(e) **[AO3 = 3]**

Up to three marks for outlining how a control group could have improved this study: it is not possible to tell if the programme has caused the improvement; improvement could have been due to the programme or due to spontaneous recovery; by using a control group would make it more scientific; scores can be taken at the same times (pre-programme / post-programme) as in an experimental condition; post programme differences between the groups can inform if programme is effective; can be more confident in inferring cause and effect .

Allow a maximum of one mark for the general purpose of a control condition: acts as comparison / baseline measure where nothing changes

Accept 'scientific' and 'validity' only if justified.

(f) **[AO3 = 5]**

Up to 5 marks for addressing both reliability and validity. One of these marks must be for reference to statistical testing.

A maximum of three marks if only one of these is addressed.

One mark for identifying a type of validity: face validity; concurrent validity.
Accept also content validity; criterion validity; predictive validity.
Only accept identification mark if it matches how the assessment would be carried out.

One mark for outlining how the assessment would be carried out. For example in concurrent validity, scores from the questionnaire are compared with those from an established but similar questionnaire known to have good validity to see if the results are similar.

One mark for the statistical testing (checking for a positive correlation / applying Spearman's rank order correlation).

One mark for identifying a way of assessing reliability. Most likely is test-retest but accept split-half reliability and item analysis.
Only accept identification mark if it matches how the assessment would be carried out.
Do not accept inter-rated / inter-observer reliability.

One mark for outlining how the assessment would be carried out. For example in test-retest, the same group of young offenders would be tested using the same questionnaire at a later date to see if the findings remained consistent.

One mark for the statistical testing (checking for a positive correlation / applying Spearman's rank order correlation).

The one mark for statistical testing can only be credited once.

Q21.

STEMUP.CO.UK

Please note that the AOs for the new AQA Specification (Sept 2015 onwards) have changed. Under the new Specification the following system of AOs applies:

- AO1 knowledge and understanding
- AO2 application (of psychological knowledge)
- AO3 evaluation, analysis, interpretation.

(a) **AO2 / AO3 = 1**

One mark for an accurate reason: The decision to use a directional hypothesis was based on findings of previous research which pointed to an effect in a particular direction ie memory is poorer with age.

(b) **AO2 / AO3 = 3**

A suitable directional hypothesis would be 'There is a negative correlation (relationship) between age and recall accuracy rating'.

- 3 marks for a fully operationalised hypothesis as above
- 2 marks for a directional correlational hypothesis that identifies age and recall as the two variables but is not fully operationalised
- 1 mark for a directional hypothesis where the variables are not identified (‘there will be a negative correlation’) or where the hypothesis lacks clarity.

Award zero marks for a non-directional or null hypothesis or any hypothesis predicting a difference or association.

(c) **AO1 = 1**

One mark for an accurate definition: The extent to which results or procedures are consistent or simply 'consistency'.

(d) **AO2 / AO3 = 3**

One mark for identification of a way of ensuring reliability. By far the most likely answer here is inter-rater reliability.

Two marks for some explanation/elaboration: using two separate psychologists and comparing them.

Three marks for an accurate and clear explanation: using two separate psychologists to rate the typed accounts for accuracy and comparing / correlating the ratings to see how similar they are.

Candidates could make a case for test retest which would involve the same psychologist re-examining the ratings after a period of time.

(e) **AO2 / AO3 = 2**

Award one mark for correct identification of one of each type of data.

- Qualitative data: the patient’s responses, the typed accounts, the doctor’s notes.
- Quantitative data: the ratings of recall accuracy on a scale of 1 – 10, ages of patients.

STEMUP.CO.UK

(f) **AO2 / AO3 = 2**

One mark for each accurate reason given:

- the researchers are testing for a correlation or a relationship between two variables.

- the data is to be treated as ordinal because the recall accuracy is in the form of ratings.

(g) **AO2 / AO3 = 2**

One mark for stating that the result is significant.

Second mark for explaining that -.52 exceeds .306 ($p \leq 0.05$, n=30 for a one-tailed test).

(h) **AO1 = 2**

One mark for a brief or muddled answer which hints at rejecting HO / accepting the H1 in error.

Two marks for explaining the term: where the researcher rejects the null hypothesis (or accepts the research / alternative hypothesis) when in fact the effect is due to chance – often referred to as an error of optimists.

(i) **AO2 / AO3 = 3**

3 marks for a clear explanation which is based on comparison of the calculated value of rs with the critical value at the 0.01 level of significance and indicates competence in use of statistical tables as follows:

- A Type 1 error is unlikely because the calculated value of rs (-0.52) exceeds the critical table value at both the 0.05 and 0.01 level for a one-tailed test.

- The chance of a Type1 error occurring is therefore less than 1%.

- This means that the researchers can be 99% certain that the results obtained are not due to chance.

Award one mark for a brief explanation (it is significant at 0.01).
Award two further marks for an explanation which refers to two of the above points.

Award one mark for stating that the obtained value (-0.52) exceeds the critical value (0.306) by a reasonable margin.

Q22.

Please note that the AOs for the new AQA Specification (Sept 2015 onwards) have changed. Under the new Specification the following system of AOs applies:

- AO1 knowledge and understanding
- AO2 application (of psychological knowledge)
- AO3 evaluation, analysis, interpretation.

(a) **AO2 / AO3 = 4**

Up to four marks are awarded for discussing

STEMUP.CO.UK

advantage(s) of using a laboratory experiment in this case.

The most likely advantages of the laboratory setting in this experiment include:

- Control over extraneous variables. The lab setting meant that extraneous variables could be minimised. In this experiment, outside factors such as waiting time, noise and stress (which would be difficult to control in a field experiment) were removed.

- Ethical issues. In this case, the testing of memory in a field experiment would have involved ethical issues including deception of patients or withholding of information.

Candidates may also refer to other advantages of the laboratory setting such as replicability. These can receive full credit if they contextualised within the scenario.

Award four marks for an answer which provides accurate and detailed discussion of relevant advantage(s) with a clear link to the scenario.

Award two or three marks for an answer which includes discussion of relevant advantage(s), with some reference to the scenario.

Award one mark only for an answer which merely identifies one or more relevant advantage(s) of a laboratory experiment appropriate to this scenario.

Advantages of laboratory experiments which are not relevant to this study cannot gain any credit eg use of technical equipment.

(b) **AO2 / AO3 = 2**

- One mark for correctly identifying the Mann Whitney U test or independent t test.

- One mark awarded for an accurate reason for choice (for Mann Whitney these are: test of difference, independent groups design / independent data or data which can be treated at an ordinal level).

Q23.

Please note that the AOs for the new AQA Specification (Sept 2015 onwards) have changed. Under the new Specification the following system of AOs applies:

- AO1 knowledge and understanding
- AO2 application (of psychological knowledge)
- AO3 evaluation, analysis, interpretation.

(a) **AO2 / AO3 = 2**

'There is an association between birth order and choice of career' = 2 marks.
A directional hypothesis is not credit-worthy. Reference to a relationship / correlation cannot gain credit.

Although technically, the psychologist is looking for an association, candidates can gain credit for expressing the hypothesis in terms of a difference eg 'There is a difference in career choice depending on birth order.'

2 marks for a clear hypothesis, 1 mark for a hypothesis which lacks clarity.

STEMUP.CO.UK

(b) **AO2 / AO3 = 3**

One mark for identifying a sampling method.
One mark for a brief explanation of how to obtain the sample eg 'by advertising for lawyers or artists to come forward'. One further mark for elaboration eg 'by explaining that adverts would have to be placed in appropriate journals etc to attract these particular categories of participants'

Candidates who identify a sampling method but describe it incorrectly can be awarded 1 mark.

(c) **AO2 / AO3 = 12**

This is a 12 mark question but marks are allocated to each of the required components as follows:

- An appropriately labelled table = 2 marks

1 mark for a table that displays the data in the question.
2 marks for a table which includes data relating to non first-born children. Totals are not required for the 2 marks.

Table: Table to show the career choices of first born and non-first born children

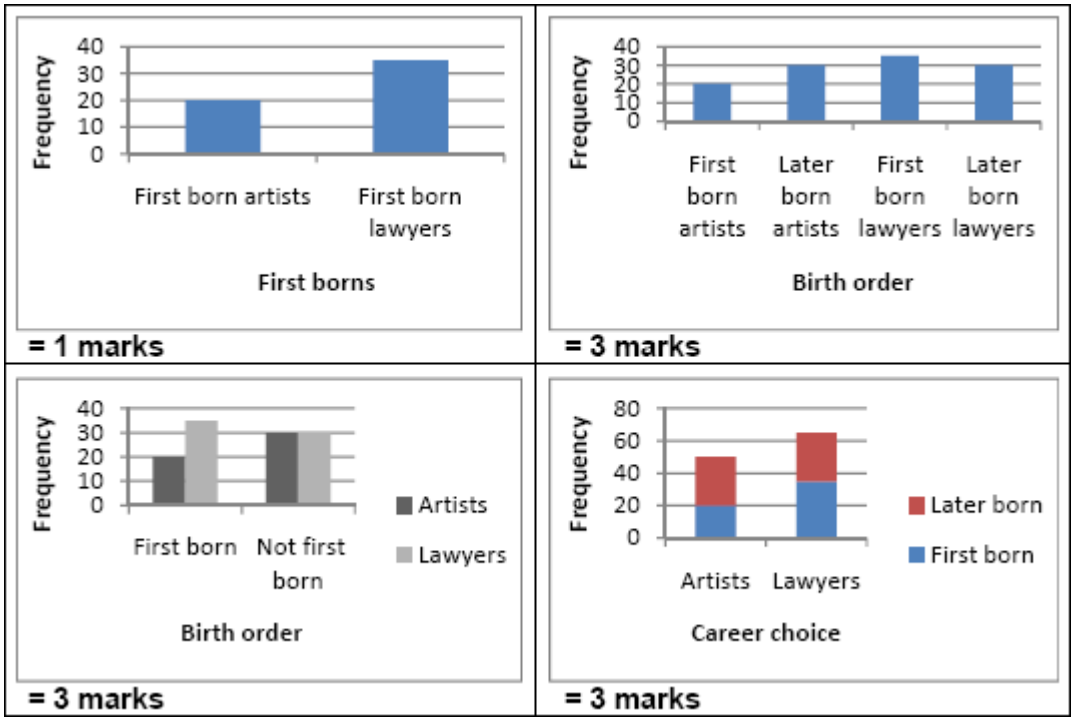
	Artists	Lawyers	Totals
First born	20	35	55
Not first born	30	30	60
Totals	50	65	115

- a sketch of an appropriately labelled bar chart = 3 marks

For 3 marks, candidates need to display the data relating to first born and non-first born career choices on a bar chart. They should label axes correctly and draw the columns to the correct approximate height for a sketch.

For 2 marks, candidates display data as above but labels are missing or lack clarity.
For 1 mark, candidates graph the data supplied in the question relating to first born career choices only.

STEMUP.CO.UK



NB Labelled axes but no bars = 0 marks.

- identification of appropriate statistical test and justification = 1 + 2 marks

An appropriate test here is the Chi-squared.
Justification gains 2 marks. Any two correct reasons from:

- data are independent
- level of measurement is nominal
- test of association / difference is required.
- identification of appropriate significance level = 1 mark.

The most likely significance level is 5% ($p \leq 0.05$). Candidates are not asked to justify their choice. Candidates who choose a more stringent level can achieve marks but they must then follow this through when they make their statement of results.

Candidates who erroneously report 0.05% or $p = 0.5$ do not gain credit for level of significance but can achieve credit for the statement of results in relation to the hypothesis.

- a statement of the results of the statistical test in relation to the hypothesis = 3 marks.

For full marks, the candidate should state whether or not they can accept the hypothesis (or they can express this in terms of rejecting the null hypothesis) at a given significance level and refer to the observed and critical values.

Where candidates choose an inappropriate value from the table but interpret that value correctly, they can gain 2 marks.

The critical value for x^2 (df =1 p 0.05 (two-tailed)) is 3.84. As the observed value of x^2 2.27 is less than the critical value, we cannot reject the null hypothesis. There is not an association between birth order and career choice.

STEMUP.CO.UK

Q24.

Please note that the AOs for the new AQA Specification (Sept 2015 onwards) have changed. Under the new Specification the following system of AOs applies:

- AO1 knowledge and understanding
- AO2 application (of psychological knowledge)
- AO3 evaluation, analysis, interpretation.

Although the essential content for this mark scheme remains the same, mark schemes for the new AQA Specification (Sept 2015 onwards) take a different format as follows:

- A single set of numbered levels (formerly bands) to cover all skills
- Content appears as a bulleted list
- No IDA expectation in A Level essays, however, credit for references to issues, debates and approaches where relevant.

(a) **AO2/3 = 6**

Candidates need to show that they understand what differentiates opinion from scientific evidence. They could mention some of the following:

- The teacher has only experienced one school in a particular catchment area so she has only observed a very limited number of 5 year-olds (issues of sampling and replicability).
- She has found out that children do not eat anything nourishing simply by chatting with the children. She has no corroborative evidence from eg parents (issues of objectivity).
- She uses vague phrases such as ‘decent breakfast’ without being clear what this means (operationalisation).
- She has generated a theory and made predictions based on flimsy evidence.
- She has not used any scientific method to lead to her conclusions eg a carefully controlled experiment, survey or observation.
- She has drawn conclusions about the effects of breakfast without considering other variables which might affect reading skills and behaviour.

AO2/3 Mark bands
6 marks Effective Explanation demonstrates sound understanding. Application of knowledge is effective and shows coherent elaboration. Ideas are well structured and expressed clearly and fluently. Consistently effective use of psychological terminology.
5 – 4 marks Reasonable Explanation demonstrates reasonable understanding. Application of knowledge is reasonably effective and shows some elaboration. Most ideas appropriately structured and expressed clearly. Appropriate use of psychological terminology.
3 – 2 marks Basic Explanation demonstrates basic, superficial understanding. Application of knowledge is basic. Expression of ideas lacks clarity. Limited use of psychological terminology.

STEMUP.CO.UK

1 mark Rudimentary Explanation is rudimentary, demonstrating very limited understanding. Application of knowledge is weak, muddled and may be mainly irrelevant. Deficiency in expression of ideas results in confusion and ambiguity. The answer lacks structure, often merely a series of unconnected assertions.
0 marks No creditworthy material is presented.

(b) **AO2/3 = 3**

In a random sample, every member of the identified population has an equal chance of selection. In this case, the sampling frame consists of the 400 five-year-old children attending ten local schools. In order to obtain a simple random sample, the researcher has to have the names of all 400 children and can then select using one of the following methods:

- **Random number tables** – random number tables are specially devised to meet the following criteria – they contain strings of numbers where each number has the same chance of being selected as any other and each number is independent of the others.
  Such tables are readily available in statistics text books etc or can be generated by the researcher using a computer program. The researcher assigns each child a number between 1 and 400. He enters the table at any place (he could close his eyes and point with a finger at a starting place) and then moves either horizontally or vertically to produce a string of random numbers. He records all the numbers which correspond to the 400 children until he has recorded a total of 100 non-duplicated numbers.

- **Computer selection** – This is a similar method where the computer does most of the work. A computer can generate an endless string of random numbers ie numbers which have no relationship to one another as a sequence. Each child's name is given a number and a random number generator program is used to produce the required sample size (in this case 100 participants).

- **Manual selection** – Using this method, the researcher has to put each name (or an assigned number) on a separate slip of paper and place them all in a container. The researcher then selects 100 slips from the container. The following conditions could apply: the container should be shaken between each draw; the slips of paper should all be the same size and folded in the same way so that one does not feel different from another; the selector draws 'blind' ie cannot see the actual slips of paper.

A simple definition of a random sample is not credit-worthy since it offers no explanation.\ Similarly, answers which only use the word 'random' as an explanation cannot gain credit eg He would choose 100 participants at random from the children. One mark for a very basic method eg 'he would take names from a hat / computer / random number table'. Two further marks for elaboration.

(c) **AO2/3 = 3**

Candidates could focus on:

- Even if a sample is random, it may not be truly representative of the population eg might all come from the same school, or be all boys or all girls.

STEMUP.CO.UK

- Practical limitations eg the time and effort needed to write out 400 slips for the manual method.

- Difficulties of obtaining a truly random sample eg even if the sample is selected randomly, parents might refuse to allow their children to participate.

Any plausible and appropriate answers should be credited.
Up to 2 marks for identification of limitations. For 3 marks, one or more limitations must be explained in reasonable detail.

(d) **AO2/3 = 5**

There are two requirements to this question, **why** operationalising variables is important and **how** to operationalise the IV and the DV. If a candidate only explains **how / why**, maximum 3 marks.

The terms' 'decent breakfast' and 'reading skills' are vague. It is important from the point of view of objectivity, replicability and control of extraneous variables to make sure that these terms are closely defined.

Suggestions as to how the psychologist might do this could include the following:

The researcher needs to specify the exact composition of the breakfast (possibly by doing a pilot study or a literature search to identify the components of breakfast most likely to bring about behavioural / cognitive change). He probably also needs to specify the time at which it is consumed. The researcher needs to use a standard reading test which should be administered to all the participants at the beginning of the study and at the end – the dependent variable is likely to be the improvement score.

(e) **AO2/3 = 2**

Reasons are:

- a test of difference

- data (scores from a reading test) are at least ordinal, this would include ordinal / interval and / or ratio

- independent design.

One mark for each appropriate reason (maximum 2 marks).

(f) **AO2/3 = 2**

It would have been more difficult to use a matched-pairs design because of the number of relevant factors that would need to be controlled (eg gender, intelligence, parental attitudes / income / education, experience of pre-school education, number of siblings in family etc). There is a relatively small pool of children available (ie 400) and it could be difficult to match on all these factors. It would also be very time-consuming; it could be quite expensive to carry out the necessary surveys; it could be quite intrusive collecting such information from parents.

One mark for a basic explanation eg "Because it is difficult to match participants appropriately".
One further mark for elaboration.

(g) **AO2/3 = 2**

STEMUP.CO.UK

One mark for identifying an appropriate issue and second mark for explaining how it could be addressed.

The most likely issue is confidentiality which could be addressed by ensuring that all scores on reading scales and all personal information are anonymised.
There are also ethical problems involved in denying the control group breakfast although it is more difficult for candidates to suggest a way of addressing this – perhaps to put only those children into the control group who do not eat breakfast anyway, restricting the study length to a short period of time and, if the study results support the hypothesis, to provide free breakfasts to these children for the rest of the academic year.
Parental consent is excluded because it is given in the stem so answers which offer this as an issue cannot gain credit.

(h) **AO3 = 12**

Design should be written clearly, succinctly and with sufficient detail for reasonable replicability.
Candidates will not receive credit for details included in the stimulus material. These include using a random sample of 100 children, gaining parental consent and selection of a Mann Whitney test.
To access marks in the top band candidates must state an appropriate hypothesis in which “playground behaviour” is clearly operationalised. The hypothesis could be directional or non-directional.
Given the wording of the question, a correlational hypothesis is not credit-worthy, however, the rest of the answer should be marked on its merits.
Likely aspects of “playground behaviour” would include activity levels, aggression, cooperative play etc.
An attempt to operationalise “a healthy breakfast” should be credited. However, candidates could assume this had already been done by the psychologist.

As this is an observational study any of the following, together with appropriate justification, would be credit-worthy:
Is the observation covert or overt?
Where are observers positioned? (In playground, watching from window?)
Is a video recording of the children used? How will this be analysed (eg content analysis)?
Do the students who observe know what the children ate for breakfast?
At what times of day does the observation take place?
How many children are observed? (Candidates could justify using a smaller sub-sample of the 100 children in the original study).
How long does each observation last?
Will the observers use a behavioural check list / tally chart?
Will more than one observer observe each child? If so, what training will be given and what checks for inter-observer reliability will take place?
Reference to time sampling or event sampling.

Credit any other relevant material.

AO2/3 Mark bands
12 – 10 marks Effective design A design that demonstrates sound knowledge and understanding of observational research. The selection and application of research techniques is appropriate. The description provides sufficient detail for most aspects of the study to be implemented. Some design decisions are justified effectively.

STEMUP.CO.UK

9 – 7 marks Reasonable design The design is reasonable and demonstrates knowledge and understanding of some aspects of observational research. The selection and application of research techniques is mostly appropriate. The description provides sufficient detail for some aspects of the study to be implemented. Some design decisions are justified.
6 – 4 marks Basic design The design is basic and demonstrates limited knowledge and understanding of aspects of observational research. The selection and application of research techniques are sometimes appropriate. Some basic design decisions/features of the study are described but there may be significant omissions, lack of clarity and possibly some implausible suggestions that severely limit implementation. Justifications of the design are limited.
3 – 1 marks Rudimentary design The design is rudimentary. Design decisions are muddled and or mostly inappropriate and are not justified. Description lacks clarity. The study could not be implemented.
0 marks No creditworthy material.

Q25.

Please note that the AOs for the new AQA Specification (Sept 2015 onwards) have changed. Under the new Specification the following system of AOs applies:

- AO1 knowledge and understanding
- AO2 application (of psychological knowledge)
- AO3 evaluation, analysis, interpretation.

(a) **AO2 / 3 = 1**

1 mark for correct answer – directional (one-tailed is acceptable).

(b) **AO2 / 3 = 3**

1 mark for correctly stating that the result is significant.
2 further marks for an explanation: the calculated value of T = 53 which is less than the value of 60 where N = 20 and $p \leq 0.05$ for a one-tailed test.

If the candidate states that the result is not significant, no marks can be awarded.

Q26.

Please note that the AOs for the new AQA Specification (Sept 2015 onwards) have changed. Under the new Specification the following system of AOs applies:

- AO1 knowledge and understanding
- AO2 application (of psychological knowledge)
- AO3 evaluation, analysis, interpretation.

(a) **AO2 / 3 = 4**

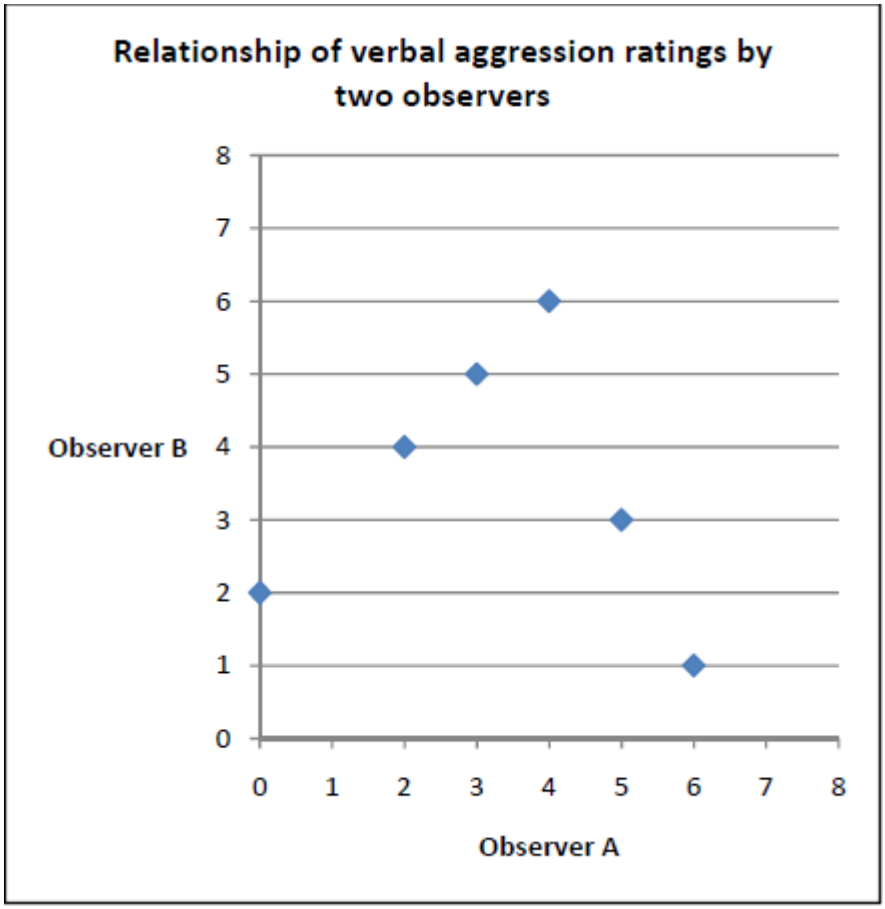
For any credit, candidates must sketch a scattergram.
For full marks, candidates should provide an appropriate title for the scattergram,

STEMUP.CO.UK

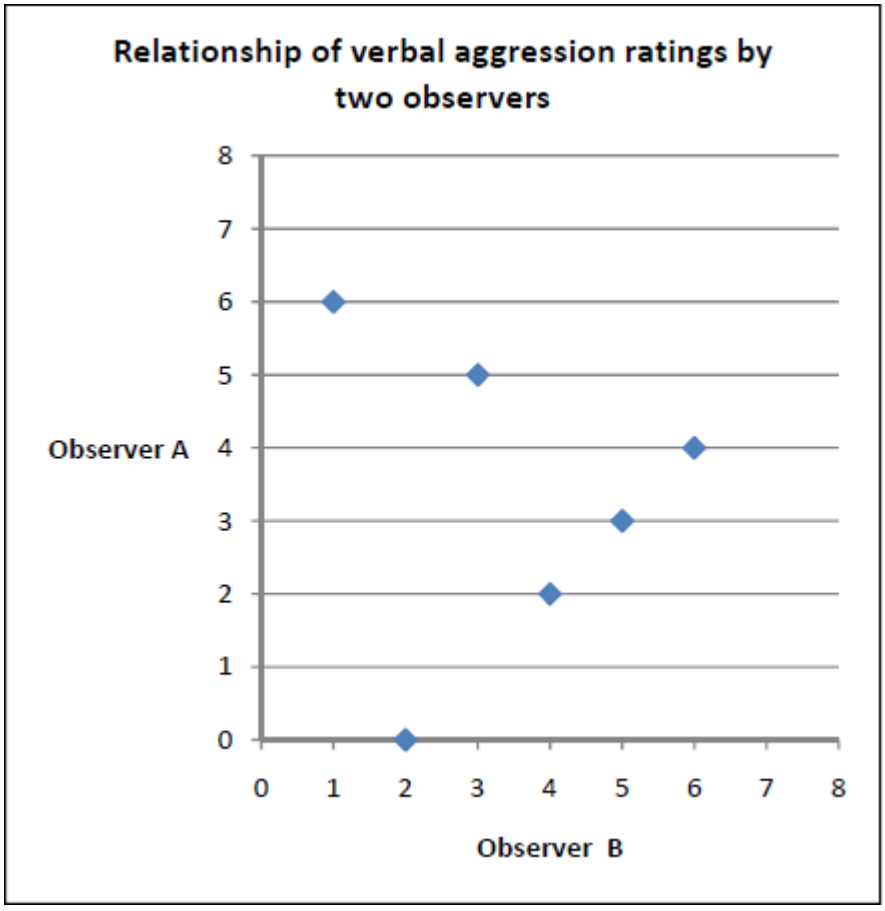
label each of the axes appropriately and plot the data accurately on the scattergram.

Each of the examples below is a full mark answer because:

- it is clearly a sketch of a scattergram
- the data are appropriately plotted
- the labels of the axes and the title taken together show full understanding of the nature of the data.



STEMUP.CO.UK



(b) **AO2 / 3 = 4**

For full marks, candidates should give a reasonably detailed explanation eg she is concerned because the observers should both recognise the same types of verbal behaviour as aggressive and you would expect their tallies to be very similar. In this case, the observers disagree in every 10-minute time interval even though they are both watching the same child and should be using the same criteria. In some time slots, there is a really big difference in the number of acts.
This suggests that the observers have interpreted the criteria differently or that, at certain times, one observer was more vigilant then the other (4 marks).

1 mark – 'because the observers do not agree with each other'.
3 further marks for elaboration.
Candidates who simply describe what is meant by inter-rate reliability can gain no marks.

(c) **AO2 / 3 = 3**

1 mark for identifying the appropriate test – Spearman's Rho or Pearson's (with appropriate justification).
2 further marks for explaining why it is appropriate ie the psychologist is testing for a correlation and the data that can be treated as ordinal.
Candidates can gain no marks on this question if their choice of statistical test is inappropriate.

(d) **AO2 / 3 = 4**

1 mark for a very brief answer eg 'better training for the observers'
3 further marks for elaboration.

STEMUP.CO.UK

There is a breadth / depth trade-off here. Candidates can elaborate on one improvement eg explain how the training might be improved or outline several improvements in less detail eg establish clearer criteria for categorising verbal aggression, filming the child so that the observers can practise the categorisation.

Q27.

Please note that the AOs for the new AQA Specification (Sept 2015 onwards) have changed. Under the new Specification the following system of AOs applies:

- AO1 knowledge and understanding
- AO2 application (of psychological knowledge)
- AO3 evaluation, analysis, interpretation.

Although the essential content for this mark scheme remains the same, mark schemes for the new AQA Specification (Sept 2015 onwards) take a different format as follows:

- A single set of numbered levels (formerly bands) to cover all skills
- Content appears as a bulleted list
- No IDA expectation in A Level essays, however, credit for references to issues, debates and approaches where relevant.

(a) **AO2 / 3 = 2**

One mark for one brief finding and a further mark for appropriate elaboration or for two brief findings or one mark for a slightly muddled answer.
On average, the treatment group showed greater improvement after the treatment than the no-treatment group. The average improvement score for the no-treatment group was very low suggesting that the treatment gains for the treatment group were not simply a result of the passage of time.

There was some variation in both groups as shown by the ranges but it was wider in the treatment group. The low range in the no-treatment group suggests that most people in this group had similar low improvement scores.

(b) **AO1 = 1, AO2 / 3 = 3**

One mark for identification of a suitable test and 3 further marks for an appropriate justification.The specification only requires knowledge of non-parametric tests. However, if a candidate names an independent t-test and justifies its use, this is perfectly acceptable. It is likely that most candidates will identify a non-parametric test. The most appropriate test is the Mann-Whitney and the justifications for its use are:

- independent groups design
- at least ordinal data
- differences.

(c) **AO2 / 3 = 2**

One mark for correctly identifying the likelihood and one further mark for an appropriate explanation or one mark for a slightly muddled answer.

The likelihood of making a Type 1 error is 5%. A Type 1 error occurs when a researcher claims support for the research hypothesis with a significant statistical test, but in fact, the variations in the scores are due to chance variables. If the level of significance is set at 5%, there will always be a one in twenty chance or less that

STEMUP.CO.UK

the results are due to chance rather than to the influence of the independent variable or some other factors.

(d) **AO2 / 3 = 4**

Two marks for each reason. One mark for a basic identification and one further mark for elaboration.

Possible reasons include:

Expectations – the patients might expect the treatment to do them some good and it becomes a self-fulfilling prophesy.

Biased sample – even though the participants were randomly assigned to groups, the treatment group might, by chance have included more people with milder symptoms that were more likely to respond to treatment.

Other support – we do not know what other support/ treatment that the participants might have had over the 8 week therapy period.

(e) **AO2 / 3 = 4**

Two marks for the advantage and two marks for the disadvantage. One mark for simply identifying an advantage / disadvantage and the further mark for elaboration in the context of the study. Answers which are not set in context cannot achieve full marks.

Advantage: Much quicker to administer and to score – could all have been given out at the same time whereas the therapist has to conduct 30 time-consuming interviews; cheaper than interviews, ie in terms of the therapist's time; people might be more comfortable, and, therefore, more honest, if they have to write responses rather than face an interviewer (could work the other way as well – see disadvantages).

Disadvantage: Self-report questionnaires might not yield as accurate data as an interview – questions can limit range of answers and there are no additional cues, eg body language, participants might be less honest on a questionnaire than in a face-to-face interview.

Marks can be awarded for any appropriate advantages / disadvantages.

(f) **AO2 / 3 = 5**

Candidates should demonstrate understanding of some of the requirements of a good consent form. For full marks, it should be succinct, clear and informative.

It is likely to include some of the following information: treatment programme that is noninvasive; requirement to be assessed on current level of functioning; use of a trained therapist to conduct interviews; duration of the programme; requirement for re-assessment at the end of the programme; random allocation to a treatment or no-treatment group.

It should show awareness of ethical considerations, eg

- no pressure to consent – it will not affect any other aspects of their treatment if they choose not to take part
- they can withdraw at any time
- they can withdraw their data from the study

STEMUP.CO.UK

- their data will be kept confidential and anonymous
- they should feel free to ask the researcher any questions at any time
- they will receive a full debrief at the end of the programme.

For full marks, candidates must include a range of both procedural and ethical points.

AO2 / 3 Mark Bands (5 marks)
5 marks Effective Consent form demonstrates sound knowledge and understanding of research ethics.
4 – 3 marks Reasonable Consent form demonstrates reasonable knowledge and understanding of research ethics.
2 marks Basic Consent form demonstrates basic, superficial knowledge and understanding of research ethics.
1 mark Rudimentary Consent form is rudimentary demonstrating very limited understanding of research ethics.
0 marks No creditworthy material is presented.

(g) **AO1 = 2, AO2 / 3 = 2**

AO1: One mark for brief description, eg ’consistency’ and one further mark for elaboration. Reliability refers to consistency over time. If a test, questionnaire, etc, is reliable, people tend to score the same on the test if they take it again soon afterwards.

AO2 / 3: One mark for a very brief answer, eg ‘do another test’ or ‘test them again’ or ‘use another interviewer to check’. Two marks for some elaboration.

Reliability could have been checked by administering a valid and reliable questionnaire to the participants as well as interviewing them and then comparing the scores on the two measures. If the interview score was reliable, there would be strong positive correlation between the scores.

The interviews could have been filmed and given to another trained therapist to assess. A strong correlation between the scores given by each therapist would demonstrate reliability.

(h) **AO2 / 3 = 10**

For full marks, the method section should be written clearly, succinctly and in such a way that the study would be replicable. It should be set out in a conventional reporting style, possibly under appropriate headings. Examiners should be mindful that there are now different, but equally acceptable reporting styles. For example, candidates should not be penalised for writing in the first person. The important factor here is whether the study could be replicated.

There should be reasonable detail with regard to:

- design
- participants
- materials
- procedures.

AO2 / 3 Mark Bands (10 marks)
10 – 9 marks Effective Effective method section that demonstrates sound knowledge and understanding of investigation design. The design decisions are appropriate and the description provides accurate detail of the design, participants, materials and procedure of the study. Effective and appropriate report style.
8 – 6 marks Reasonable The method section demonstrates reasonable knowledge and understanding of investigation design. The design decisions are generally appropriate and the description provides reasonable detail of the design, participants, materials and procedure of the study. Generally appropriate report style.
5 – 3 marks Basic The method section demonstrates basic knowledge and understanding of investigation design. Some aspects of the design are appropriate. The description provides basic detail of some features of the study or rudimentary outline of the main features. Expression lacks clarity.
2 – 1 mark Rudimentary The method section demonstrates rudimentary knowledge or understanding of research. The report is weak, muddled or incomplete. Deficiency in expression results in confusion and ambiguity.
0 marks No creditworthy material is presented.

Q1.

There were many correct answers to part (a) and it was apparent that the majority of students had the knowledge to be able to identify the description of an overt observation.

In part (b) there were many correct answers and it was apparent that the majority of students had the knowledge to be able to identify the description of an overt observation.

In part (c) most students appeared to have a good knowledge of the strengths and limitations of the two types of observation. However, many failed to actually use this knowledge to answer the question on discussing reasons why covert may have been more beneficial than overt. There were many limited responses which either explained limitations of overt observations or strengths of covert observations but without comparison. Equally there were many responses which had implicit discussion of benefits which was not clear and thus not sufficient for full marks.

Q2.

Part (a) was very straightforward and answered very well.

In part (b) many students were able to provide appropriate scattergrams, with accurate title, axes, and plotting. Others missed out on one or two marks with vague titles and / or axes. However some students, despite the questions in this section mentioning ‘relationship’, Spearman’s rho and ‘correlation’, provided completely inappropriate graphical displays e.g. histograms and bar charts. A small minority did not attempt this question at all.

Part (c) required reference to ‘level of measurement’. Although the majority of answers could identify ordinal data for 1 mark, very few went on to characterise ordinal data or why this studyproduced ordinal data, which would have fully justified the use of Spearman’s rho.

Part (d) was very straightforward if the scattergram was plotted accurately, but a significant minority of students were clearly unaware that 0.15 is a negligible correlation and 0.95 a virtually perfect correlation (straight line).

Q3.

Part (a) was generally quite poorly answered. Although many students could gain some credit for providing detail from the study and / or giving ethical issues, few were able to demonstrate sound understanding of the requirements of a practical consent form with many not referring to consent at all. Ethical issues were generally covered reasonably well and were the strongest of the three elements, whilst format / style was the weakest and was often inappropriate (eg ‘You will’, or ‘You must’, or ‘You have taken part’). Although there were some excellent responses, generally it appeared that the students had not really thought about the nature of giving informed consent, with sparse experimental detail and the incorrect tone used, often reading more like a brief / debrief from or a legal disclaimer. This is perhaps an area that schools / colleges could look at in more detail.

In part (b) despite many students showing an understanding of directional hypothesis, just less than half failed to achieve any marks, mainly due to a focus on horror films leading to more nightmares with no reference to the romantic comedy condition. Additionally, many students failed to fully operationalise the variables and some struggled to write an appropriate hypothesis for a repeated measures design, employing a writing frame for an

independent measures design.

In part (c) most students identified individual differences as a reason to use a repeated measures design but only about half of these students managed to clearly link their reason to the experiment. Surprisingly, there was some confusion over the term ‘repeated measures’, with some students focussing their answers on reliability and others on order effects. Responses to this question again suggest that students need to gain more practical experience of designing and conducting experiments as part of their course, rather than focussing on theoretical learning of these concepts.

Part (d) offered a great opportunity to distinguish genuine understanding of reasoning behind the use of counterbalancing as opposed to whether students had merely rote learned what it is and why it is used. Overall, there appeared very little understanding with approximately half of students failing to achieve any marks and approximately one tenth of students accessing both marks. Most students achieving marks identified the issue of order effects but application to the experiment was limited, with many rote learned, inappropriate responses such as boredom / tiredness given rather than engaging with this study.

Part (e) was generally well answered with over half of the students accessing all 3 marks. Generally, answers referring to a hat / container were more successful than those involving a random number generator / computer, as these responses often failed to fully describe the process.

Overall, students were able to answer part (f) well, although understanding of mean scores was generally better than standard deviation. It appeared that students often understood the data but failed to appropriately justify their suggestions, often simply restating information given in the table. There were also a few costly examples of justifications given without suggestions.

Part (g) was a good differentiator, with some excellent responses by students who clearly explained what ‘significant at $p<0.05$’ meant and successfully managed to do so in the context of the experiment. Weaker students generally gave a rote learned reason for why psychologists generally use the 5% level or failed to contextualise their response. Unfortunately, many students failed to explain that the ‘<0.05’ sign meant less than 5%. Teachers should ensure that students are familiar with the mathematical symbols required by the new specification.

Part (h) was generally not well answered. A huge range of modifications to the design of the experiment were given but unfortunately a lot of these were inappropriate, such as ‘use a control group’, ‘use a blind procedure’, ‘use a repeated measures design’ or ‘use an independent measures design’, suggesting either a lack of understanding of the reasoning for experimental design or a failure to recall the experimental protocol. Some students suggested using a matched pairs design and went on to explain effectively how this could improve validity by reducing the likelihood of participants guessing the aim of the study which would reduce demand characteristics. More commonly students justified the modified design by inappropriately suggesting that it would avoid individual differences. Changing the sample was a common answer but it was often not clear how this would improve validity, with few being able to give more than a generic explanation. Stronger responses generally focussed on the problem of using a text message to ascertain whether a nightmare had been experienced, suggesting a questionnaire/interview to reduce the chances of participants lying or to help distinguish nightmares from unpleasant dreams.

Q4.

(a) Despite 36% of students not having the mathematical skills to answer this question

STEMUP.CO.UK

appropriately and 3% not even attempting it, those who did have the knowledge answered it well. Some missed or did not understand the requirement for three significant figures.

(b) Most answers provided an accurate statement of the likely outcome of the experiment and could provide some sort of explanation. However, many students missed the final mark by not providing accurate detail of the type of social influence being displayed, especially in condition 2. Weaker answers were largely common sense, lacking psychological terminology

(c) This was a demanding question as students had to explain how the researchers might have addressed this issue in sufficient detail for 4 marks. Many chose matched pairs as an appropriate method, but were vague about what participants might be matched on, how matching might be carried out, and how participants would then by distributed across conditions. There were, however, some very good answers covering all these aspects. A significant minority chose repeated measures as an appropriate method, but this was unlikely to be practical given the nature of the study. Such answers earned credit only if they showed awareness of the need, for example a very long interval between testing, the use of different but comparable questionnaires, etc.

(d) Despite the injunction a number of students still referred to the level of measurement as a reason for using chi-square. Otherwise this question was done well, with many answers covering the requirement for independent data and a test of difference (between conditions) or association (between variables). Answers referring to correlation did not receive credit.

(e) The majority of students demonstrated understanding the relationship between calculated and critical values, and the link to the chosen significance level. A few were confused as to which was the calculated and which was the critical value.

Q5.

(a) This question was well answered with over 90% of students correctly identifying the type of experiment used in the study.

(b) Almost half of the students gained full credit for this question. Unfortunately, a few gave the independent variable instead of the dependent variable, or got confused with speed or the distance they had to run. Many students failed to state that the running time was measured in seconds.

(c) This question was generally poorly answered, with 57% of students failing to achieve any marks and 3% not attempting it. There was an occasional reference to *‘“the need for continuous data”’*, but very few students gained the second mark. Overall, there was a lack of understanding about what a histogram is or when they should be used and schools/colleges need to address this. A common incorrect answer was to assume a histogram is used for a correlation.

(d) This question was reasonably well answered. Unfortunately, a number provided a full title instead of naming the type of graph, although when these included reference to a bar chart they received appropriate credit. A number of the axis labels were vague, omitting *“seconds”* or *“mean/average”*, or just writing *“conditions”*. There was also some confusion regarding the type of graph with *“scattergram”* being the most common incorrect response.

(e) Although there were some strong responses, generally students found this harder than anticipated. A number failed to receive any credit due to simply defining the

STEMUP.CO.UK

mean and the standard deviation. It was also far too common for students not to understand and answer the 'justify' component of the question. Many students simply restated information from the table or provided possible explanations or conclusions, as opposed to justifications. Although students generally saw the mean as showing a difference, there was often the claim that music hindered performance, with confusion regarding time being a higher score, meaning slower. Worryingly, some students still have little understanding of standard deviation.

(f) Despite 36% of students not having the mathematical skills to answer this question appropriately and 3% not even attempting it, those who did have the knowledge answered it well. Some missed or did not understand the requirement for three significant figures.

(g) This question was generally answered very well, with the majority of students achieving all five marks. Impressively, nearly all students identified the points in mark scheme for choice of test and most could justify why the result was not significant, although a few picked the wrong critical value from the table.

(h) Many students could correctly define a Type II error, although there was some limited descriptions and some confusion between Type I and Type II errors. Few students managed a coherent response to the second part of question, requiring an explanation as to why the 5% level is normally used in research. Many were not clear about the balance between making Type I and Type II errors, suggesting that a 5% level prevented these errors from occurring. Weaker students did little beyond stating that it is used because of convention.

(i) From the stem of the question, a clearly uncontrolled variable was the type of music participants listened to in condition B. Surprisingly few students identified this extraneous variable but there were a variety of environmental factors given that received credit, if they could have feasibly changed within one week and affected the running times. Unfortunately, over half of the students appeared not to have read the stem properly and therefore offered inappropriate extraneous variables, thus gaining zero marks. The most common inappropriate variables were participant variables, which would not have feasibly changed within the week, for example running ability, fitness level, age or gender, or issues regarding order effects, which would have been addressed by the design of the study.

(j) This question was generally quite poorly answered. Although many students could gain some credit by defining what peer review is, very few demonstrated a detailed understanding of theprocesses or the purposes of it, with the majority achieving level 2. Many students made no reference to publication and saw it as a simple checking method, often with the misconception that it involves the reviewer repeating the study. Stronger students gave the process and then proceeded to give its purpose, whereas weaker ones wrote all they knew, including pre-learned evaluative points that did not gain credit. Despite some impressive answers that showed detailed knowledge of both process and purpose, overall students had very limited or no practical understanding of what peer review involves and teachers are encouraged to address this.

Q16.

(a) This question required a definition of content analysis which proved challenging for many students. Almost half of the answers achieved no marks at all. This was made more remarkable by the fact that most were able to gain some marks on part (b) where they were asked to explain how to carry out a content analysis for the data in question.

(b) Most students were able to gain some marks here despite poor performance on part (a) and could identify in a basic way how to carry out a content analysis on the video recordings. Some were able to provide a clear description of the process but few appreciated that behavioural categories need to come from somewhere, whether that is from pilot work or previous research.

(c) Most students were able to identify an appropriate method of testing the reliability of the content analysis and collect at least one mark. The most popular answers were test-retest and inter-rater reliability. Many failed to gain the three marks available as their explanation of how the method of checking reliability would be carried out lacked detail. A few students became side-tracked into improving reliability and a small number used split half which was inappropriate in relation to content analysis and gained no marks.

(d) This question required students to explain why a repeated measures design was used in the experiment. Many students provided a basic answer referring to the need for less participants or the removal of individual differences but were unable to provide further explanation of why this would be important in this experiment. Students who thought about the scenario and elaborated their explanation with reference to reaction times, concentration or driving skills, achieved full marks.

(e) There was a broad range of answers to this question and about 75% of students achieved no marks at all. Many students contradicted their previous answer to part (d) and referred incorrectly to individual differences in reaction times and a similar proportion referred to order effects which had been controlled by counterbalancing or driving experience. Some students picked up on the possibility of differences in the nature of the ‘chat’ on the phone which was encouraging. However, few students showed any awareness of the need to match the two hazard perception tests (stimulus materials / tasks) in this repeated measures design.

(f) Many answers to this question displayed a marked lack of common sense. Despite referring to a simple hazard perception test, which is a key component of the driving test, many students claimed that watching a 3-minute film of a road would be traumatic, leading police drivers to suffer psychological harm. Others referred to possible deception and failed to appreciate that the purpose of the experiment is rather obvious in a repeated measures design. Better answers took issues such as informed consent / right to withdraw and explained how these related to this research.

(g) The question on writing instructions was answered well, with around half of students achieving four or five marks. Some failed to gain full credit as their instructions referenced both conditions or failed to include a check of understanding. Very weak answers failed to refer to the conversation or made no reference to reacting as quickly as possible.

(h) This question required students to identify an appropriate statistical test and justify their choice. About one third of students gained the full marks for identifying the Wilcoxon test with appropriate justification but just under half gained one mark only for identification of the test. Common problems included justification as a test of difference which gained no credit as it was included in the question. Other students were confused about the type of data required for the Wilcoxon test and many answers referred to ‘not nominal’ data.

(i) There is still evidence that few students understand the concepts of statistical error and well over half failed to gain any marks here. Some became confused between type 1 and type 2 errors and others referred to the number of hazards detected rather than reaction times.

STEMUP.CO.UK

(j) This question was answered reasonably well, with many students referring to the greater potential for generalisation in a larger sample of inexperienced drivers. Some also referred to the general importance of replication to check findings in the context of the experiment, which was creditworthy.

Q17.

(a) The majority of students gained the mark for this question by outlining the relationship. A small number of students simply stated 'positive correlation' which did not gain the mark as reference to the variables was required for an outline.

(b) Most students were able to name a correct test (Spearman's or Pearson's) although not all could justify the choice of test with reference to the levels of measurement. Some answers simply stated ordinal or interval, which was not enough as it was not clear which variable was being referred to. One could argue that the empathy scale was not equal interval and therefore should be treated as ordinal data. However, the hours spent reading fiction per week could be treated as interval. Some suggested an incorrect test e.g. Wilcoxon or Chi-square.

(c) Most students gained at least one mark on this question as they could outline a type of validity. However, some just named a type which was not enough for even one mark. Where students gained both marks, they generally referred to either face or concurrent validity (although occasionally to predictive or criterion validity). For full marks, answers had to explain how the type of validity would have been implemented. Many failed to do this part of the question and thus gained only half marks. Some students confused validity and reliability, often referring to split-half or test-retest, and a noticeable number tried to answer the question with reference to pilot studies.

(d) Most students were able to identify a limitation of the study and answered the question well. The majority of the problems identified referred to the sample. A number of answers referred to ethical issues and therefore gained no credit.

(e) Answers which simply stated that correlation looks for a relationship and experiments investigate differences (or similar) gained only one mark. Students that did access further marks were able to explain issues around ethics and manipulation of variables in such a study.

(f) This question was answered well by those students who had clearly had practice at designing and implementing their own investigations. It was pleasing to note that some students gained full marks on this question and it was evident that some schools and colleges had prepared students well. Students who did not gain high marks on this question usually:

- wrote very brief answers or ran out of time
- did not address all 5 bullet points. This was a shame because a number of the points were often addressed comprehensively but then students failed to address certain sections, in particular materials. The best answers addressed each point in turn in a structured and comprehensive manner
- provided little or no justification. The question specifically asked for justification of design decisions, but some answers gave little or no justification, e.g. for design, sampling techniques etc. Without justification the question was not properly addressed and this led to lower marks.

Q18.

(a) Hypothesis writing continues to be a problematic for many students, despite the

STEMUP.CO.UK

requirement to do this at AS level. Less than a third students achieved the full 2 marks available and a further third scored no marks at all, having mistakenly written a directional hypothesis or one which predicted a difference between happiness scores and intelligence scores. Many responses were lacking in clarity or failed to operationalise the variables sufficiently. The best answers were concisely and clearly worded such as “There is a correlation (relationship) between pupils’ scores on a standardised intelligence test and their scores on a questionnaire measuring happiness”, which achieved the full 2 marks.

(b) Most students were able to identify an appropriate alternative method to collect data about happiness, the most popular choices being interviews, observations and diary studies. Some were able to provide a clear explanation of why this would be better, but weaker students became side-tracked into describing the possible method in detail (i.e. observational categories that might be used) and lost focus on the question of comparison. Better students were able to refer to precise advantages of their chosen method over questionnaires which were contextualised in relation to measuring happiness.

(c) This question required students to give two reasons why Spearman’s rho was used to analyse the data. Almost all students were able to accurately identify one reason, which is encouraging!

(d) Weaker students still struggle to interpret critical and obtained values appropriately. About a third of students managed to pick up the full 3 marks. Most of the remainder were able to say the result was significant, gaining one mark but went on to select the incorrect critical value from the table. A small number erroneously compared the correlation coefficient (0.42) with 0.5 demonstrating misunderstanding throughout the entire reasoning.

(e) This question required students to interpret a further correlation coefficient (which was statistically insignificant) and put both pieces of information together to draw an overall conclusion to the reported study. This proved challenging and less than 5 % of answers achieved the full 4 marks here. Many students were able to identify that the scores demonstrated different kinds of relationships (positive and negative) but were unable to take this further and think about possible explanations. The better answers focused on the inability to establish cause in correlational research and the role played by other variables, in this case age. Some made use of their knowledge about reliability which was creditworthy.

Q19.

(a) Hypothesis writing continues to be problematic for many students, despite the requirement to do this at AS level. Around 40% of students achieved zero marks on this question , having mistakenly written a directional hypothesis or one which predicted a difference between mathematical ability and musical ability as opposed to a relationship. Many responses lacked clarity or failed to operationalise the variables sufficiently. The best answers were concisely and clearly worded such as “There is a correlation (relationship) between pupils scores on a test of mathematical ability and their scores on a test of musical ability”, which achieved the full 3 marks.

(b) This question was answered well, with most students scoring two or all three marks. Weaker students were able to spot the test was based on a subjective judgement and some also made the point that singing was a poor measure of all round musical ability. Stronger students identified the lack of control (different choices of song) and were able to link this appropriately to investigator bias. Some students also made the point that the test lacked validity as it had not been standardised.

(c) There was a broad range of answers to this question and about 40% of students achieved no marks at all. Some confused reliability with validity, suggesting various methods such as comparing the scores with another measure of maths ability. Few contextualised this by identifying alternate forms which would have been creditworthy. Others made reference to running a pilot study which received no marks.

The remaining 60% had some idea of ways of assessing reliability of the maths test, the most common methods being test-retest and split-half. Some used inter-rater reliability appropriately suggesting that two separate markers could be used for the maths test: others became sidetracked into assuming that the study was observational. Stronger students were able to explain two or three methods of checking reliability in reasonable detail.

(d) This straightforward question on a random sample caught out quite a few students. Most were able to achieve 1 mark by referring to the method as being likely to yield a more representative sample. The weakest students simply defined random sample and went no further.

(e) This question required students to draw a scatter graph to display the data. About half achieved all three marks here. Many students failed to gain full marks by inaccurate or missing labels or title. About one third of students drew an incorrect graph, the most common error being to draw a bar chart.

(f) Most students were able to make some commentary on the generally negative correlation shown in the graph and table. Better students noted the presence of two outliers which weakened the overall strength of the relationship and some commented on the impact of outliers in a small sample. A small number of students made a rough calculation of Rs which was impressive but unnecessary to gain full marks.

(g) This question had a range of answers from students that covered marks from 0-10. The mark scheme allowed students to argue for different ways of designing the experiment (independent measures or matched pairs) and of generating a sample (volunteer or random selection from the two groups) provided these were workable and justified. Some common errors included:

- suggesting an inappropriate design (repeated measures) which did not take account of the information relating to left and right handers
- suggesting a sampling method but not explaining how it would yield an appropriate sample of left and right handers
- assuming that a maths test also needed to be completed (ie incorrect IV)
- failing to provide any procedural information
- producing a debrief which was not suitable to be read out to participants
- providing standardised instructions and claiming they were a debrief.

Some schools and colleges had clearly prepared their students well and many showed an impressive understanding of experimental design. Others struggled with the question and/or, failed to read the instructions and therefore gained very few marks. Once again, advice to teachers is: to do practical work. It was clear that some students were very familiar with designing experiments and they had a strong advantage here.

STEMUP.CO.UK

(h) This question required students to follow through their design from (g) and give some indication of how the data would be recorded and analysed. Most managed to sketch an appropriate table to record data, although a few misread the question and produced a summary table. Some students were able to follow through their design / data type with an appropriate test which could have been Mann Whitney (independent design) or Wilxocon (matched pairs). Students who had collected nominal data or recorded data in nominal form were credited if they suggested chi square to analyse it.

Q20.

(a) Students often struggled with this question. Very few understood why measures of dispersion are used in addition to measures of central tendency and a number used the term ‘dispersed’. Applications to the stem lacked the necessary detail to attract a mark.

(b) Many students were able to make correct use of the table and draw an appropriate conclusion about the statistical significance of the T value. Stronger students were able to present this information well, with some even correctly stating that the results were not significant at the 0.02 level and explaining why. A few less successful students were confused about the critical and calculated values of T.

(c) It was heartening to see so many students being able to explain why this test was used, with many students scoring the full 3 marks. A few students, however, stated that the data was nominal or interval.

(d) Most students were able to score at least one mark on this question even if they scored poorly on preceding and subsequent questions. Students had to think carefully about this answer and many were able to suggest one or two sensible reasons for the use of a diary by each offender. Elaboration and explanation of each reason proved more of a challenge, sometimes resulting in overlap and repetition across the two reasons offered. Some students simply repeated the stem.

(e) This question produced some answers that showed a lack of understanding of the use of a control group with quite a few students suggesting that a control group would consist of ‘a group of non-offenders’ or ‘normal people / people with no anger issues from a normal population’. More informed students were able to explain why a group of people who would not have the anger management programme would have improved the study although some simply said that ‘it would make it more scientific’ without explaining why.

(f) Answers to this question were most disappointing with almost two fifths of students failing to score a single mark. What should have been a straightforward question proved challenging for many students who seemed unfamiliar with how to establish the reliability and validity of the questionnaire. Some students confused reliability with validity, others simply provided definitions of each and, where an attempt was made to apply reliability and validity to the stem, students very often referred to pilot studies and peer review. Reliability was sometimes mistaken for replicability, the ‘split-half method’ was frequently explained as dividing the total score in half to see if the two halves correlated and test- retest was sometimes explained as testing one group of offenders and retesting another group of offenders. Face validity was rarely applied to the questionnaire ie anger scores. Even students who correctly addressed the issues of reliability and validity, failed to explain how statistical tests of correlation would be used in this context. The value of carrying out practical activities to enable students to ‘think like a psychologist’ and apply their knowledge of practical activities in answer to questions such as this one cannot be overstated. It is clear that where students had been presented with such opportunities, they

STEMUP.CO.UK

were able to write confidently and in an informed manner.

Q21.

(a) This question was answered well with most students aware that a directional hypothesis was appropriate due to the existence of previous research. A minority of students provided rather more detail than required for one mark.

(b) Hypothesis writing is still a problematic area for many students, despite the requirement to do this at AS level. Many students achieved zero marks on this question, having mistakenly written a non-directional hypothesis or one which predicted a difference between older and younger patients. Many responses were lacking in clarity or failed to operationalise recall adequately. The best answers were concisely and clearly worded such as “There is a negative correlation (relationship) between age and recall accuracy rating”, which achieved the full three marks.

(c) Although this question was worth only one mark, many students produced lengthy answers. Some distinguished between specific types of reliability such as external or internal. A small number of students became confused between validity and reliability.

(d) There was a broad range of answers to this question, with students in roughly equal measure being awarded marks across the full range. The majority had at least a rough idea of ways of assessing reliability (the most common being inter-rater) but found it difficult to select an appropriate method for the study detailed. The weakest answers were those where the student focussed on reliability of the study overall, rather than reliability of the ratings which was what the question required. Answers that achieved the full three marks generally selected the most straightforward idea; to take two independent psychologists who rated the typed accounts separately and then correlated their ratings. Students who achieved only one mark suggested test retest as a method but most were unable to carry this through and indicate that the psychologist would need to return to the data after a suitable interval and re-rate the accounts.

(e) This question was answered well with the majority of students achieving two marks. There was a range of both kinds of data to draw on here including the doctor’s notes and the patients responses (qualitative data) and ages and accuracy scores (quantitative data).

(f) Answers to this question demonstrated an understanding of the use of the Spearman’s rho statistical test with the majority of students achieving two marks.

(g) This question confused many students who were unaware that the critical value relates to the magnitude of rho not the direction. So negative correlation drops the minus sign when compared with the critical values. About half of students were clearly aware of this and could compare the obtained value with the correct figure from the table. The remainder made a number of errors, some comparing -.52 with 0.05, others claiming that the figure was smaller than .306. Some incorrectly used the values relating to a non-directional hypothesis.

(h) Full marks were achieved by stating that the null is rejected and the experimental hypothesis accepted, when in fact results are due to chance. Good understanding was shown among students who referred to the level of significance being set too leniently or the 5% likelihood of a Type 1 error occurring with the 0.05 level of significance. In about one in three cases students confused Type 1 and Type 2 errors.

STEMUP.CO.UK

(i) This question was challenging for students with many achieving no credit or not answering the question. Even students who were able to explain what was meant by a Type 1 error were unable to apply this knowledge in this question and compare the obtained value with the 1% significance level. A small number gained one mark for identifying that the obtained value was substantially larger than .306. However, far too many relied on a rote learned response that the 5% significance level avoids Type 1 errors therefore one could not have occurred.

Q22.

(a) In this question, students were required to discuss the advantages of carrying out the experiment described in the stem, in a laboratory. Fewer than half of students made any reference to the stem and the most common mark awarded was one out of four. Those who referred to an advantage (eg control of extraneous variables) and linked it appropriately to the scenario (eg posters on the walls) were able to access the full range of marks. A small but significant minority insisted on writing about disadvantages and achieved no marks. Once again, schools and colleges should advise students to read stems carefully and apply knowledge in Section C.

(b) Most students achieved full marks, identifying the Mann-Whitney as the appropriate test and giving and ordinal data or independent groups as a reason. Some students provided two or three reasons going beyond the requirements of the question. There were a minority of cases where an incorrect answer was given, most commonly Spearman's rho or Wilcoxon's signed ranks test.

Q23.

(a) Hypothesis writing is still a problematic area for many candidates – despite the requirement to do this at AS level. Many candidates achieved zero marks on part (a), having mistakenly written a directional or a null hypothesis. Many responses were lacking in clarity or failed to include an operationalised DV so only achieved 1 mark. The best answers were concisely and clearly worded responses such as "There will be an association between birth order and career choice", which achieved the full 2 marks.

(b) Virtually all candidates identified an appropriate sampling technique. However, a large number did not score full marks because their account of how to obtain the sample was confused or insufficiently linked to the study in question on artists and lawyers. Candidates who chose a random sample needed to explain how the target population would be identified.

(c) Some centres had clearly prepared their candidates very well and many showed an impressive understanding of inferential statistics scoring 11 or 12 marks. However, other candidates struggled with the question and collected very few marks. Some of the most common errors were as follows.

A number of candidates did not know how to express the statistical conclusion of a research study, by referring to observed and critical values and probability. There were errors in correctly identifying the observed and critical values and their relationship to the hypothesis. A large number of candidates did not label the axes of the graph or only showed data relating to first born career choices.

Some candidates chose the wrong statistical test; some did choose the correct statistical test but did not then state the reasons why the test was appropriate.

Yet again, advice to teachers is to do some practical work. It was clear that some candidates were very familiar with the rationale for selecting a test and deciding if an

STEMUP.CO.UK

observed value is significant or not. These candidates had a strong advantage on part (c).

Q24.

(a) A challenging question because candidates needed to apply their knowledge. They often knew about what makes something scientific (objectivity, replicability, etc) but seemed unable to engage with the stem. There were lots of answers involving paradigm shift, which were not relevant to this question.

(b) Most candidates had some idea about how a random sample could be obtained, but often failed to explain the methods fully. They could suggest all the names should be put in a hat, but did not make it clear that the names were then selected “without looking” or “without bias”. There was some confusion with systematic sampling.

(c) Many answers displayed some confusion here, eg saying that a limitation was that it was not representative of the whole population, when the point is that is might not be representative of the target population of 400. Some answers referred to problems of allocation to conditions, rather than random sampling. A good point was made by those who said that if some parents did not give consent, the psychologist would have to select again, and that would not be random.

(d) This question was not answered well. Most candidates seemed very unclear about why it is important to operationalise variables. How to actually operationalise the two variables was beyond many candidates. Some effective answers referred to food content eg fat, sugar etc.

(e) Most candidates answered this correctly.

(f) There was some serious confusion about what exactly matched pairs design is. Few could go beyond “it’s time consuming” or “difficult to match on all variables”. Some referred back to the random sample and said it would not be possible; others felt that at five-years-old children are either too similar to match or too different.

(g) Most could identify an ethical issue such as confidentiality, the right to withdraw and protection from harm (those who did not get any breakfast or who were embarrassed at their poor reading). Some seemed to forget that they also had to explain how the issue would be dealt with, or they simply repeated that the right to with draw could be dealt with by giving the right to withdraw.

(h) This question was not answered well. Many candidates failed to read the question carefully before they attempted it. They were given the information that they were using the same group of children (ie the 5-year olds in the previous study). Despite the fact that the ethical issues and sampling had already been addressed in the plan for the original study many wrote at great length about sampling and ethics. The majority of candidates were unable to write a fully operationalised hypothesis, and often simply restated the aim. Many seemed to think the IV was breakfast versus no breakfast, rather than healthy versus unhealthy breakfast. Some of their ideas were totally impractical, especially given that the children were only 5 years old. In many answers lack of detail would have made any kind of replication very difficult. However, some candidates did understand the need for some sort of training for the observers, the need for clearly identified behaviour categories to record, and the importance of being able to distinguish the two groups in the playground. Designing a study is clearly a difficult task for candidates, and one that they need to practice.

Q25.

(a) Almost all answers were correct – however, surprisingly, some answers were left blank or the answer ‘yes’ was provided.

(b) Many candidates clearly understood how to read the table and to interpret results and so gained the full 3 marks here. Some gained 1 mark for saying that the result was significant but then demonstrated a complete lack of understanding in the rest of their answer.

Q26.

(a) This question proved to be a good discriminator. Candidates who understood scattergrams were able to make a reasonable sketch with appropriate labels and accurately plotted data and so gained full marks. However, a disappointingly large number of candidates clearly had no understanding of scattergrams and drew a frequency polygon instead for which they could gain no marks. The requirement to present and understand graphs is clearly stated on the AS specification: ‘presentation and interpretation of quantitative data including graphs, scattergrams and tables.’

(b) Some candidates gave full answers in which they made good use of the data contained in the table. However, fewer candidates were able to make use of the information in the scattergram and very few referred to correlation. There were 4 marks available for this question which should have made candidates realise that some detail was required. Answers such as ‘she was concerned because the observers gave different ratings’ could not gain much credit. Quite a few candidates wasted time by defining inter-rater reliability. Answered included suggestions of how to improve reliability which, of course, was addressed in part (d).

(c) Relatively few candidates identified an appropriate test - almost every reasonably familiar test was quoted. Experimental designs were often quoted as incorrect reasons for test selection. Many candidates did not even suggest an inferential test but suggested calculating the range, mean or standard deviation. Candidates who did identify the appropriate test were usually also able to offer an appropriate justification.

(d) This was a good discriminator. Most candidates could offer at least one solution to this issue but many stopped after making their initial point eg ‘give them more training’. Some were able to elaborate on this effectively to gain full marks but many showed little understanding. Very common errors were ‘get more observers’ or ‘average the results’ or ‘only use one observer’.

Q27.

Perhaps surprisingly, many candidates achieved higher marks for this section than for their other two questions. Most candidates attempted all parts of the question although a significant minority did not complete part (h) suggesting that they might have run out of time. This was a pity since this part carried 10 marks.

(a) This was a straightforward question and many candidates accessed full marks but a surprising number were confused by median and range and some did not understand what the range indicated about the data.

(b) Many candidates were very well prepared and got full marks here but some wrote very confused answers showing little understanding eg ‘Spearman’s Rho because it was nominal data and repeated measures’.

(c) There was a centre effect here. Some candidates had a good understanding of Type

STEMUP.CO.UK

1 error while others had clearly never heard of it. Some understood what is meant by the term and offered a definition but were not able to apply their knowledge to answer the question.

(d) Candidates offered a wide range of answers, although some were a little bit too brief or poorly explained to get both marks. There was some confusion about what is meant by a placebo and some candidates offered two explanations which were essentially the same as one another. Other candidates offered factors which could apply equally to the treatment and non-treatment group. It is important in this kind of question to read the stem carefully. Some candidates said that the therapists might have been biased in favour of the treatment group, but the stem clearly states that the therapist did not know who had been in which group.

(e) A lot of candidates missed the point that the advantage / disadvantage needed to be in comparison to interviews. Many candidates gave advantages/disadvantages that could apply equally well to both self-reports and interviews. This was acceptable only if the candidate made it clear ie '*People are less honest in a questionnaire*' was not credit-worthy because it could apply to both interviews and questionnaires. However, '*People are less honest in a questionnaire because they are anonymous and feel they can lie about themselves without being found out*. *In an interview where they are face-to-face with the interviewer, they might find it more difficult to lie*'.

(f) Some candidates wrote excellent consent forms containing both ethical and procedural information and expressed them in appropriate language. Some candidates had a very vague understanding of what needed to be included here, either only focusing on all the ethical issues (*you will have the right to withdraw your data, yourself etc*) with no mention of the procedures, or vice versa. Many adopted a rather inappropriate tone eg '*Once you have signed this form, you are committed to being in the study' or 'You have to subject yourself to an interview*'. It was surprising to see that a few candidates seemed to think a consent form acted as some kind of legal disclaimer – '*you may suffer harm but if you sign this you can't sue us.*'

It was notable on this question that candidates who were able to express themselves clearly and succinctly were much more likely to access full marks. Many answers were so poorly constructed that the content was difficult to understand. Many switched confusingly between pronouns eg *They will have to have an interview. You can withdraw at any time. I agree to be part of this study*.

(g) There were some very muddled answers to this question. Candidates often didn't read the question carefully, and wrote something like '*Reliability means if you do the study again you will get similar results*' for their definition and then didn't know what to write for the next part of the question. Those candidates who explained it in terms of inter-rater reliability generally gained full marks. Some candidates did not read the question carefully and did not relate their answer to checking the scores in this particular study. Many candidates thought incorrectly that test-retest involved using different participants. Some candidates suggested split-half methods indicating a lack of thought about the question. Some candidates confused reliability with validity.

(h) Many candidates showed limited awareness of a conventional reporting style. While it was not necessary to divide the method section into sub-sections, this strategy might have helped candidates to include all the relevant details. Weaker answers made no mention of gender or eating disorders and simply repeated details from the stimulus material. Many candidates completely lost sight of the fact that gender differences were being investigated and suggested randomly allocating participants to groups. A lot of time was wasted in including aims / hypotheses and statistical

analyses which do not form part of a method section. Better answers included appropriate detail of IV, DV, design, sampling method, materials / equipment and procedure which would have enabled replication to take place. As in (f), poor expression and grammatical errors often obscured meaning.