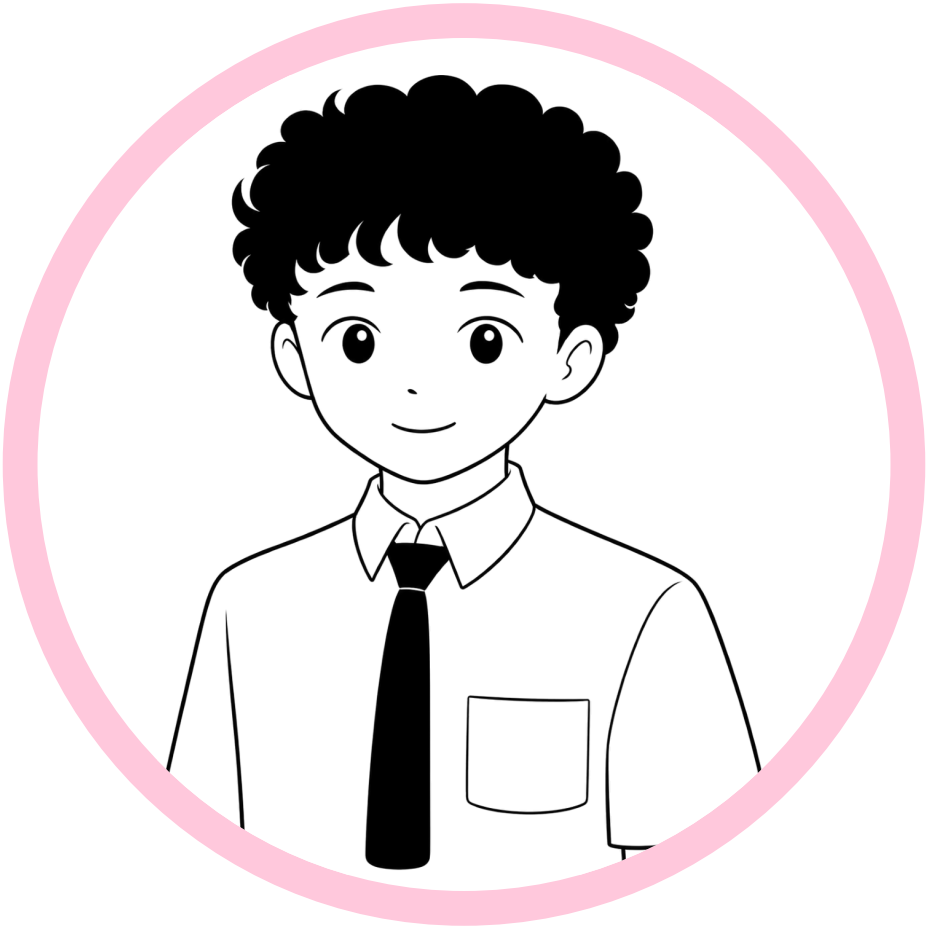
AQA A-LEVEL PSYCHOLOGY · 7182

PSYCHOLOGY



MARK SCHEME

3.2.3.2 DATA HANDLING AND ANALYSIS

SCORE	PERCENTAGE
/813	%

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) **[AO2 = 2]**

1 mark for identifying an appropriate type of observation from the following: naturalistic or covert or nonparticipant.

1 mark for relevant justification in this study.

Possible applications:

- it was carried out in the school canteen which was the normal environment for the students
- the teacher did not place the recycle bins in the canteen they had already been placed there
- the teacher was hidden from view of the students
- the students didn't know they were being observed
- the teacher observed from a distance and did not interact with the students.

2

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

2 marks for relevant explanation of event sampling applied to the scenario, e.g. the teacher recorded each time one of the three types of item was recycled.

1 mark for naming event sampling or limited / muddled explanation

2

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

Possible content:

Age

- it suggests that younger children are likely to recycle more than other children **(1 mark)**. More year 7-9 (than any other age group) recycled for all of the categories (cardboard, plastic and food) **(1 mark)**

or

- it suggests that older children are least likely to recycle **(1 mark)**. Fewer year 12-13 students recycled any of the items (cardboard, plastic or food) **(1 mark)**

Plus

Type of item:

- it suggests that children are most likely to recycle plastic **(1 mark)**. More children in each age group recycled plastic (than any of the other items) **(1 mark)**

or

- it suggests that children are least likely to recycle food **(1 mark)**. Fewer children recycled food (compared to all other items) for all age groups **(1 mark)**

Accept other possible suggestions / conclusions that apply to the data e.g. children do seem to recycle or that are derived from totalling the data. Credit the implications of these findings.

4

STEMUP.CO.UK

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

2 marks for a clear, coherent explanation of an appropriate improvement with some details e.g. more locations, more lunch times, more observers, further ethical considerations

1 mark for brief / muddled / limited explanation.

Accept other appropriate explanations that are linked to the design.

2

[10]

Q2.

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

2 marks for 250 hours OR for 1000/10 x 2.5 = 250 hours OR for 100 x 2.5 = 250 hours

1 mark for correct workings but incorrect answer e.g. 15000 minutes.

2

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

1 mark for correct identification

plus

1 mark for correct explanation

Possible content:

- primary data because the results came directly from the eye witnesses / because the data was collected by the psychologist specifically for the purpose of the investigation
- quantitative data because the psychologist used closed questions
- qualitative because the psychologist recorded what they said (in words).

Accept answers referring to levels of measurement with appropriate justification.

2

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

2 marks for relevant modification that would reduce investigator effects in this study.

1 mark for a brief or muddled explanation.

Possible content:

- have an interviewer who had not witnessed the event / did not know the aims of the study so that they would not be affected by their own perception of the event
- use open-ended questions so that the interviewees were able to give a more detailed and accurate version of what they saw
- use questionnaire (or other means) to collect data without face to face interaction.

Credit other relevant suggestions that would reduce investigator effects in this study.

2

STEMUP.CO.UK

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

2 marks for a clear, coherent outline with some detail.

1 mark for a limited / muddled outline, e.g. 'other psychologists assess / check / review her research report'.

0 marks for 'other psychologists look at the research'.

Possible content:

- her report would be sent for independent scrutiny (checked / reviewed) by other psychologists
- they would consider e.g. validity, ethics, errors, significance, originality and possible improvements
- to see whether it should be published.

Credit other relevant answers that can be applied to the stem.

2

[8]

Q3.

(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

STEMUP.CO.UK

		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
- 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]

Q4.

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

2 marks for a clear explanation of why it was important to match the students for this variable.

1 mark for a limited or muddled explanation, or for an answer without application.

Content:

- level of interest could affect their satisfaction scores so needs to be controlled; matching students' levels of interest in business studies keeps this participant variable constant.

2

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

1 mark for a suggestion of an appropriate variable – interest in using computers (e.g., for studying), past experience of using computers (e.g., for studying), disability issues, prior knowledge of this business studies topic.

Accept other appropriate suggestions.

1 mark for an explanation of how it might relate to satisfaction scores.

2

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

STEMUP.CO.UK

2 marks for a clear outline of one benefit of random allocation in this situation.

1 mark for a limited or muddled explanation, or for an answer without application.

Relevant points:

- removal of researcher bias in determining which participant has the most to least number of online hours
- increased validity as the researcher will not be biased in how students are allocated
- removal of participant bias as each participant has the same chance of having a particular number of online hours as any other participant.

2

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

3 marks for a clear and coherent explanation of how stratified sampling would improve the design.

2 marks for some explanation of how stratified sampling would improve the design.

1 mark for a muddled / limited explanation.

Content:

- stratified sampling ensures each identified sub group of the whole target population of possible participants is represented in the sample so generalisation of the results to the target population is more valid than if the participants were just available at the time of asking
- stratified sampling is an improvement because people who are identified by opportunity are likely to be present at same time, doing similar things, possibly of similar personality types etc which may limit representativeness of sample.

Accept other appropriate explanation points.

3

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

1 mark for a question which would generate qualitative data: what did you like / dislike about the online / classroom lessons?

Accept any appropriate question or item that generates qualitative data.

1

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

2 marks for a clear and coherent explanation of a possible problem.

1 mark for a muddled / limited explanation.

Possible problems:

- analysis of qualitative data relies on interpretation by the researcher which could be inaccurate / biased / subjective
- qualitative data may not be easy to categorise / collate into a sensible number of answer types
- the researcher could be left with lots of individual responses that cannot be summarised.

STEMUP.CO.UK

Do not accept answers which simply identify a problem with no further appropriate elaboration e.g. ‘hard to analyse’, ‘time consuming’.

2

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

1 mark for an outline of an appropriate ethical issue that would be relevant when using a questionnaire: issues such as anonymity, protection from harm, treating with respect, parental consent.

Possible issues:

- the students should be treated with respect by explaining that their answers will be kept confidential
- the students must be assured that they can withdraw their answers if they want to
- the students must be told exactly what the researcher was trying to find out about the relationship between satisfaction and the way information was delivered.

Answer must be an outline and not just a name or term.

1

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

Award **1 mark** for each of the following points.

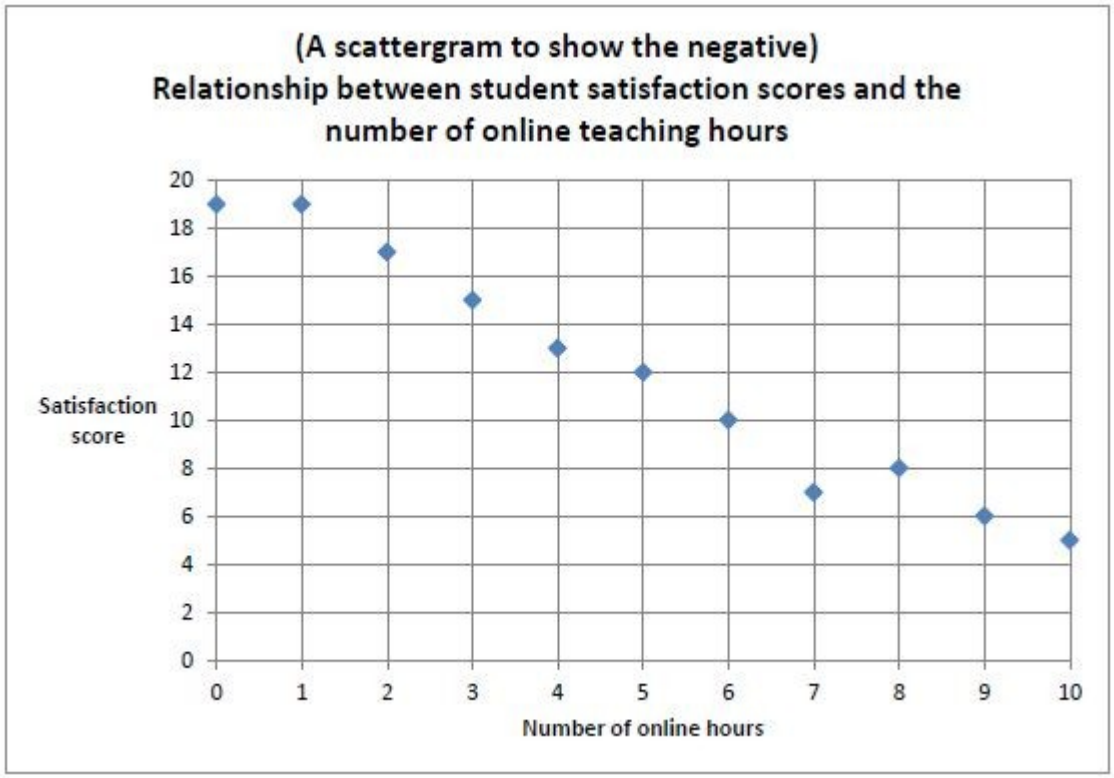
Content:

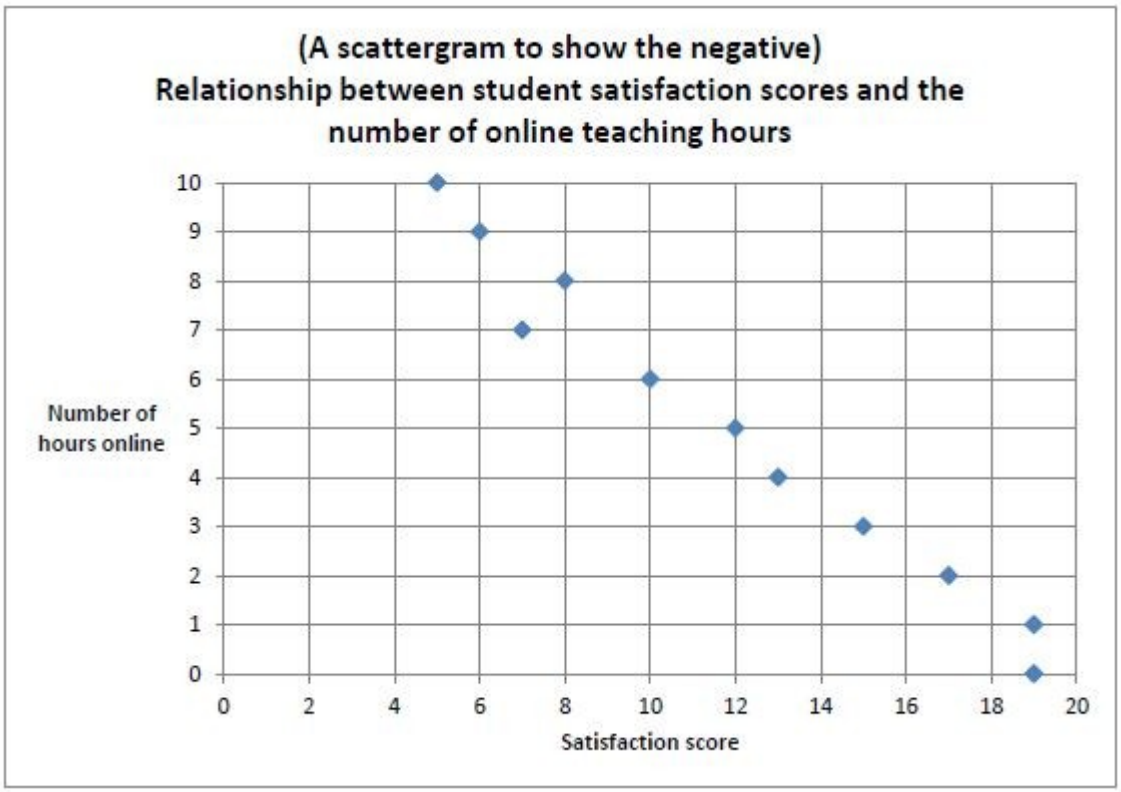
- appropriate title, including reference to (negative) relationship and the two variables
- axis 1 – number of online hours and suitable scale 0-10
- axis 2 – satisfaction score and suitable scale 0-20
- accurate plotting of 11 points on scattergram.

No penalty for drawing a line of best fit.

If display is not a scattergram award max 1 mark if title is appropriate.

STEMUP.CO.UK





4

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

1 mark for saying there is a negative relationship / correlation (between the number of online hours and satisfaction scores).

1 mark for further detail e.g. the **more** hours they have online the **less** they are satisfied with the course (or vice versa), however this does not indicate a cause and

STEMUP.CO.UK

effect relationship.

Accept other valid conclusions.

2

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

1 mark for the mode = 19

1 mark for the median = 12

2

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

2 marks for a clear and coherent explanation.

1 mark for a muddled or limited explanation.

- The median is more representative than the mode because the mode is at the top end of the set of scores, whereas more of the scores are numerically closer to the median.

Accept other explanations, for example, in terms suitability in relation to ordinal data vs nominal data.

2

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

1 mark for B

1

[24]

Q5.

(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

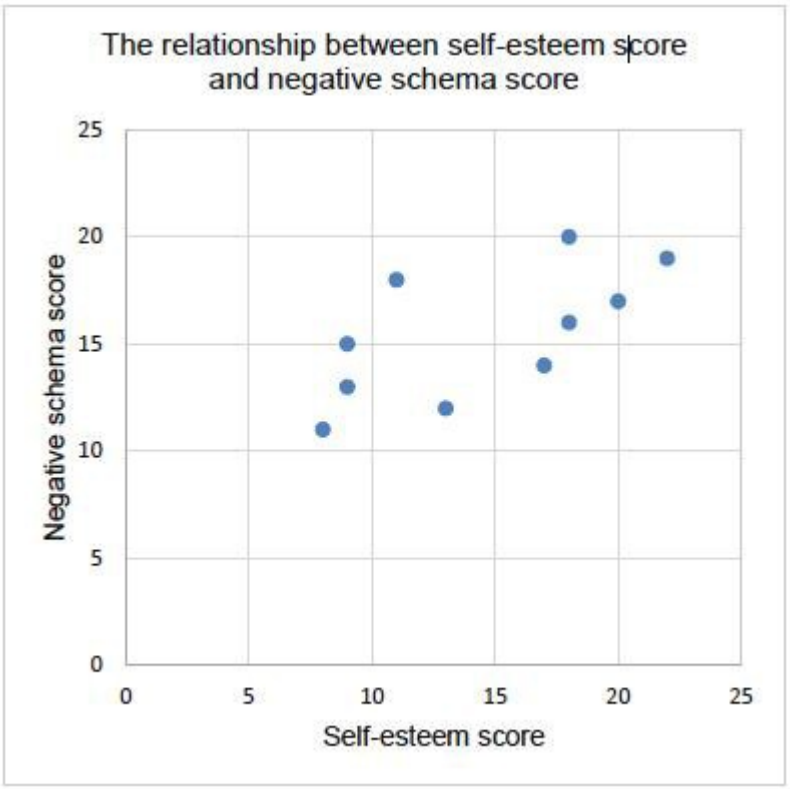
Note: co-variables should be operationalised ('score') either in the title or on the axes.

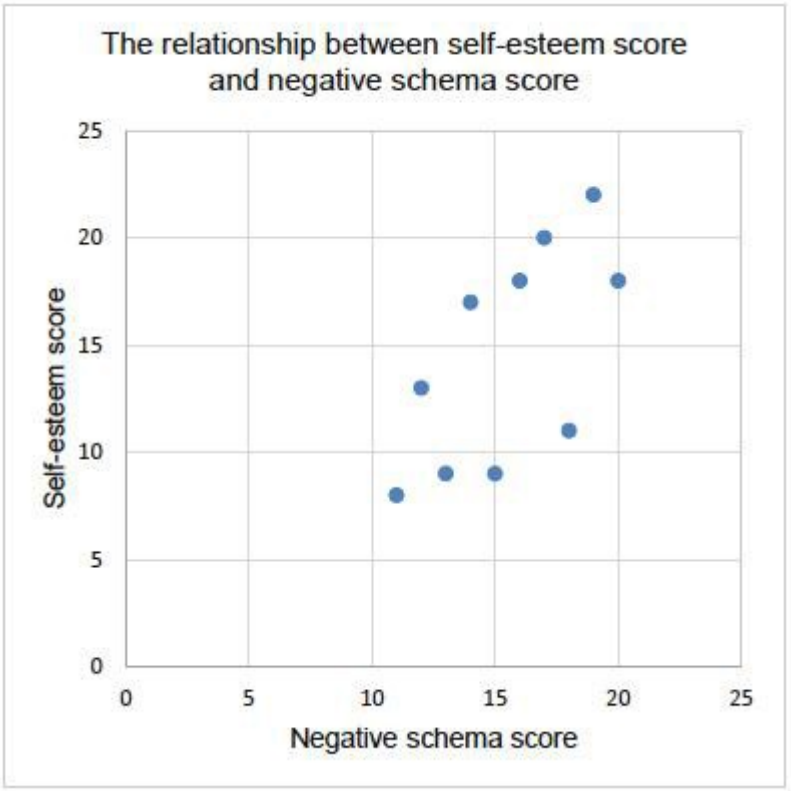
Accept: line of best fit.

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]

Q6.

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

C 27%

1

(b) **[AO1 = 1 AO2 = 2]**

1 mark for a clear and coherent definition of a pilot study.

Pilot studies are small-scale investigations conducted before research.

Plus

2 marks for a clear, elaborated reason for why the psychologist decided to conduct a pilot study for this investigation.

1 mark for a limited / muddled reason for why the psychologist decided to conduct a pilot study for this investigation.

Possible reasons:

- to identify whether there needed to be any modifications in the design, e.g. whether the interview questions were sufficiently relevant to dream content
- to identify whether it would be feasible to conduct a study on the differences in dream themes across gender.

Credit other possible reasons.

3

(c) **[AO1 = 1 AO2 = 1 maths]**

1 mark for a clear and accurate description of qualitative data.

Qualitative data is non-numerical / descriptive data.

Plus

STEMUP.CO.UK

1 mark for a clear strength of collecting qualitative data in this study.

Possible strengths:

- enables the students to talk in detail about their dreams to develop deeper understanding of dream content
- enables the researcher to develop a deep understanding of the content of the dream to improve classification and identification of dream themes.

Credit any other relevant strengths.

Note – zero marks should be awarded for a generic strength (not applied to the stem).

2

(d) **[AO1 = 1 AO3 = 2]**

1 mark for a clear description of investigator effects.

Any (unintentional) influence of the researcher's behaviour / characteristics on participants / data / outcome.

Plus

2 marks for a clear and elaborated way that the psychologist could minimise investigator effects in this study.

1 mark for a limited / muddled way that the psychologist could minimise investigator effects in this study.

Possible ways:

- provide a standardised script for the interviewers to use so that they all asked the same questions in the same way to avoid any bias in the students' responses
- the interviewers could have been trained to greet the students in the same way and ask questions with a neutral tone
- ensure all interviewers were female or all interviewers were male
- ensure that the students were interviewed by someone of the same gender as themselves.

Credit any other relevant ways.

Note – Do **not** credit any answers referring to using a blind procedure (e.g. 'ensuring the interviewer was not aware of the purpose of the study') as this is already stated in the stem.

3

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

Award **1 mark** for each of the following points:

- identify / devise important categories / themes (from the interviews)
- example of possible category / theme (e.g. references to aggressive interactions)
- work through transcribed conversations / repeatedly listen to segments of the recording
- count / tally the number of occurrences of each of the categories / themes.

STEMUP.CO.UK

4

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

Test-retest reliability - Award **1 mark** for each of the following points:

- content analysis repeated on a second occasion using the same interview data
- compare the results of the two separate analysis (number of occurrences of each)
- researchers could calculate the correlation between the two ratings
- researchers generally accept 0.8 correlation (accept 0.7-0.9) between the test and the re-test.

Inter-rater reliability - Award **1 mark** for each of the following points (up to 4 marks):

- use a second person to work with the original researcher
- they could read the interviews (separately) and devise a set of categories (and agree operational definitions)
- they could tally the occurrences of each of the categories of the interviews (separately)
- they could compare their tally charts looking for agreement
- researchers could calculate the correlation between the two ratings
- researchers generally accept 0.8 correlation (accept 0.7-0.9) between the test and the re-test.

4

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

2 marks for correct answer: 225 (even if no correct workings are shown).

1 mark if incorrect answer is provided but all workings are correct.

Correct workings:

60/100 = 0.60

0.60*375 = 225 dreams

2

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

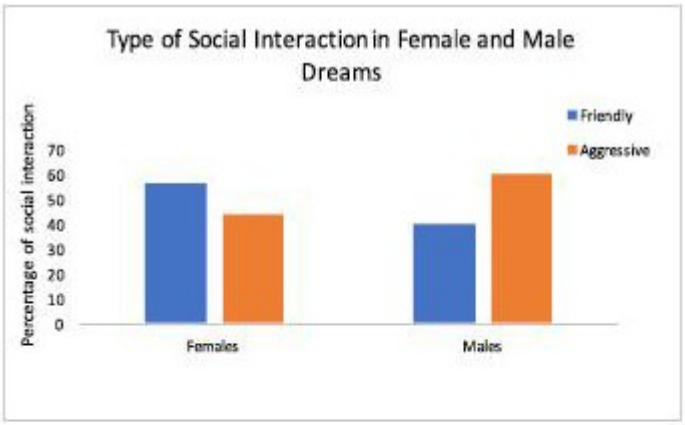
1 mark for each of the following:

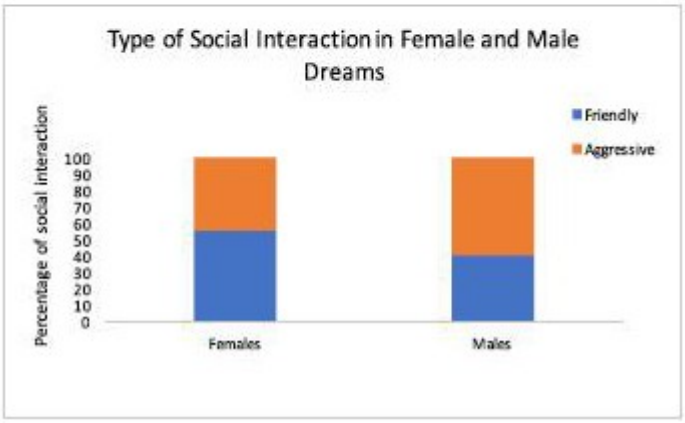
- appropriate title with reference to the IV and the DV
- accurately plotted data, bars must not touch
- accurate y-axis label, percentage / % is sufficient
- accurate x-axis label with key, where necessary.
-

Note – If the graphical display is not suitable candidates can be awarded a maximum of **1 mark** for an appropriate title.

Note – Do not credit the title if it explicitly refers to a histogram.

STEMUP.CO.UK





4

[23]

Q7.

(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.

STEMUP.CO.UK

	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 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.

STEMUP.CO.UK

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:

- 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.

STEMUP.CO.UK

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.

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

STEMUP.CO.UK

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

3

[25]

Q8.

[AO2 = 4]

1 mark for **each** of the following points:

- group A scores indicate a normal distribution
- because in Group A the mean, median and mode are almost the same (as in a bell-shaped curve)
- group B scores indicate a (positively) skewed distribution
- because the mean is higher than the median / mode or to the right hand side of the distribution.

[4]

Q9.

[AO2 = 4]

1 mark for **each** of the following points:

- group A scores indicate a normal distribution
- because in Group A the mean, median and mode are almost the same (as in a bell-shaped curve)
- group B scores indicate a (positively) skewed distribution
- because the mean is higher than the median / mode or to the right hand side of the distribution.

[4]

Q10.

[AO2 = 4]

1 mark for **each** of the following points:

- group A scores indicate a normal distribution
- because in Group A the mean, median and mode are almost the same (as in a bell-shaped curve)
- group B scores indicate a (positively) skewed distribution
- because the mean is higher than the median / mode or to the right hand side of the distribution.

STEMUP.CO.UK

[4]

Q11.

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

Boys = between 26% and 37% inclusive.

Girls = between 63% and 74% inclusive.

Maximum **1 mark** if the two figures do not add up to 100.

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

2 marks for the correct answer: 90

If the answer is incorrect, can award **1 mark** for the correct workings:

150 × 0.6

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

Level	Marks	Description
2	3 – 4	Knowledge of two differences between qualitative and quantitative data is clear and well detailed. The answer is generally coherent with effective use of terminology.
1	1 – 3	Knowledge of two differences is evident. Specialist terminology is not always used appropriately. Or **one** difference at level 2.
	0	No relevant content.

Possible differences (expressed as comparison):

- Quantitative data involves numbers (number of boys and girls), whereas qualitative data involves words, e.g. quotes from the interview
- Quantitative data can be measured objectively, whereas qualitative data is based on the subjective interpretation of language
- Quantitative data is immediately quantifiable, whereas qualitative data has to be transformed and is only quantifiable if the data is put into categories and the frequency is counted.

Credit any other relevant differences e.g. ease of analysis, easier to identify trends/patterns, ease of comparison.

Q12.

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

2 marks for explanation that a non-directional hypothesis is suitable or ‘it should not be directional,’ (1) as there is no reference to evidence that allows the psychologist to predict the direction of the results (1).

STEMUP.CO.UK

1 mark for a muddled/limited explanation of why the hypothesis should be non-directional or 1 mark for stating non-directional/'no'.

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

2 marks for correct answer 55(%)

1 mark for correct workings only 22/40 = 0.55 × 100

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

2 marks for a clear explanation of why stratified sampling might improve the study.

1 mark for a muddled/limited explanation

Relevant points:
- stratified sampling increases representation/generalisation by sampling many subsets of students
- stratified sampling reduces research bias as the subsets of students are selected randomly.

Credit other relevant answers, e.g. stratified sampling increases external validity/population validity by sampling many subsets of students, or justification through comparison with volunteer sampling.

Note: No marks for simply stating increases validity/reliability

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

2 marks for a clear explanation of primary data.

1 mark for a muddled or limited explanation (e.g. reference to 'first hand' or focus on researcher not data).

Content
- primary data is original data/first hand response of the participants/source in the research.
- collected specifically for the research being carried out.

Credit answers that refer to the data in the study described.

Note: do not credit an answer which simply states collected by the researcher

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

2 marks for identification and a clear explanation of one reason why the mean is the most appropriate measure of central tendency.

1 mark for identification of one reason.

Possible reasons:
- the mean can be said to be **representative of all the data** collected as it is calculated using all the individual values.
- the mean is the **most sensitive measure of central tendency** as it uses all the values in set of data.

Accept other valid reasons.

STEMUP.CO.UK

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

Task A

2 marks for correct answer 63 (secs)

1 mark for correct workings only 630/10

Task B

2 marks for correct answer 77.4 (secs)

1 mark for correct workings only 774/10

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

3 marks for a clear and coherent discussion of the purpose of counterbalancing.

2 marks for some relevant discussion.

1 mark for a muddled/limited discussion.

Possible content:
- it controls the impact of order effects (practice, fatigue or boredom)
- allows order effects to be distributed evenly across both conditions
- making each condition of the IV occur as the first task and the second task equally
- counterbalancing does not eliminate order effects which will be present because there are two separate tasks to be completed by each person.

Note: No marks for answers which only state decreasing/reducing/eliminating/preventing order effects without reference to any of the above points.

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

1 mark for identification of **one** appropriate extraneous variable.

Possible variables: that can be controlled and need to stay constant for all participants or from Task A to Task B to avoid affecting the dependent variable:
- level of difficulty in the two 'spot the difference tasks'
- wearing of headphones
- the music/tracks
- instructions for participants
- characteristics of the psychologist
- change in environmental conditions.

Accept other relevant variables
Do **not** accept participant variables

Possible explanations:

1 mark for elaboration of the effect of the EV.

1 mark for how it might have affected the DV (the effect on time).

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

STEMUP.CO.UK

Level	Marks	Description
2	3 – 4	There is a clear and detailed explanation of how the second study improves on the first study. The answer is generally coherent with effective use of appropriate terminology.
1	1 – 2	There is limited or partial explanation of how the second study improves on the first study. The answer lacks coherence and use of appropriate terminology.
	0	No relevant content.

Possible content:

- the second study is more objective. In the first study participants self-reported what they believed to be the case and this is refuted by the follow-up study
- the measurement of time allows for more detailed analysis of data
- use of experimental method involves manipulation of IV, establishing cause and effect in follow-up study which is not a feature of the self-report method used in the first study
- possibility of socially desirable answers as a feature of first study which might be unreliable and this is not present in follow-up study where measurements are more direct.

Credit other possible explanations.

Note: no marks for simply stating more valid/reliable.

Q13.

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

1 mark for an accurate, relevant disadvantage of quantitative data eg lacks detail (which may decrease validity of findings) etc.

Plus

1 mark for clear application regarding the negative impact of using quantitative data in this study (such as rating a patient's improvement).

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

1 mark for an appropriate question which produces quantitative data, eg 'On average how many hours of physiotherapy does the patient receive per week?'

Plus

1 mark for an appropriate question which produces qualitative data, eg 'How does physiotherapy help your patients?'

Note: can accept relevant items that are not phrased as questions, eg describe… or rate....

Q14.

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

STEMUP.CO.UK

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.

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.

STEMUP.CO.UK

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 ÷ 123 =0.048780
0.048780 × 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:

- 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.

STEMUP.CO.UK

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.

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

STEMUP.CO.UK

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.

Q15.

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

Level	Marks	Description
2	3 – 4	Knowledge of the difference between primary and secondary data is clear and mostly accurate. The knowledge is applied appropriately. The answer is generally coherent with effective use of terminology.
1	1 – 2	Knowledge of the difference is evident but the answer lacks accuracy and detail. Application is limited, inappropriate or absent. Use of terminology is either absent or inappropriate.
	0	No relevant content.

Application:
- primary data is gathered directly/first-hand from the participants, and is specific to the aim of the study whereas secondary data has previously been collected by a third party (another researcher or an official body), not specifically for the aim of the study, and then used by the researcher
- in this study, the interview recordings are primary data –gathered specifically for the purpose of investigating causes of aggressive behaviour whereas the school records are secondary data – pre-existing and not gathered for the purpose of investigating causes of aggressive behaviour.

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

2 marks for a clear, coherent explanation of how thematic analysis could be carried out in this study

1 mark for a limited/muddled explanation.

Possible content

- use the recordings to make a transcription of the interview
- use coding to initially analyse the transcripts
- review the transcriptions/codes looking for emergent themes/ideas that might be linked to later aggressive behaviour, eg family violence, parental argument, alcohol misuse.

No credit for answers referring to **pre-determined** categories, codes, themes and/or counting frequencies.

STEMUP.CO.UK

Q16.

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

Level	Marks	Description
2	3 – 4	Knowledge of the difference between primary and secondary data is clear and mostly accurate. The knowledge is applied appropriately. The answer is generally coherent with effective use of terminology.
1	1 – 2	Knowledge of the difference is evident but the answer lacks accuracy and detail. Application is limited, inappropriate or absent. Use of terminology is either absent or inappropriate.
	0	No relevant content.

Application:

- primary data is gathered directly/first-hand from the participants, and is specific to the aim of the study whereas secondary data has previously been collected by a third party (another researcher or an official body), not specifically for the aim of the study, and then used by the researcher
- in this study, the interview recordings are primary data – gathered specifically for the purpose of investigating causes of offending whereas the court/prison records are secondary data – pre-existing and not gathered for the purpose of investigating causes of offending behaviour.

(b) **AO3 = 2**

2 marks for a clear, coherent explanation of how thematic analysis could be carried out.

1 mark for a limited/muddled explanation.

Possible content

- use the recordings to make a transcription of the interview
- use coding to initially analyse the transcripts
- review the transcriptions/codes looking for emergent themes/ideas that might be linked to violent crime eg family violence, parental argument, alcohol misuse.

No credit for answers referring to **pre-determined** categories, codes, themes and/or counting frequencies.

Q17.

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

Level	Marks	Description
2	3 – 4	Knowledge of the difference between primary and secondary data is clear and mostly accurate. The knowledge is applied appropriately. The answer is generally coherent with effective use of terminology.

STEMUP.CO.UK

1	1 – 2	Knowledge of the difference is evident but the answer lacks accuracy and detail. Application is limited, inappropriate or absent. Use of terminology is either absent or inappropriate.
	0	No relevant content.

Application:

- primary data is gathered directly/first-hand from the participants, and is specific to the aim of the study whereas secondary data has previously been collected by a third party (another researcher or an official body), not specifically for the aim of the study, and then used by the researcher
- in this study, the interview recordings are primary data – gathered specifically for the purpose of investigating causes of addiction to gambling behaviour whereas the debt counsellor reports are secondary data – pre-existing and not gathered for the purpose of investigating causes of addiction to gambling.

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

2 marks for a clear, coherent explanation of how thematic analysis could be carried out.

1 mark for a limited/muddled explanation.

Possible content

- use the recordings to make a transcription of the interview
- use coding to initially analyse the transcripts
- review the transcriptions/codes looking for emergent themes/ideas that might be linked to later gambling addiction, eg family poverty, sensation-seeking, influence of peer group.

No credit for answers referring to **pre-determined** categories, codes, themes and/or counting frequencies.

Q18.

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

1 mark – A

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

1 mark – value of research score = 6

1 mark – ethical concern score = 7

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

1 mark each for identifying an appropriate descriptive statistic.

Plus

1 mark each for an explanation how it could be calculated.

Possible content:

- Median could be calculated by arranging the scores in order then choosing the

STEMUP.CO.UK

numerical midpoint (between 5th and 6th score).
- Range could be calculated by taking the lowest score from the highest (and adding 1).
- Standard deviation could be calculated by using the distance of each score from the mean (credit formula for SD).

If student provides calculations that illustrate 'how' then can award the explanation mark.

Can also credit answers based on mean - Mean could be calculated by adding all the scores and dividing by the number of scores (10).

Credit other relevant descriptive statistics with appropriate explanation (eg percentages, graphs, correlation). If no calculation is involved eg with graphs then accept some description of how the representation could be arranged.

Q19.

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

2 marks for a clearly outlined disadvantage that is explicitly related to the study – sample is biased, lacking representativeness, more confident/helpful/curious which could affect how they respond to therapy or represent other offenders.

1 mark for knowledge of a disadvantage of using volunteers which could be related to the study but this is not explicitly presented.

Credit other relevant disadvantages.

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

Award **1 mark** for each of the following points up to a maximum of 3 marks.

- All the volunteers are identified either by name or number.
- The 50 names/numbers are put in a container or computer.
- Assign alternate names/numbers drawn to Group 1 then Group 2 and so on until there are 25 in each group or alternative system **OR** set parameters for two groups of 25 to be randomly generated.

Credit alternative descriptions of a practical procedure which would result in the same outcome.

(c) **[AO2 = 3]**
3 marks for a clearly stated and appropriate non-directional (or directional) operationalised hypothesis: there is a difference in the reduction in anger scores of offenders who complete Therapy A and those who complete Therapy B.

2 marks for a statement with both conditions of the IV and the DV that lacks the clarity of the 3 mark answer.

1 mark for a muddled statement with both the IV and DV present.

0 marks for expressions of aim/questions/correlational hypotheses or statements with only the IV or DV or one condition of the IV present.

Full credit can be awarded for a null version of the hypothesis.

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

STEMUP.CO.UK

Award **1 mark** for any of the following points up to a maximum of 3 marks.

Possible points:

- Both therapies seem to be effective.
- The pre-therapy scores for both groups suggest that the offenders in Group 1 were more angry on average than those of Group 2 at the start of study.
- The post-therapy scores for both groups suggest that the offenders in Group 1 were more angry on average than those of Group 2 at the end of study.
- Anger scores are lower at the end of the study for both groups suggesting both therapies reduced feelings of anger.
- The improvement in Group 1 is 50-30, a decrease of 20 and in Group 2 is 40-15, a decrease of 25, which suggests that there was a greater reduction in feelings of anger in Group 2.

Credit other relevant points.

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

Level	Marks	Description
2	3 – 4	The explanation for how matched pairs would have improved the study is clear and the answer identifies the difference in the pre-therapy scores as the issue. The answer is generally coherent with effective use of terminology.
1	1 – 2	There is limited/partial explanation for how matched pairs would have improved the study. The answer may lack coherence. Use of terminology may be either absent or inappropriate. **OR** the answer only xplains matched pairs design in the context of this study.
	0	No relevant content.

Possible points:

- The researcher wants to be sure that the final anger scores are a result of the therapy and not participant differences between the groups.
- With independent groups participant variables might have caused the lower post-therapy scores for Group 2 rather than the therapy.
- Matching the anger scores of the participants at the pre-therapy stage would reduce this difference – seen in the Group 1 median of 50 which is ten points higher than the Group 2 median of 40.
- Other variables could also affect response to therapy e.g. gender, type of offence.

N.B. Answers that only explain a matched pairs design in the context of this study cannot be awarded above level 1.

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

2 marks for a clear outline of an ethical issue that might have occurred in the study.
1 mark for a brief or muddled outline of an ethical issue that might have occurred in the study.

STEMUP.CO.UK

- In this study the participants are all in custody so despite the suggestion that they are volunteers they may have been coerced in some way to participate, this could relate to issues of consent or stress or right to withdraw.
- Distress/psychological harm such as: finding out they have high-anger scores; the questionnaire might trigger more anger; finding out there are two therapies and finding out one might be better; therapies might cause more harm than good; the participants might have some expectation that the programme would be effective and this might not be the case.
- Reducing the anger of the participants may be more beneficial for staff than the participants, manipulation of their behaviour for this reason could be unethical.

Plus
2 marks for a clear explanation of how the issue described above could be dealt with. It is expected that withdrawal of data or debriefing would be appropriate answers but these should be clearly matched with the chosen ethical issue.

1 mark for a brief or muddled explanation of how the issue could be dealt with.

Credit other relevant information.

(g) **AO2 = 1**

1 mark for knowledge of an appropriate disadvantage of the median.

Possible answers
- Any outlier values/extreme values (such as the scores of 50) would be ignored/would not form part of the average measurement.
- Less sensitive than the mean.
- It does not represent all the findings.

(h) **AO2 = 2**

2 marks for a clear explanation of how demand characteristics might have occurred in the study. The explanation will make clear that participants alter their behaviour in some way because of a belief they have about their participation in the study.

1 mark for a limited or muddled explanation that is linked to the study.

Possible points

- Participants might respond to the questionnaire with answers giving a reduction in anger as they expect the therapy is to reduce anger/they know this is what the researcher expects.
- Participants might ‘respond’ with an increase/decrease in anger in order to ‘ruin’ results.
- When the researcher gives out the questionnaire for the second time, this sets up a demand characteristic that the second answers should be different to the first answers.

Credit other relevant information.

(i) **AO3 = 2**

2 marks for a clear discussion of one strength of using questionnaires in research which might include: the ability to collect large amounts of data relatively quickly and conveniently; discussion of how they might be easy to score/collate – when items are closed; reference to standardisation/replication of method.

STEMUP.CO.UK

1 mark for limited or muddled explanation of a possible strength.

N.B If the strength discussed relates to a particular type of questionnaire, the type of questionnaire must be made clear for the second mark.

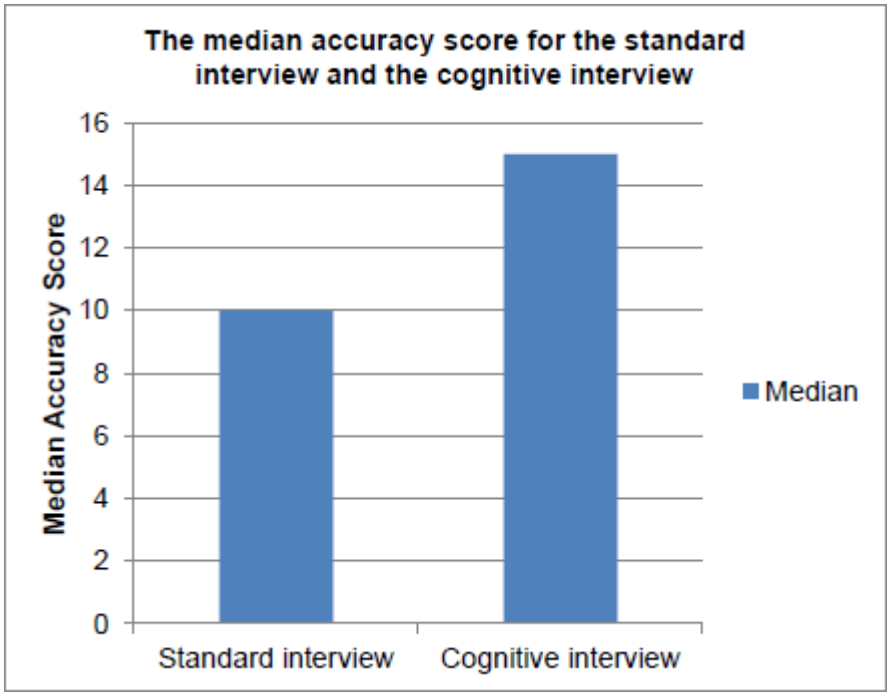
Credit other relevant strengths.

Q20.

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

1 mark for each of the following:

- display as a bar chart
- both axes labelled correctly
- an informative title with reference to the IV and DV
- y axis has appropriate scaling
- bars are separate
- bars are plotted reasonably correctly.



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

Award **one mark** for each of the following points:

- the researcher needs to ensure that the two groups are matched for key variables
- example of at least one key variable – any that might reasonably be expected to affect memory in this situation, eg eyesight, age, intelligence
- all participants should be pre-tested / assessed for the key variable / variables
- for each person in one condition, the researcher should assign a 'matched' person in the other condition.

Credit other relevant points or this information embedded in the example.

STEMUP.CO.UK

Q21.

(a) **[AO1 = 1 AO3 = 2]**

1 mark for identification of the correct experimental design – independent groups / independent measures.

Plus

2 marks for a clear and coherent outline of an advantage using appropriate terminology.

OR

1 mark for a brief / vague / muddled outline of an advantage.

Possible advantages:

- performances not affected by order effects as people only do one condition
- demand characteristics less likely as participants only aware of own condition
- same task / materials can be used in both conditions as participants are always naïve to the task.

Credit other relevant advantages.

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

2 marks for a clear and coherent outline of how participants are used in either a repeated measure or a matched pairs design.

1 mark for a vague, muddled or incomplete outline of a repeated measure or a matched pairs design.

If the answer to **(a)** is incorrect, credit a different design to that given.

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

1 mark for an appropriate and plausible suggestion.

Plus

1 mark for an appropriate justification.

Likely suggestions:

- testing all participants in the same room
- making sure that all participants hear the same instructions
- ensuring that all participants are tested by the same researcher.

Credit other relevant suggestions.

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

3 marks for an appropriate non-directional (or directional) operationalised hypothesis: ‘There is a difference in the number of ideas generated when participants work alone and when they work in groups.’

STEMUP.CO.UK

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

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

0 marks for expressions of aim / questions / correlational hypotheses or statements with only one condition.

Full credit can be awarded for a hypothesis expressed in a null form.

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

1 mark: 3 (in each group)

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

1 mark for naming a suitable measure of dispersion (range or standard deviation).

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

1 mark for stating that the statistic calculated (either the range or the SD) would be greater in **Condition A** than in **Condition B**.

or written as

1 mark for stating that the statistic calculated (either the range or the SD) would be less in C**ondition B** than in **Condition A**.

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

Marks for a clear description of a practical way as follows:

1 mark – all the participants allocated a number from 1 to 15.
1 mark – the 15 numbers are put in a hat.
1 mark – assign first three numbers drawn to a group and repeat process for other 4 groups.

Accept other valid descriptions that would be practical and produce the same outcome.

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

1 mark : for each condition, the overall number of ideas generated should be divided by the overall total of 185.

Plus

1 mark : the result for each condition should then be multiplied by 100 to give the percentage.

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

Level	Marks	Description
3	5 – 6	Both elements of required content are clear and mostly

STEMUP.CO.UK

		well detailed. The debrief is all in verbatim format.
2	3 – 4	Both elements of required content are present. The answer lacks detail and / or clarity in places. Some of the answer is in verbatim format.
1	1 – 2	There is some information about at least one element of required content. The answer lacks clarity. Verbatim format is lacking. For one mark there must be some relevant content, eg an optional point about ethics.
	0	No relevant content.

Required content:

- explanation of the aim: to see if creativity is affected by the presence or absence of others
- information about the other condition – in an independent design people need to know about the condition in which they did not take part.

Optional content:

- specific ethical issues, eg right to withdraw data / be informed of results / check of welfare
- general ethical considerations, eg respect for participants.

Q22.

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

1 mark for each of the following points:
- The total observation time for each parent was 10 minutes.
- The psychologist made 20 observations for each parent.
- To generate 20 observations for each parent she must therefore have recorded her observation every ½ minute or every 30 seconds.

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

1 mark for the correct answer: 30%.

Plus

1 mark for showing correct workings: 12 divided by 40 multiplied by 100.

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

1 mark for primary data.

Q23.

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

2 marks for a clear and coherent explanation of the usefulness of the standard deviation in this study.

STEMUP.CO.UK

1 mark for a weak or muddled answer in which the impact of the difference in the SDs is alluded to.

- Useful to inform about the spread of scores.
- Indicates participant variables – as a group the people in Condition 1 are quite different / are more variable than those in Condition 2.

Credit answers which suggest that the SDs can be used to look for similarity or differences in variance.

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

2 marks for a clear, coherent outline of a relevant problem.

1 mark for a weak, muddled or very limited outline.

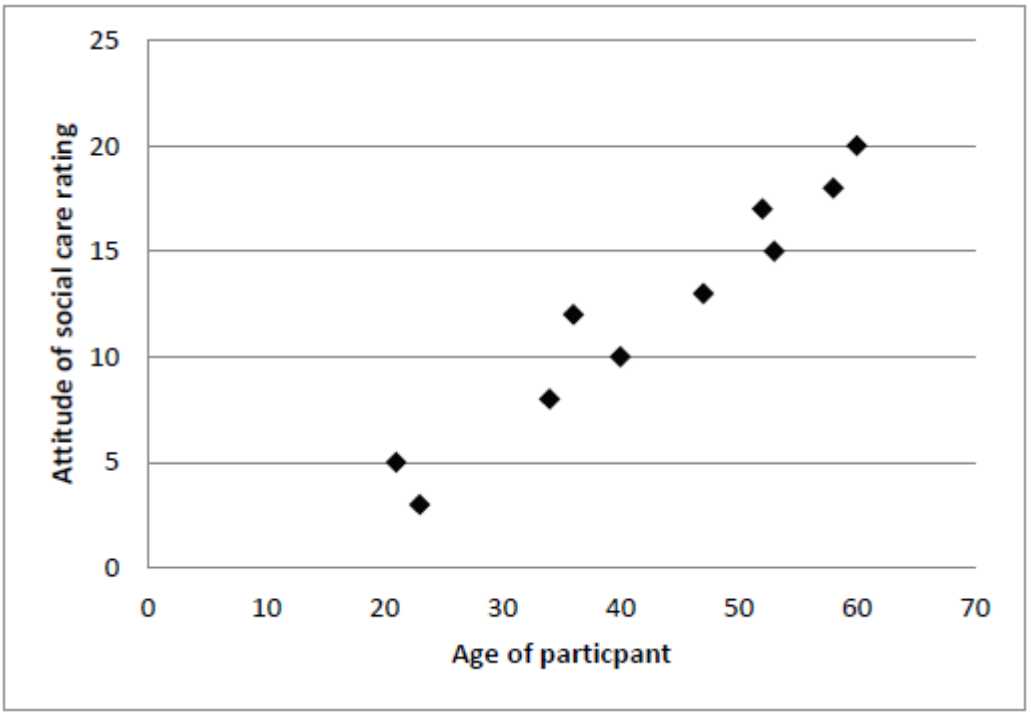
Possible problems:

- Direct observation of memory is not possible and must be inferred from the results / behaviour of the participants – this inference could be mistaken.
- The task given is rarely how normal memory functioning occurs because it is specifically designed to make measurement possible – the researcher therefore collects data that is only related to memory processing under experimental conditions.

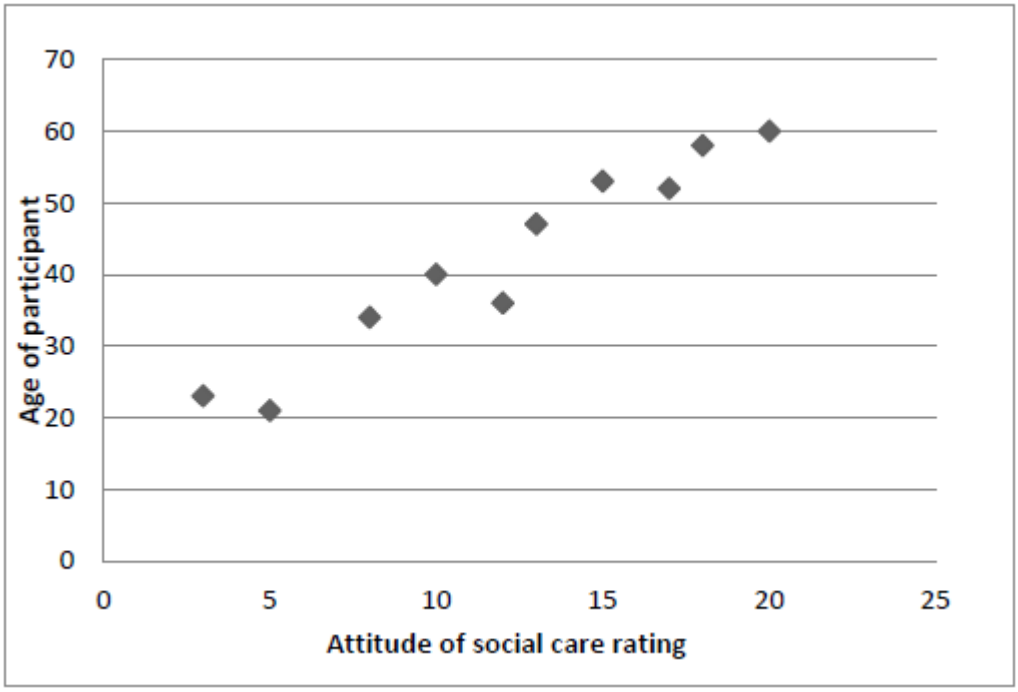
Credit other valid problems.

Q24.

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



STEMUP.CO.UK



3 marks for the following points:

- Axes correctly labelled as Age of participant and Attitude to social care rating.
- Scales are suitable.
- Points plotted accurately.

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

2 marks for: there is a **positive relationship** between age and interest in social care issues / as people get older their interest in social care increases (1) this is because as the values on one co-variable increase, so do the values on the other co-variable (1) OR as age increases so does attitude to social care **rating / score**.

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

1 mark for knowledge of an investigator effect – this is when the person collecting the data has knowledge of what the research aim is / traits and that knowledge / those traits affect the data obtained.

1 mark for a brief explanation of how investigator effects might have occurred in this study.

If the researchers believed that older people would be more interested in social care they could have just given scores based on the age of the person.

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

2 marks for explaining how investigator effects could have been avoided in the study. The answer needs to explain what could be done and how that would decrease / eliminate the effect.

Possible content:

- Discussion of separate observation by the two researchers and comparison – inter-rater reliability.

STEMUP.CO.UK

- Having ‘blind’ rating of the discussion by someone who is unaware of the aim or research hypothesis.
- Filming the discussions so there is a permanent record that can be checked by peer review of the data to confirm the scores / ratings.

Credit other relevant procedures.

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

Level	Marks	Description
2	3 – 4	Explanation of how closed and open questions are beneficial is clear. The answer is generally coherent with effective use of terminology.
1	1 – 2	There is limited / partial reference to the benefit(s) of closed and open questions. The answer lacks accuracy and detail. Use of terminology is either absent or inappropriate. **OR** answer only refers to either closed or open questions at Level 2.
	0	No relevant content.

Possible content:

- Closed questions would present participants with options for their response so the researchers would be able to collate and display the information collected easily.
- Closed questions make it easy to compare specific response to questions the researchers wanted answered – they can be sure there will be certain information because they have restricted the options to include that information.
- Open questions allow respondents to interpret the question as they wish to and develop their response with detail or depth – so there is lots of information received.
- Open questions allow the researchers to pursue a line of enquiry that they may not have predicted but which comes to light because of a response by an interviewee.

Credit other relevant procedures.

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

- **1 mark** for an appropriate open or closed question – requiring information about a social care issue.
- **1 mark** for correct identification of this as an open or closed type of question.
- **1 mark** for a suitable explanation for why the choice was appropriate – this could relate to producing a type of data (closed – ease of analysis, open – lots of detail or depth to response / allows respondent to elaborate her / his reasoning for the response given) or it could focus on an issue of social care introduced by the candidate and not in the stem.

(g) **[AO1 = 2 AO2 = 2]**

AO2

STEMUP.CO.UK

1 mark: the responses to the open questions in the interview constitute qualitative data.
Plus
1 mark: the attitudes ratings **AND / OR** the collated responses to the closed questions in the interview constitute quantitative data.

AO1
1 mark for an explanation of how the responses to the open questions is qualitative data ie is non-numeric / descriptive / retains detail of actions / thoughts / feelings.
Plus
1 mark for an explanation of how the ratings / collated responses to closed questions is quantitative data ie numerical such as a score / behaviour is represented in the form of a score on a scale.

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

2 marks for each explanation of how the chosen ethical issue could be dealt with.

1 mark for a brief muddled explanation.
2 marks for a clear explanation.

Consent – to be part of what is in essence two studies. Participants should be forewarned – a briefing.

Protection from harm – at the end of participation all will have to be fully aware that they were rated for their social care interest and a low score might indicate they are 'uncaring'. They may wish to withdraw their data.

Right to withdraw – being made aware that they can at any time stop participating and at the end of their participation they can withdraw detail of their behaviour in the research.

The explanation must demonstrate an appreciation that people should be dealt with, with respect and competence.

Credit other relevant ethical issues.

Q25.

[AO2 = 2 AO3 = 2]

Credit application of knowledge of normal distribution and how to achieve this through modification of test as follows:

1 mark each for any four of the following points:

- Adjust the difficulty of the test
- To make the test more demanding/harder
- So that fewer people get a very high mark
- Most people get a medium mark (a few people get a very low mark and a few people get a very high mark)
- So it will achieve a more normal distribution (credit use of diagram to illustrate)

STEMUP.CO.UK

Q26.

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

1 mark – correlation

Plus:

1 mark – she is investigating the (numerical) relationship between two co-variables

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

1 mark – scattergram

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

1 mark – 10 children

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

1 mark – pattern shows a (strong) negative correlation

Plus:

1 mark – suggesting that the more years are spent in an institution, the lower the language ability (or the fewer years spent in an institution), the better the language ability

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

1 mark for correct answer 14 (or 15)

Plus:

1 mark for correct workings (18 − 4) or (18 − 4 + 1)

Q27.

(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.

STEMUP.CO.UK

(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]**

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.

Q28.

(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

STEMUP.CO.UK

		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 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:

STEMUP.CO.UK

- 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:

- 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.

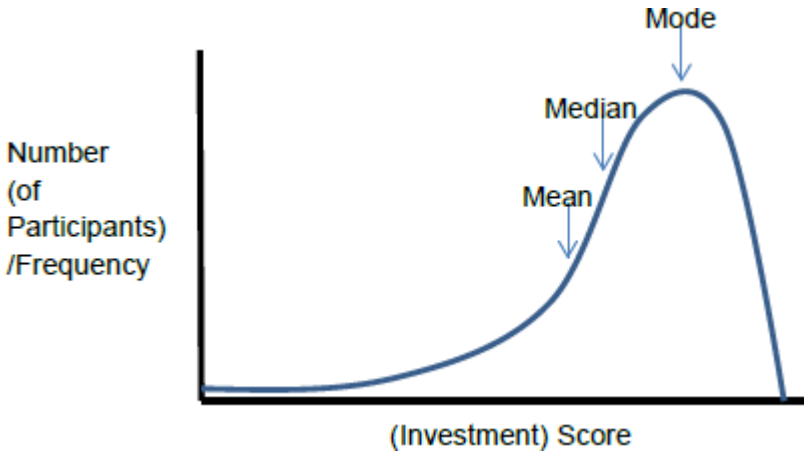
Credit answers based on any type of validity. Most answers will refer to either face or concurrent as follows:

- 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).

Q29.

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

STEMUP.CO.UK



Credit a rough sketch of a negatively skewed distribution as follows:

1 mark for shape of curve with tail to the left.

1 mark for axis labels – '(Investment) Score' on horizontal axis, 'Number (of Participants)' / 'Frequency' on vertical axis.

1 mark for positioning the mean, median and mode appropriately in relation to one another.

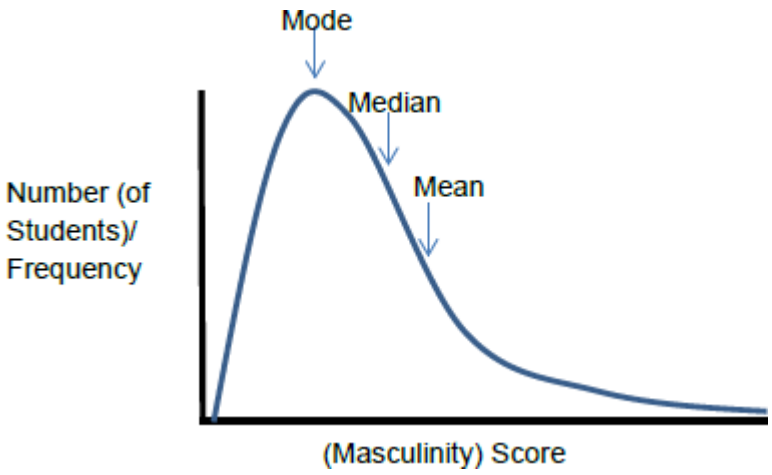
(b) **[AO2 = 1]**

1 mark for stating negative skew.

If graph sketched in (a) does not show a negative skew, credit answers that match the sketch given.

Q30.

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



Credit a rough sketch of a positively skewed distribution as follows:

1 mark for shape of curve with tail to the right.

1 mark for axis labels – '(Masculinity) Score' on horizontal axis, 'Number (of Students)' / 'Frequency' on vertical axis.

STEMUP.CO.UK

1 mark for positioning the mean, median and mode appropriately in relation to one another.

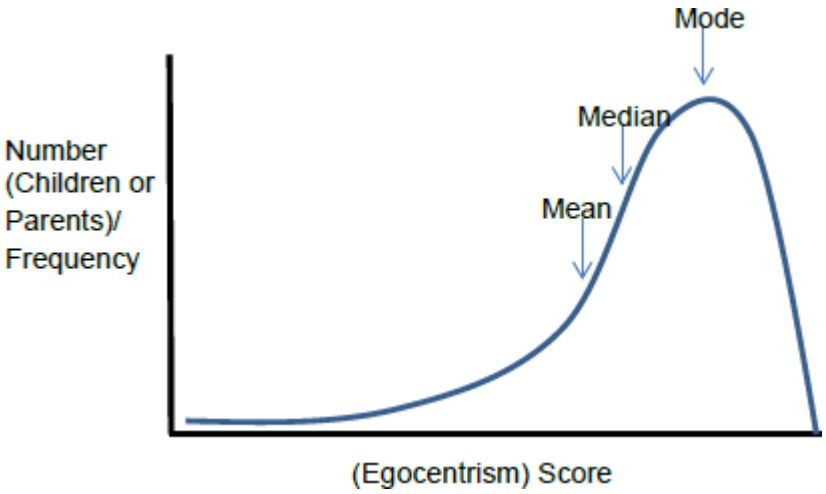
(b) **[AO2 = 1]**

1 mark for stating a positive skew.

If the graph sketched in (a) does not show a positive skew, credit answers that match the sketch given.

Q31.

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



Credit a rough sketch of a negatively skewed distribution as follows:

1 mark for shape of curve with tail to the left.

1 mark for axis labels – '(Egocentrism) Score' on horizontal axis, 'Number (Children or Parents)' / 'Frequency' on vertical axis.

1 mark for positioning the mean, median and mode appropriately in relation to one another.

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

1 mark for stating a negative skew.

If the graph sketched in part (a) does not show a negative skew, credit answers that match the sketch given.

Q32.

[AO1 = 4]

1 mark each for a correct definition of both a Type I and a Type II error
Plus
Up to 2 marks for a clear distinction between these two errors.

Possible content:

- A Type I error occurs when a researcher claims support for the research hypothesis with a significant result when the results were caused by random

STEMUP.CO.UK

variables

- A Type II error occurs when the effect the researcher was attempting to demonstrate does exist but the researcher claims there was no significance in the results/erroneously accepts the null hypothesis
- The difference is that in a Type I error the null hypothesis is rejected when it is true and in a Type II error it is retained when it is false.

Q33.

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

2 marks for an accurate comment about the means for both males and females
Plus
2 marks for an accurate comment about the standard deviations for both sets of data

Means: the mean score for males is almost 3 times larger than that of the females which suggests they are very much better at map reading than the females

Standard deviations: sds are quite similar to each other suggesting the spread of performances of the male participants and the female participants is similar within each group.

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

Award 2 marks for a correct calculation of the percentage for the male participants and 2 marks for a correct calculation of the percentage for the female participants.

If the calculation for one or both of the groups is incorrect but the procedure used is correct award 1 mark for each time this occurs to a maximum of 2 marks.

Males – 13/20 = 65%

Females 5/20 = 25%

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

Up to 2 marks for a clear comment on the data

Possible content: the difference in the percentages confirms the earlier suggestion that men are much better at map reading than women.

Q34.

[AO2 = 4]

Level	Marks	Description
2	3 – 4	Knowledge of the effectiveness of atypical and typical antipsychotics on positive and negative symptoms is clear and mostly accurate. The findings in the table are used appropriately. The answer is generally coherent with

STEMUP.CO.UK

		effective use of terminology.
1	1 – 2	Some knowledge of the effectiveness of atypical and typical antipsychotics and positive and negative symptoms is evident. Use of findings from the table is not always effective. The answer lacks accuracy and detail. Use of terminology is either absent or inappropriate.
	0	No relevant content.

- Atypical and typical antipsychotics are equally effective against positive symptoms with more than half of patients responding well
- The main difference is that negative symptoms respond better to atypical antipsychotics, 30% improve compared with typical antipsychotics 16%
- Atypical antipsychotics are more effective against negative symptoms
- These findings support the view that they act on different neurotransmitters

Q35.

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

Level	Marks	Description
2	3 – 4	Clear understanding of the notion of social sensitivity is demonstrated through effective application to the stem. Explanation of how the researchers could have dealt with the issue of social sensitivity in this case is clear. The answer is generally coherent with effective use of terminology.
1	1 – 2	Some understanding of the notion of social sensitivity is demonstrated through limited application to the stem. There is limited/partial explanation of how the researchers could deal with the issue of social sensitivity in this case. The answer lacks accuracy and detail. Use of terminology is either absent or inappropriate.
	0	No relevant content.

Content:

- Awareness of issue: Researchers should be aware of the implications of their research: possible negative impact for the children in the sample; possible negative implications of the research for the reputation of Crayford school and the wider community; possible self-fulfilling prophecy
- Dealing with the issue: Researchers should take adequate steps to counter the above: sensitive briefing/debriefing of participants, parents, teachers etc; care in relation to publication, disclosure of results and confidentiality/anonymity.

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

1 mark for nominal level/categorical level

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

STEMUP.CO.UK

1 mark – categorical data is crude/unsophisticated/does not enable very sensitive analysis

Plus

1 mark – because it does not yield a numerical result for each participant

Q1.

In part (a) Most students scored well and were able to identify the type of observation accurately. In part (b) Very few students could identify the sampling technique as event sampling. It seemed that they interpreted the word "sampling" to refer to the way in which participants were obtained. Those who did correctly identify event sampling were able to give an accurate and detailed explanation. In part (c) It was pleasing to see how well the majority of students performed. Many were able to produce accurate and detailed answers, drawing some sensible conclusions from the data. For example, the school needs to do more to encourage the recycling of food, the initiative seems to work better with younger years, so they need to try harder with the older groups. The pitfall for some students was to go beyond the data in the table and refer to conformity and obedience. For others it was to interpret the table as if it was a correlation. Part (d) asked for one way in which this particular observation could have been improved. This is not the same as asking for a completely different study, which was not creditworthy. Better answers referred to the use of more than one observer, or to carry out the observation over several different days or locations.

Q2.

Part (a) was an easy part of the question for most students. However, they do need to be very careful with the clarity of their handwriting in questions such as this one. It was sometimes difficult to determine whether the answer was 250 or 230. There were some careless maths errors, such as "losing" a zero, or by giving the answer in minutes rather than hours.

Part (b) was another part of the question students found easy. The most common answer was quantitative data because she used closed questions. Credit was also given to those who chose qualitative data because she recorded what they said; or primary data because she collected the data specifically for this study.

In part (c), while some students were able to provide accurate answers showing good understanding, others merely described what the investigator effect was, rather than how to reduce it. This is yet another example of the need to read the question carefully. Better answers were able to suggest getting someone else who had not been at the concert to ask the questions, or to avoid having the investigator present by using a questionnaire. It was very clear from the answers to part (d) that a large number of students had no idea what the purpose of a peer review is. They seemed to think that a journal was something the psychologist kept herself. However, a small minority of students were able to explain that the process was to determine whether or not the journal should publish her report. This would involve a group of experts in the same field evaluating her report to make this decision.

Q3.

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.

Q4.

Part (a) was generally answered very well and most students were able to explain why matching the students was important. However, some failed to apply their response to this specific situation as thus were limited to 1 mark.

In part (b) many students were able to identify an appropriate variable to match the students on and then go onto explain how this might affect satisfaction scores. There were many possible accurate responses with the most common being how difficult they may find the topic (often linked to IQ) and prior knowledge of business studies. A significant minority of students identified an appropriate variable but then failed to apply it to this situation and explain how it may have affected satisfaction scores. There was also a significant number of students who failed to engage with the question and simply suggested one of a number of generic variables such as IQ, gender or age. Unless these variables were clearly linked to an appropriate variable (such as linking age to interest/experience of using computers) they would not be appropriate variables to match students on in this study and thus would not receive credit. It is important to remind students to engage with the question posed and tailor their answer to that specific question not to simply produce a response of non-specific rote-learned material.

In part (c) it was clear from student responses that the vast majority of students appeared to have some understanding of the process of random allocation. However, fewer students seem to clearly understand the reasons why random allocation is used and thus failed to answer the question posed. Of those students who did answer the question correctly some failed to apply it to this situation and thus were limited to 1 mark. It is important to note that simply referring to removal of bias without any outline or application is not sufficient to gain any credit. Reference to bias needed to be clearly focused on either removal of researcher bias or removal of participant bias and should be outlined and applied for full credit.

In part (d) most students clearly understood the strengths of stratified sampling and / or the weaknesses of opportunity sampling. However, many failed to tailor this knowledge to the actual question posed and did not explain why stratified sampling might be better than opportunity sampling in this study. Most students produced limited responses where they simply explained the strengths of stratified sampling or the weaknesses of opportunity sampling. Some students did explain why stratified sampling was an improvement in general but failed to apply their reasoning to this specific design. Equally there were many responses which had implicit explanation of improvements which was not clear and so not sufficient for full marks. A significant minority of students either produced responses which had no relevant content or did not attempt the question suggesting that some students did not have the knowledge to enable them to answer this question. There was also some confusion with systematic sampling. This is particularly disappointing considering there was a similar question on the previous examination paper in the June 2017 series which also asked about how stratified sampling could be an improvement (although in comparison to volunteer sampling). It is important to remind students that just because a topic appears on an exam in one series does not mean that the same topic cannot appear on the exam in the following series. It is also important to use previous examination questions with students to ensure they are familiar with the style and type of questions which they may be presented with in their actual examination.

In part (e) the overwhelming majority of students answered correctly and there was a wide range of appropriate questions which would generate qualitative data seen in student

STEMUP.CO.UK

responses. Unfortunately, a very small minority of students wrote Likert scale questions which would have generated quantitative data.

In part (f) there were some excellent responses with a range of different appropriate problems with analysing qualitative data seen. Most students who identified an appropriate problem went on to clearly explain it and thus gain full credit. However, some students failed to address the actual question posed and explained problems either with collecting qualitative data (such as ‘it takes time for participants to complete’) or with the validity of the qualitative data (such as ‘participants might lie in their responses’). Responses must be based on analysis to gain credit and the most common error seen was where students were focusing on subjectivity of the participants when answering the questions to produce the qualitative data rather than subjectivity of the researcher when analysing the qualitative data.

In part (g) most students answered correctly and a wide range of outlined ethical issues was seen in responses. Unfortunately, a few students simply named an ethical issue without outlining it.

In part (h) most students were able to draw an accurate scattergram with two appropriately labelled axes on suitable scales. A small minority of these students also produced an accurate and appropriate title which referred to the relationship between the two co-variables. However, many students failed to either provide a title for their graph or produced an inaccurate or vague title which was not sufficient for credit. There were also errors in plotting points in some responses and some students produced very small graphs with inappropriate scaling. Some students failed to read the question correctly and did not draw a scattergram. A variety of incorrect types of graph were seen but the most common was a line graph where students had joined up the plotted points ethical issue without outlining it.

The vast majority of students scored one mark in part (i) as few responses had both components required for two marks. Most responses only included a description of the relationship between online hours and satisfaction but did not identify the relationship as a negative correlation. The description of the relationship was expressed in several different accurate ways but the most common was more online hours the less satisfied they were. A small number of students produced confused conclusions based on causal relationships which were incorrect.

Part (j) was answered very well with the vast majority of students correctly calculating both the mode and median satisfaction scores and thus being awarded full marks. A very small minority made a mistake in their calculations leading them to produce an incorrect median and/or mode. Unfortunately, a very small number of students failed to fully answer the actual question posed and calculated a mean value instead of either the median or mode.

In part (k) there was evidently much confusion amongst students as to why the median would be more representative than the mode in this data set. Most responses simply consisted of statements of how to calculate the median and the mode rather than explaining a reason why the median might be more suitable to represent the average score in this study. Some students were able to explain why the mode was less representative of the average satisfaction scores and thus produced a limited explanation but very few were able to explain why the median was more representative. Only a very small minority of students were able to explain that the median was more representative as it is numerically closer to more of the other scores in the data set compared to the mode and thus few students gained 2 marks. Some students simply repeated the question stating that the median is more representative than the mode without any explanation.

There were a lot of correct answers to part (l) and it was apparent that many students had the knowledge to be able to identify the negative skew in this data. However, when

compared to the other MCQ on this paper, more students did not attempt this question and the number of correct responses was lower. This suggests that their knowledge of negative skew was poorer than their knowledge of introspection and overt observations.

Q5.

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).

Q6.

Part (a) was answered correctly by the vast majority of students.

In (b) most students understood what is meant by a pilot study, with the vast majority of students achieving at least 1 mark but only half of these students were able to provide a reason why a pilot study should be conducted in the context of the investigation, with many giving elaborated generic responses but unable to use their knowledge flexibly to apply it appropriately. Of the approximate one quarter of students achieving full marks, the most common route was to focus on whether participants would be able to recall their dreams by 9am each morning or whether the questions asked by the interviewers were clear and appropriate for gathering the necessary detail regarding dream themes.

In (c) the vast majority of students were able to give an accurate description of qualitative data but less than half of these students were able to apply this to the practical scenario to gain the second mark. Many students wasted time giving generic strengths of collecting qualitative data.

Part (d) was overhall poorly answered. Although over two thirds of students managed to give a description of investigator effects, it was clear that many students do not fully understand the term. Students often failed to distinguish between investigator effects and deliberate fraud with others muddling investigator effects up with demand characteristics or simply stating ‘the effects of the investigators’. Application was poor with less than a third of students achieving marks for their suggestions. Most common errors were to suggest employing a blind procedure despite the stem stating that interviewers ‘did not know the purpose of the study’. Students either did not read the stem material carefully enough or do not actually understand what a blind procedure is and therefore did not spot this in the stem but instead just provided a generic response to deal with issues in research.

When students understood the concept of content analysis, part (e) was well answered; however, over one third of students failed to achieve any marks. A worrying number of students had no idea what content analysis is, with students frequently describing

STEMUP.CO.UK

correlational analysis, statistical analysis, concurrent validity, descriptive statistics and even peer review in an attempt to explain how the content (data) from an experiment could be analysed. Those that understood the term generally explained how occurrences of each category could be tallied, with better students able to explain how categories would be identified from the interview data and giving appropriate examples of these. A surprisingly small number of students seemed to have a sound practical understanding of how content analysis should be conducted and teachers should be encouraged to offer further opportunities for students to develop their understanding of content analysis through practical research.

Those students who did not know what content analysis was focussed on how reliability of experiments could be improved more generally, focussing on repeating experiments, peer review or statistically analysing the results. Those who did understand what content analysis was often demonstrated excellent knowledge, although there was some muddle between test-retest and inter-rater reliability. Unfortunately some responses were poorly focussed, concentrating on how reliability can be ensured rather than assessed.

Part (g) was well answered with the vast majority of students achieving full marks. Those students who lost marks generally lost them due to not answering the question posed, with students working out friendly rather than aggressive dreams or trying to calculate the number of males opposed to number of dreams or assuming that 375 dreams were classified as friendly as opposed to including social interactions. There were a small number of students who provided the right workings but gave the wrong answer, wasting valuable time unsuccessfully calculating the answer by hand. Students need to take care when reading questions and data given in tables and should use a calculator to check their answers to avoid losing unnecessary marks.

Part (h) was generally answered well with over three quarters of students achieving at least three marks. Most commonly students lost marks for not giving both conditions of the IV in their titles or referring incorrectly to histograms or for having the bars touching. Teachers should make students aware that with discontinuous/discrete data bars should not be touching on a graph.

Q7.

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.

Q8.

This was without doubt the worst answered question on the whole paper, with a modal mark of 0. Most students failed to address the issue of distribution and instead gave lengthy interpretations of the data in the table, comparing the measures of central tendency for the two groups and attempting to draw conclusions in relation to the study. The poor performance on this question illustrates the importance of reading questions carefully and thinking carefully about what the question means. It is likely that, having seen many examples of data interpretation questions with tables on previous papers,

students automatically thought that this was a similar question and responded according. It is also likely that many students left the examination hall believing that

Q9.

This was without doubt the worst answered question on the whole paper, with a modal mark of 0. Most students failed to address the issue of distribution and instead gave lengthy interpretations of the data in the table, comparing the measures of central tendency for the two groups and attempting to draw conclusions in relation to the study. The poor performance on this question illustrates the importance of reading questions carefully and thinking carefully about what the question means. It is likely that, having seen many examples of data interpretation questions presented in tables on previous papers, students automatically thought that this was a similar question and responded according. It is also likely that many students left the examination hall believing that they had answered this question well. A lucky minority of students who were not actually answering the question at all nevertheless managed to gain 1 mark by incidentally noting that the measures of central tendency for Group A were all about the same. An even smaller minority who understood the point of the question gained 4 marks very swiftly, often in a single sentence. A handful of students illustrated their answers with sketches of a bell-shaped curve and a positive skew.

Q10.

This was without doubt the worst answered question on the whole paper, with a modal mark of 0. Most students failed to address the issue of distribution and instead gave lengthy interpretations of the data in the table, comparing the measures of central tendency for the two groups and attempting to draw conclusions in relation to the study. The poor performance on this question illustrates the importance of reading questions carefully and thinking carefully about what the question means. It is likely that, having seen many examples of data interpretation questions presented in tables on previous papers, students automatically thought that this was a similar question and responded according. It is also likely that many students left the examination hall believing that they had answered this question well. A lucky minority of students who were not actually answering the question at all nevertheless managed to gain 1 mark by incidentally noting that the measures of central tendency for Group A were all about the same. An even smaller minority who understood the point of the question gained 4 marks very swiftly, often in a single sentence. A handful of students illustrated their answers with sketches of a bell-shaped curve and a positive skew.

Q11.

(a) A straightforward question with many students getting both marks (more than 60%). Those that didn't tended to not appreciate that their two numbers must total 100, while others got the boys and girls the wrong way round.

(b) The majority of students (more than 90%) gained both marks.

(c) Most students understood what is meant by quantitative data and what is meant by qualitative data. The difficulty they appeared to have was being able to describe a difference between them. A strategy for answering such questions could be to select an attribute and then describe how it differs. For example, ease of analysis; quantitative data is easily analysed and can be put into graphs, whereas qualitative data is much more difficult to analyse and has to be converted into quantitative data before it can be easily analysed. Reference to methods used to collect the data was not relevant.

Q12.

(a) There was evidently much confusion amongst students as to when a directional or non-directional hypothesis should be used and many failed to get any marks on this question. Those who did correctly state the hypothesis should be non-directional portrayed this is many ways, ranging from ‘the hypothesis should be non-directional’ or ‘the hypothesis should not be directional’, to simply ‘no’ in response to the question stated. However, only around half of these correct responses went on to accurately explain the reason for this in relation to the fact that there was no previous research. Many students stated that ‘the hypothesis should be directional as it was clearly seen from the data that participants…’. It appears that many have not had sufficient practical experience of selecting an appropriate hypothesis and thus did not understand how to apply their knowledge in order to accurately answer this question.

(b) The overwhelming majority of students clearly knew how to calculate a percentage and most were awarded two marks. Virtually all also provided their workings to support their answer. Those students who gave an incorrect answer but included correct workings were awarded one mark.

(c) Most students clearly understood that stratified sampling was more representative or reduces research bias but many did not refer to subsets of students in the correct proportions. A minority confused stratified sampling with systematic sampling. There were also a number of students that didn’t attempt the question, suggesting that they did not have the knowledge.

(d) The vast majority of students scored one mark for this question with their understanding that primary data is first hand data. Very few referred to it as being ‘original’ data. Even fewer students referred to primary data as data collected specifically for the research being carried out and thus few were awarded two marks. There was significant confusion in responses illustrating a lack of understanding surrounding primary data as many students were incorrectly focusing on who was collecting the data.

(e) Some good responses were provided by the minority of students who understood why the mean was the most appropriate value and could refer to it as being more representative. Of these correct responses, only around half went on to clearly explain why it was the most appropriate and thus few students achieved full marks. It was evident that the vast majority did not understand why the mean would be the most appropriate measure and stated incorrect reasons, such as ‘the mean is the most appropriate as it shows the overall average’ and so scored no marks.

(f) This question was answered very well with the vast majority of students producing correct mean values for both tasks and thus being awarded full marks. A small minority were awarded partial credit for stating an incorrect answer but with correct workings.

(g) It was very clear from responses to this question that most students did not understand the purpose of counterbalancing and so scored zero marks. Most responses incorrectly stated that the purpose was to ‘reduce, eliminate or prevent’ order effects. Students seemed to be unaware that the purpose of counterbalancing is to control or equally distribute the impact of order effects. Other responses were tailored towards an explanation of how you would do counterbalancing rather than discussing the purpose. There were also a significant number of blank spaces suggesting that some students did not have the knowledge to enable them to attempt this question.

(h) There were some excellent answers to this question with appropriate extraneous

STEMUP.CO.UK

variables (EVs) identified and elaborated. However, many students failed to achieve the third mark as they did not explicitly state what effect the EV would have on the time taken to complete the spot the difference task (DV). Many correct responses referred to the type or volume of music being played, or the difficulty of the spot the difference tasks. However, a minority of students failed to achieve any marks as their responses referred to participant variables, such as poor eyesight or IQ, which would have been accounted for through the use of a repeated measures design. Additionally, some answers were vague and referred to 'environmental conditions' as an EV without explicitly referring to the change in these conditions between the tasks and thus limited the marks awarded.

(i) Most students produced some limited explanation of how the follow up study improved upon the first study and thus achieved level 1. However, few went on to give a detailed explanation of this improvement, often only focussing on either why the follow up study was positive or why the original study was negative without comparison. Those students who achieved level 2 tended to focus on the follow-up study being more objective, a more scientific measurement or allowing cause and effect to be established, and then compared this with the relevant opposing weakness of the original study (more subjective, less scientific and only opinion so does not establish cause and effect). A minority of students mistakenly assumed that the first study generated qualitative data as it was self-report and so focused on the limitation of that versus the quantitative element of the follow-up study, and so scored no marks.

Q13.

(a) Most students were able to give a disadvantage of quantitative data, but some failed to provide appropriate application. Some common incorrect answers included disadvantages of qualitative data, advantages of quantitative data and advantages/disadvantages of questionnaires.

(b) Nearly all students managed to gain at least one mark and many gained full credit. The quantitative question suggested was generally effective, but the qualitative question suggested could have produced a quantitative response, so did not gain credit. Some students got qualitative and quantitative the wrong way round and consequently gained no marks. The most popular method of answering the quantitative part was either to give a rating scale or to pose a question followed by the words *"Yes/No"*. Although giving a closed answer response that is not officially quantitative, it was credited for this series. However, schools and colleges should ensure students are able to distinguish between quantitative and closed questions for future series.

Q14.

(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

STEMUP.CO.UK

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 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

STEMUP.CO.UK

limited or no practical understanding of what peer review involves and teachers are encouraged to address this.

Q15.

(a) Most students could happily identify the primary and secondary data in the stem material and gained marks for doing so. However, very few answers received full credit for properly explaining the difference between primary and secondary data. There was widespread misunderstanding that the distinction hinges on who collects the data, with many answers stating that primary data is that which is ‘collected by the researcher’ whilst secondary data is ‘collected by someone else’. Such answers failed to recognise the proper distinctions, ie that primary data comes directly from the source (the participants) and is collected for the purposes of the study, whereas secondary data does not come directly from the source and is not specific to the purpose of the study. Various ways of making either of these points were accepted, for example, reference to primary as original data and secondary as not original.

(b) Many students left this question blank and most of those who did attempt an answer based their response on some form of quantitative analysis, referring to categories being established prior to analysis and tallying or counting frequencies. Very rarely did students refer to transcription of the recordings, coding or emergent themes.

Q16.

(a) Most students could happily identify the primary and secondary data in the stem material and gained marks for doing so. However, very few answers received full credit for properly explaining the difference between primary and secondary data. There was widespread misunderstanding that the distinction hinges on who collects the data, with many answers stating that primary data is that which is ‘collected by the researcher’ whilst secondary data is ‘collected by someone else’. Such answers failed to recognise the proper distinctions, ie that primary data comes directly from the source (the participants) and is collected for the purposes of the study, whereas secondary data does not come directly from the source and is not specific to the purpose of the study. Various ways of making either of these points were accepted, for example, reference to primary as original data and secondary as not original.

(b) Many students left this question blank and most of those who did attempt an answer based their response on some form of quantitative analysis, referring to categories being established prior to analysis and tallying or counting frequencies. Very rarely did students refer to transcription of the recordings, coding or emergent themes.

Q17.

(a) Most students could happily identify the primary and secondary data in the stem material and gained marks for doing so. However, very few answers received full credit for properly explaining the difference between primary and secondary data. There was widespread misunderstanding that the distinction hinges on who collects the data, with many answers stating that primary data is that which is ‘collected by the researcher’ whilst secondary data is ‘collected by someone else’. Such answers failed to recognise the proper distinctions, ie that primary data comes directly from the source (the participants) and is collected for the purposes of the study, whereas secondary data does not come directly from the source and is not specific to the purpose of the study. Various ways of making either of these points were accepted, for example, reference to primary as original data and secondary as not original.

(b) Many students left this question blank and most of those who did attempt an answer based their response on some form of quantitative analysis, referring to categories

being established prior to analysis and tallying or counting frequencies. Very rarely did students refer to transcription of the recordings, coding or emergent themes.

Q18.

(a) Just under a third of the responses were incorrect.

(b) The majority of answers were correct, but a number of students suggested incorrect modes. A few students calculated means here.

(c) Many students gained full marks for this question by offering the mean and median as alternatives. Some explanations suggested that it is necessary to arrange scores in chronological (rather than numerical) order to work out the median. Of those using the range, several mistakenly asserted that the range should be calculated by taking the highest number from the lowest. Very few students referred to the standard deviation or graphical representations. Where students showed explicit calculations these were used to credit the mark for explaining how the statistics should be calculated in the absence of any verbal explanation. Those few students who did not answer were presumably confused by the term 'descriptive statistics'.

Q19.

(a) For full credit students needed to apply their knowledge of a problem associated with volunteers in this particular study. Many only referred to bias or representativeness and did not relate their responses to type of offender/ possible effect on therapy or on anger scores.

(b) There were many different suggestions given for random allocation and all suggestions that were practical and would result in the required groups were credited. However, many responses never went beyond the idea of putting names in a hat or into a random number generator, meaning that all the researcher would have is names in a hat or in a generator. Students should be able to describe accurately what they would do to reach the final stage of having the required number of people in the required groups. It was very clear that many students had no idea of what they would have to instruct the computer or random number generator to do in order to reach this stage.

(c) Full marks were rarely achieved as most responses failed to show understanding of what was being measured in this study (the DV) which was a **reduction** in anger scores not anger scores or anger. Some students thought the study was looking for a difference between the start and end of therapy rather than between two different therapies.

(d) Many students failed to write enough about the data. Instead, they seemed to be providing answers to a question that required justification of one chosen suggestion, often just referring to the greater reduction in Group 2/Therapy B indicating that it was more effective. There were many points that could have been extrapolated from the data in the figure.

(e) Many students overlooked the most significant improvement to be derived from using a matched pairs design, ie the matching on pre-therapy anger scores. These students resorted to explaining matched pairs in a generic way without reference to the criteria which could be used for matching. Instead they often suggested IQ, age and sex as the crucial points, rather than pre-therapy anger scores. While credit was available for demonstrating an understanding of matched pairs design these responses did have to be contextualised and unfortunately many answers were generic.

STEMUP.CO.UK

(f) It was surprising that a number of students missed the second part of this question. An appropriate ethical issue was outlined (often fully) but then there was a failure to explain how this issue could have been dealt with. The importance of keeping the whole question in mind is highlighted in this case. Some responses presented a muddle of possible issues and markers had to try to match an appropriate issue with its appropriate solution. More worrying were the responses where one issue was described and a different issue was dealt with.

(g) Answers were often disappointing as students seemed unaware that median values are a form of average and stated that a problem was the median is not an average. Others made vague statements about how the median doesn't use all the values, but it would be difficult to find a median without the other values.

(h) This was generally answered well by the majority of students. However, there were quite a number of responses where knowledge was evident but marks were not obtained because the answer was not contextualised.

(i) It was again difficult to give full credit to responses in some cases as they were often vague or undeveloped so the information about the strength of using questionnaires as a technique for data collection often lacked clarity.

STEMUP.CO.UK

Q1.

[AO2 = 4]

Content

- Median is 11 for Group A (9 + 13/2) and 8.5 for Group B (8 + 9/2)

1 mark for each accurately calculated median

Plus

2 further marks for explaining that the median is used because of the outlying/extreme scores (one in each group) which would have distorted the mean.

Also accept answers based on unsafe level of measurement.

Q2.

[AO2 = 4]

Content:

- Median is 34.5 for Group A (32 + 37/2) and 50.5 for Group B (45 + 56/2)

1 mark for each accurately calculated median

Plus

2 further marks for explaining that the median is used because the level of measurement is not interval – ratings data with units of variable size.

Q3.

[AO2 = 4]

Possible content:

- Median is 29.5 (29 + 30/2) for Group A and 24.5 (24 + 25/2) for Group B

1 mark for each accurately calculated median
2 further marks for explaining the median is the more appropriate measure because of the outlying extreme scores in each group which could have distorted the mean.

Accept answers based on unsafe level of measurement.

Q4.

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

One mark for knowledge of inter-rater (observer) reliability (this may be named or implicit in the application)

Plus

Up to two marks for applying knowledge of how reliability of the category system

STEMUP.CO.UK

could be assessed in this study.

For example:

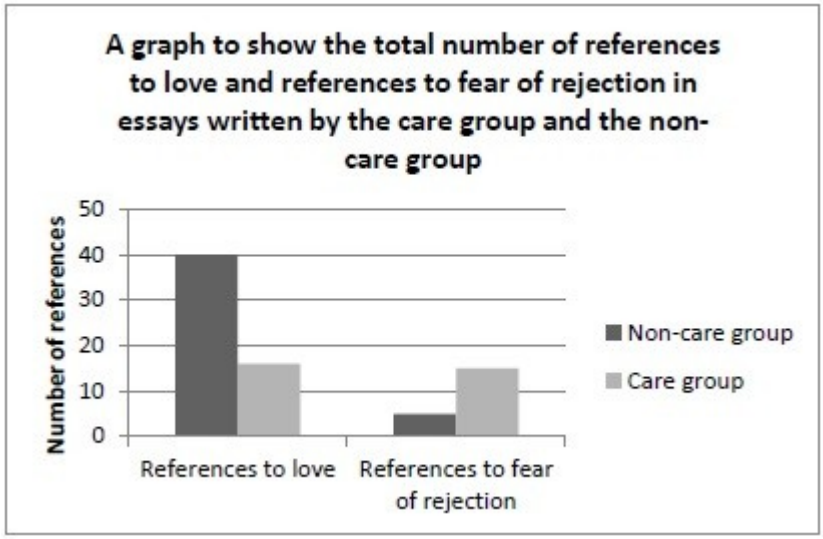
The consistency between the recordings of two researchers would be assessed by asking each researcher to categorise the essays independently for 'references to love'. The data would then be compared and if similar then the analysis is reliable.

Credit reference to the fact that the correlation coefficient between the researchers' data should exceed 0.8.

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

Up to 3 marks for an accurate sketch of a bar chart to represent the data in the table

- **1 mark** for an appropriate title
- **1 mark** for correctly labelled axes
- **1 mark** for accurately plotted bars (4 in total)



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

- **1 mark** for calculating the total number of references to love overall (16+40 = 56)
- **1 mark** for representing the total number of references to love in the care group as a fraction of the total number of references to love overall (16/56)
- **1 mark** for dividing both parts of the fraction by the lowest common denominator (16/8 = 2; 56/8 = 7) to produce the correct answer (2/7)

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

3 marks for clear explanation plus coherent elaboration

2 marks for explanation plus some elaboration

1 mark for vague/muddled explanation

STEMUP.CO.UK

Students may make a single, elaborated point or a number of points in less detail.

Possible content

- The interviewer could have asked follow-up questions to gain greater insight into some of the points raised in the essay
- The interview would produce more detail/depth of information than the participant would have been able to reveal within the essay
- Interview data would have provided a further measure of reliability/validity of the original essay content
- The interviewer may have gained rapport with the participant so they have felt comfortable revealing more personal/sensitive information face-to-face

Accept other valid points.

(e) **[AO1 = 4 AO2 = 2]**

Level	Marks	Description
3	5 – 6	Knowledge of Bowlby's monotropic theory is clear and accurate. There is appropriate use of the data in the table and clear links made to Bowlby's theory. The answer is clear and coherent. Specialist terminology is used effectively.
2	3 – 4	Knowledge of Bowlby's monotropic theory is present though there may be some inaccuracy/lack of clarity. There is some attempt to use the data in Table 1 though the links to Bowlby's theory may not always be clear. There is some appropriate use of specialist terminology.
1	1 – 2	Knowledge of Bowlby's theory is briefly stated with little elaboration. The use of data in Table 1 in the context of Bowlby's theory may be inappropriate or absent. The answer may include inaccuracies and be poorly organised. Specialist terminology is either absent or inappropriately used.
	0	No relevant content.

Possible content:

Knowledge of Bowlby's monotropic theory:

- Unique 'monotropic' relationship with mother-figure
- Notion of critical/sensitive period up to approx. 2 years
- Importance of internal working model for future relationships
- Social releasers to facilitate bond
- Reciprocal process

STEMUP.CO.UK

Accept other valid points.

Possible application:

- The care group made fewer references to love in their essays/more reference to fear of rejection which suggests their lack of monotropic bond may have affected the quality of their future relationships
- Failure to form internal working model in the care group
- Adopted at 5 so missed critical period for formation of bond

Q5.

(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.

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.

STEMUP.CO.UK

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

STEMUP.CO.UK

(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
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.

STEMUP.CO.UK

- 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

Q6.

[AO2 = 6 AO3 = 6]

Level	Marks	Description
4	10 – 12	**All three** elements are present. Suggestions are generally well detailed, practical and justified showing sound understanding of experimental design and data analysis. There is sufficient information for the study to be implemented. The answer is clear and coherent. Specialist terminology is used effectively. Minor detail and/or justification is sometimes lacking.
3	7 – 9	**All three** elements are present. Suggestions are mostly sensible and practical showing some understanding of experimental design and data analysis. There is some appropriate justification. Implementation of some aspects is possible but detail is lacking. The answer is mostly clear and organised. Specialist terminology is mostly used

STEMUP.CO.UK

		effectively.
2	4 – 6	**At least two** elements are present. Some suggestions are appropriate but others are impractical or inadequately explained. Justification is often incomplete or inappropriate. Implementation would be difficult. The answer lacks clarity, accuracy and organisation in places.
1	1 – 3	**At least one** element is present but knowledge is very limited. Justification is either absent or inappropriate. Implementation would not be possible. The answer as a whole lack clarity, has many inaccuracies and is poorly organised.
	0	No relevant content.

Three elements to be credited:

Design

- The experimental design to be used (repeated/independent/matched).
- IV and DV – note the DV must be suitable for ordinal level analysis
- Any relevant aspect of control eg duration of study, control of relevant environmental variables eg office heating, noise pollution – note this need not be exhaustive

Materials/Apparatus

- The self-devised instrument for measuring mood should be one that yields ordinal level data
- A rating scale is the most suitable measure eg ratings from 1–10 where 1 = very unhappy and 10 = very happy. Statement(s) from the rating scale should be outlined
- Alternatively students could describe a questionnaire and give examples of suitable items

(Note – although essential to the study plants need not be described)

Data analysis

- Descriptive statistics should include a measure of central tendency and dispersion (given the requirement for an ordinal level measure the most appropriate here would be the median and range but can award some credit for other measure of central tendency and dispersion)
- Suitable inferential analysis would be a test for differences between two conditions suitable for data at the ordinal level (Mann Whitney or Wilcoxon). Whichever test is chosen it should be consistent with the proposed experimental design.

(Note - descriptive statistics might also include appropriate graph/bar chart but this is not essential)

Q7.

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)

Q8.

(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)

Q9.

(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)

Q10.

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.

AO3 = 4

- A smartly dressed confederate elicits more obedience (1 mark) than a casually-dressed confederate (second mark)
- The type of task (request) also influences rate of obedience (1 mark)
- If told to do something that requires effort (e.g. heavy task) obedience levels are not affected by what the person is wearing (2 marks)

AO3 Interpretation of data
4 marks Effective interpretation of data Effective interpretation that demonstrates sound knowledge of what the data shows, with reference to both what the confederate is wearing and type of task.
3 marks Reasonable interpretation of data Reasonable interpretation of what the data shows, with reference to what the confederate is wearing and the type of task, but one in more detail.
2 marks Basic interpretation of data Basic interpretation of what the data shows; in terms of, for example “more” or “less”.

STEMUP.CO.UK

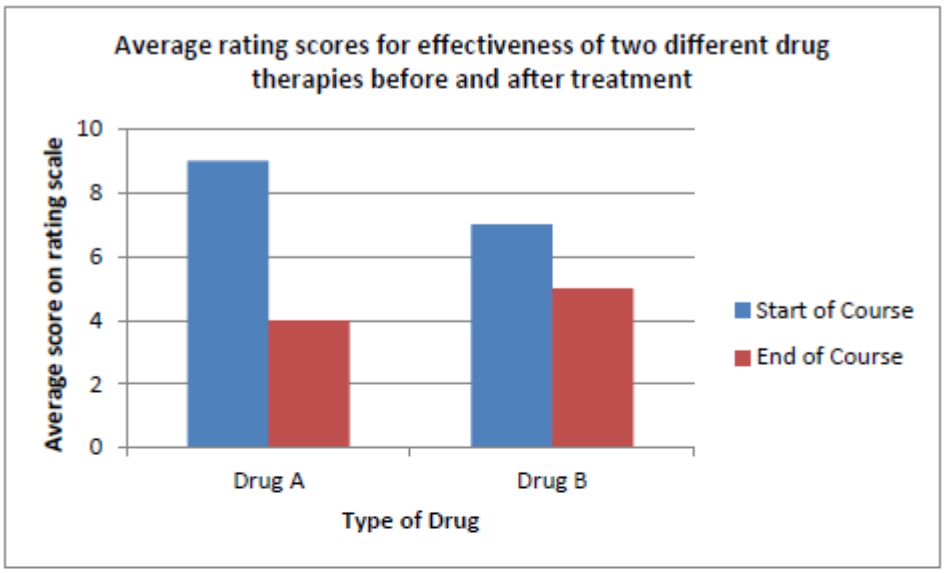
1 mark Rudimentary interpretation of data Rudimentary, muddled interpretation of the data, demonstrating very limited knowledge.
0 marks No creditworthy material.

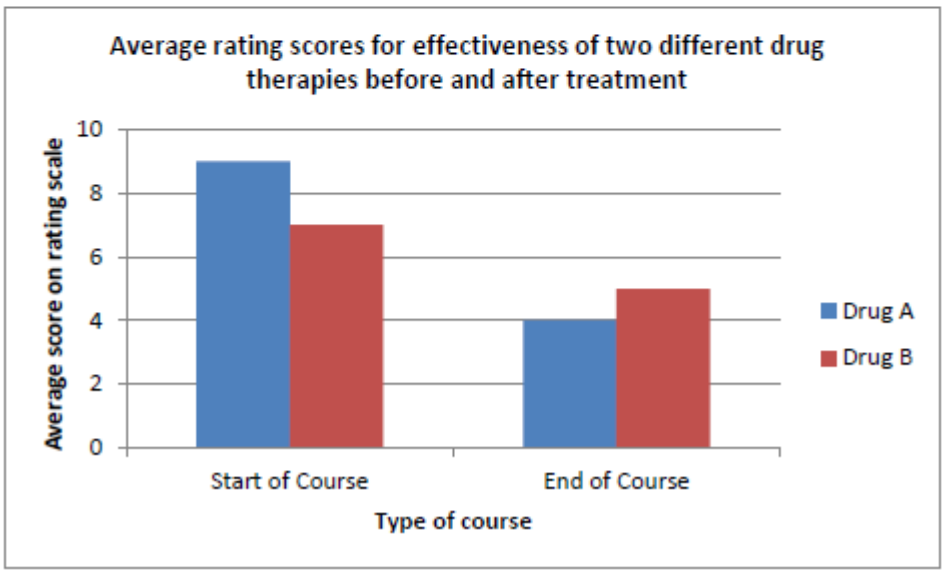
Q11.

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.

AO3 = 4





1 mark for correctly labelled x-axis: either with over-arching label, e.g. Type of Drug, or by clearly labelling the 2 conditions e.g. Start and End of course.

STEMUP.CO.UK

1 mark for correctly labelled y-axis: average score on (rating) scale.

Up to 2 marks for clearly sketching a bar chart. For full marks, there needs to be an appropriate use of graph paper and bars labelled correctly.

Q12.

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.

Q13.

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]**

A median score is calculated by putting all the scores in order from lowest to highest (or vice versa) and finding the middle score in the set.

Credit explanations that refer to sets with even numbers of values, ie by finding the numerical mid-point between the two middle scores.

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

One mark for identification of the dependent variable:

The estimate of how many marks would be scored on the maths test.

Accept alternative wording.

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

STEMUP.CO.UK

2 marks for a clear, testable statement containing both conditions of the IV and an operationalised DV.

There is a difference in the estimates men will give of their scores on a maths test and the estimates women will give of their scores on a maths test. (Accept a null version)

Men will give higher (or lower) estimates of their scores on a maths test than the estimates women will give of their scores on a maths test.

For 1 mark – a statement with both conditions of the IV and a DV which may not be operational or testable.

No marks for expressions of aim / questions / correlational hypotheses or statements with only one condition.

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

One mark for identification of the experimental design as independent groups / measures / samples. No credit if incorrect design is chosen.

One mark for an explanation of this design.
The participants in the male group are different people from those in the female group or reference to only participating in one condition.

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

Up to 3 marks for a clear explanation of the procedure for obtaining a random sample in this study:

- put all the names / numbers of the men (from the factory) in a hat / computer (1)
- draw out 15 names for the sample or get computer to randomly generate 15 numbers (1).
- repeat for all the women (of the factory) (1) or vice versa.

Accept other ways that would clearly generate a random sample.

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

Up to 2 marks for a suitable conclusion drawn from the data in **Table 2.**

One mark for the conclusion:
There is a difference in the expectations men and women have of their own numeracy skills.
Accept a directional statement such as:
Men think they are likely to achieve better scores on a maths test than women think they will achieve.

One mark for the justification:
The median estimate for men was much higher than that for the women.

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

Up to 3 marks for what comparison of the estimated and actual maths scores of the men and women indicates.

Men overestimated their numeracy skills / numerical ability / score they would

STEMUP.CO.UK

get (or similar) (1 mark)

Women underestimated their numeracy skills / numerical ability / score they would get (or similar) (1 mark)

Overall conclusions:

- People / Men / Women are not very good at estimating our ability
- Expectations were wrong
- (Although estimates are different) ability was the same
- Men are over confident re ability
- Women under confident re ability

(1 mark for any of these)

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

Information must be written in verbatim form for more than 1 mark.

Essential points	Purpose / aim of the study Ethical point
Optional points	Background information / Elaboration of the aim and conditions Any questions? Thanking for participation Interest in the results?

4 marks	**Verbatim** Both essential and at least one optional point are addressed clearly such that an understanding of the study is achieved. Information should be clear, relevant, sensible and logically structured.
3 marks	**Verbatim** Both essential points are addressed such that there is reasonable understanding of the study. There may be deficiencies in clarity, some irrelevance, illogical sequencing or inappropriate content.
2 marks	**Verbatim** Any 2 points are addressed. There may be omissions / irrelevances / muddle such that understanding of the study might be limited.
1 mark	There must be at least one relevant point. Information may be unclear / inappropriate / irrelevant such that understanding of the study would be very limited or most points addressed but not in verbatim form.
0 marks	No relevant information. Understanding of the study would not be possible.

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

One mark for a valid reason and a further mark for elaboration of the reason given.

STEMUP.CO.UK

Reason: to identify any possible flaws in (the design of) the study (1 mark).
Elaboration: to provide an opportunity to improve the study / or so that the researcher does not waste time collecting data that will have to be discarded (1 mark).

Accept any other valid answer.

Q14.

(a) **AO2 = 3**

The answer must clearly relate to one or more of the main techniques used in a cognitive interview (other than report everything):-

Context reinstatement
Recall from a changed perspective
Recall in reverse order

Some of the main additional features of the enhanced cognitive interview could be relevant, as long as it could be explained to the participant: – eg Encourage to relax

1 mark for identification of a relevant cognitive technique.
1 mark for very brief statement eg “tell me what you saw in reverse order”.
Second mark for appropriate elaboration eg “Tell me what you saw on the film in a different order to how it actually happened.” If instructions are not suitable to be read out maximum 1 mark for this part.
For 3 marks technique and instructions must match.

(b) **AO3 = 2**

The researcher might conclude that the cognitive interview was effective because more correct items were recalled, but it did not affect the number of incorrect items recalled.

0 mark - the cognitive interview was effective with no explanation.
1 mark - it was effective because there were more correct items recalled or it was not effective because the number of incorrect items stayed the same.
2 marks - it was effective because there were more correct items recalled **and** the number of incorrect items stayed the same / didn’t increase.

1 mark for stating there were more correct items recalled with the cognitive interview than with the standard interview **and** the number of incorrect items recalled was the same. (There is no reference to effectiveness).

Q15.

AO3 = 2

The graph indicates a fairly strong, positive correlation between scores on a stress questionnaire and days off through illness. The following can all receive a mark: direction, strength and a description of their relationship. Credit can also be given for mentioning the flattening of the graph at higher stress levels.

Q16.

Please note that the AOs for the new AQA Specification (Sept 2015 onwards) have

STEMUP.CO.UK

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 / 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

STEMUP.CO.UK

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.

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

STEMUP.CO.UK

- 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.

AO3 = 4

The data suggest that the confederates have a considerable influence on whether or not the participant obeys; candidates could consider the implications of the difference between 92.5% and 10%. They might consider whether the confederates are acting as role models, informing the participant how to behave. Credit could also include comparison of power of confederates with power of having the experimenter in the same room.

The question is not just asking candidates to describe the data in the table, but to consider the effect that the confederates have, to access the top bands answers need to be shaped to fit the question.

AO3 Interpretation of data
4 marks Accurate and reasonably detailed Accurate and reasonably detailed answer that demonstrates sound knowledge and understanding of what the data suggest about obedience. There is appropriate selection of material to address the question.
3 marks Less detailed but generally accurate Less detailed but generally accurate answer that demonstrates relevant knowledge and understanding. There is some evidence of selection of material to address the question.
2 marks Basic Basic answer that demonstrates some relevant knowledge and understanding but lacks detail and may be muddled. There is little evidence of selection of material to address the question.
1 mark Very brief/flawed or inappropriate Very brief or flawed answer demonstrating very little knowledge. Selection and presentation of information is largely or wholly inappropriate.
0 marks No creditworthy material.

Q18.

(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.

STEMUP.CO.UK

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**

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.

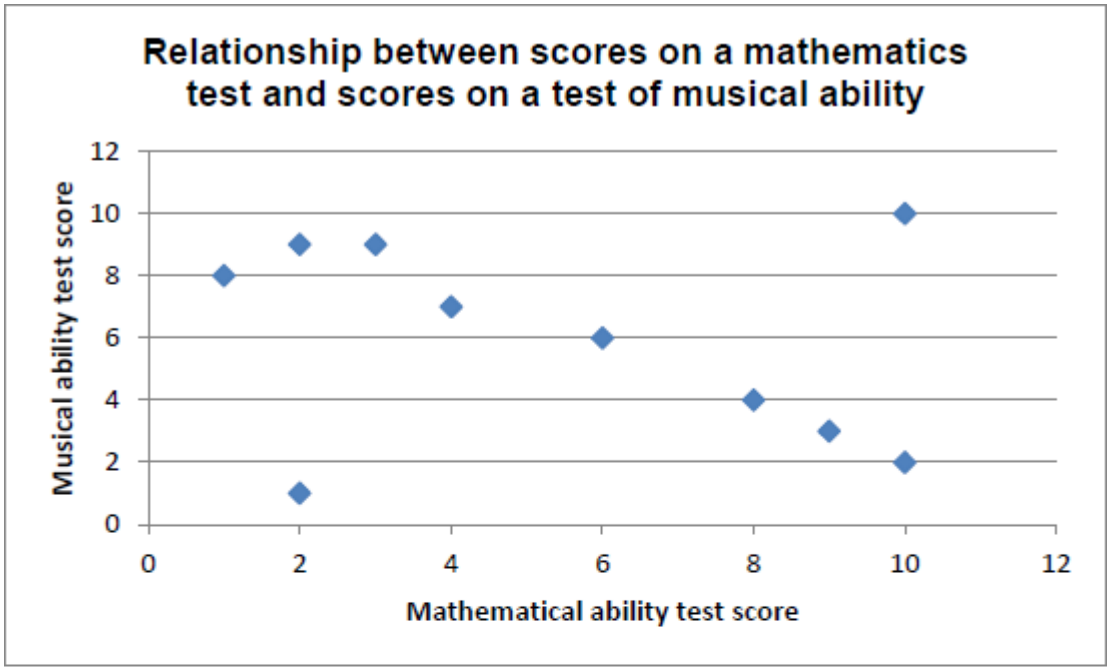
No credit for definition of a random sample.
1 mark for a brief or muddled reason (it is not biased).
2 marks for a reason that clearly points to an advantage of random sampling. This could be achieved through a comparison with another method (it is less likely to be biased than a volunteer sample).

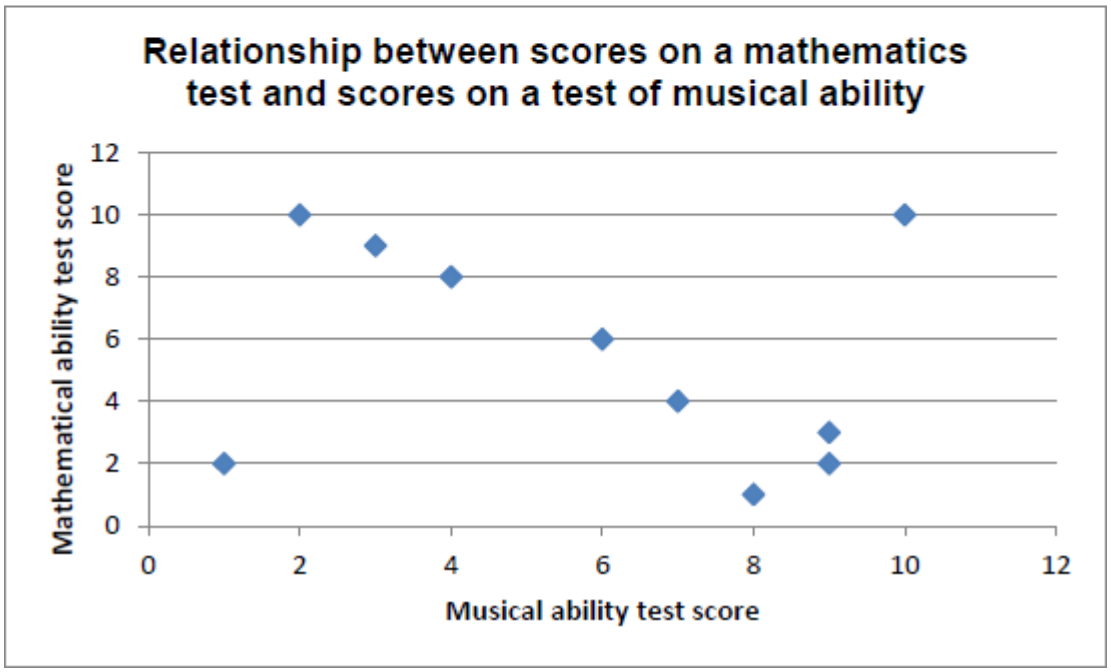
STEMUP.CO.UK

(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.





(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.

STEMUP.CO.UK

- 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.

- 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		
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.

Q19.

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.

[AO3 = 2]

One mark for knowledge of what is meant by quantitative data: data in numerical form; data representing how much there is of something.
The second mark is for an example of a specific behaviour: such as Pavlov - a measure of the amount of saliva produced by the dog; or Skinner, the time taken to respond to a stimulus (to peck a button / press a lever / jump over a barrier / reach a goal box etc) or number of responses to a stimulus (pecks, lever presses in a given time period).

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.

[AO3 = 2]

One mark for each reason applied to the behaviourist approach. For two valid points only - allow 1 mark. Likely answers will probably include: subjectivity of qualitative data / open to interpretation; cannot be replicated; not open to quantification and statistical analysis; specific so not amenable to generalization; not associated with the scientific approach.
Credit relevant comparison to quantitative data collection.

STEMUP.CO.UK

Q21.

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 / 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

STEMUP.CO.UK

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.

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) **AO3 = 4**

The independent variable is age at which the children started day care, or whether the children started day care before or after age 2.
1 mark for age.

STEMUP.CO.UK

2 marks where the IV is operationalised as above.
The dependent variable is aggression score as assessed by the researcher.
1 mark for aggression.
2 marks for aggression score, measure of aggression, level of aggression.

If either IV or DV is identified but not entirely clear – 1 mark.

(b) **AO3 = 2**

1 mark for pointing out the difference is small or the age of starting day care didn't make much difference to mean aggression score.
1 mark for stating the children who started day care before age 2 had a higher mean score than those who started after the age of 2.
1 mark for saying both groups mean score was approximately half the maximum.
Maximum 2 marks.
Eg "The mean aggression score was slightly higher for children who started day care before the age of 2." 2 marks
Candidates can gain 2 marks by two brief points or one point elaborated.

(c) **AO3 = 1**

1 mark for: range
semi-interquartile range
interquartile range
standard deviation or
variance.

Do not credit: deviation or interquartile.

(d) **AO3 = 3**

0 marks if the candidate has not drawn a bar chart.
1 mark if the candidate has drawn a bar chart but the scale is clearly inappropriate and not correctly labelled.
2 marks if the candidate has drawn a correctly labelled bar chart but the scale is clearly inappropriate; or the candidate has drawn an appropriate bar chart but the labelling is incomplete eg vertical axis refers to mean score or aggression score rather than mean aggression score.
For full marks the bar chart should indicate a small difference. Both bars and the vertical axis should be correctly labelled.

(e) **AO3 = 2**

0 marks for a non-directional hypothesis or a correlational hypothesis.
1 mark if either variable is not operationalised eg day care makes children more aggressive or the answer is slightly muddled.
2 marks for eg Children who start day care before age 2 have higher aggression scores than those who start day care after age 2, or Children who start day care at a younger age will be assessed as more aggressive than children who start day care at an older age.
Credit a directional hypothesis in the opposite direction.

Q23.

AO3 = 4

The graph shows a strong negative correlation between score on depression scale and

STEMUP.CO.UK

weeks of treatment. The more treatments the lower the depression. However, there also seems to be a plateau, where between 2-3.5 weeks there is very little change in depression.

1 mark for each of the following:

- Strength (it is a moderately strong / strong correlation)
- Direction (negative)
- Description of the relationship (the longer the treatment the lower the depression score)
- Indication of plateau / change in direction.

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) **[AO3 = 2]**

For two marks, a clear, testable statement containing both conditions of the IV and an operationalised DV.

For one mark, a statement containing both conditions of the IV and a DV.

No marks for expressions of aim, questions, correlational hypotheses or statements without two conditions.

Possible answers:

Directional: Participants will successfully shoot more netballs through the hoop in the presence of an audience than in the absence of an audience.

Note: Also accept a directional hypothesis opposite to the above.

Non-directional: There will be a difference in the number of netballs successfully shot through the hoop for participants who perform in the presence of an audience and participants who perform in the absence of an audience.

Credit null hypothesis.

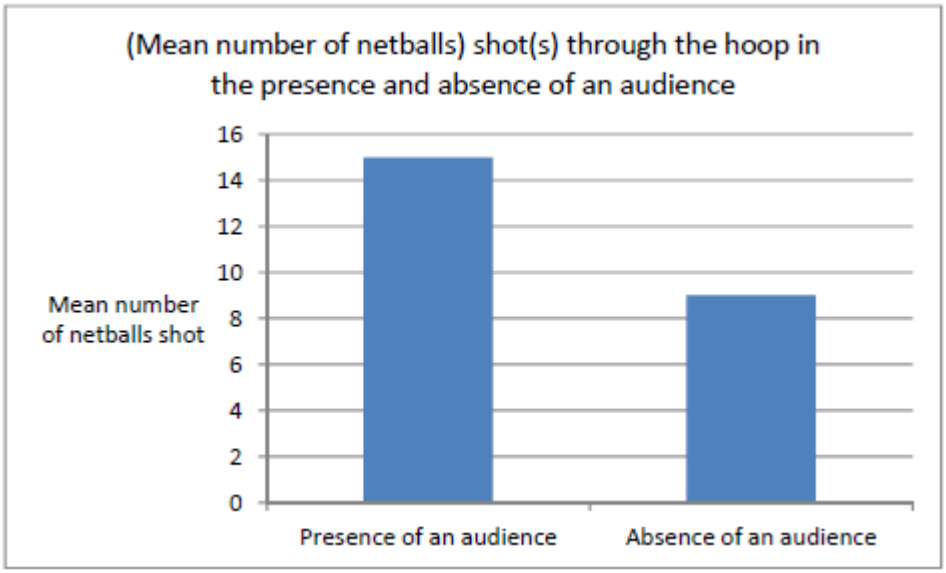
(b) **[AO3 = 2]**

STEMUP.CO.UK

One mark for an appropriate conclusion eg the presence of an audience has a positive effect on the performance of the task (or similar response) (AO3, 1).

One mark for the justification of the response eg the mean number of netballs successfully shot through the hoop is higher when there is an audience than when there is no audience (or similar response) (AO3, 1).

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



To gain the maximum three marks, candidates must provide the following:

- An appropriate title for the graphical display
- Appropriate axes and labelled eg presence / absence of an audience
- Plotting of data using a sensible scale / no penalty for joined bars.

Line graphs can be credited for title and axes only.

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

One mark for a relevant strength.
One mark for how / why it is a strength.

Likely strengths: more likely to be representative of the target population; able to generalise to the target population; to exclude researcher bias etc.

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

One mark for a relevant limitation of an independent groups design.
One mark for how / why in relation to this study eg individual differences in ability / height etc.

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

Up to two marks for an appropriate explanation of how the problem may have been overcome.

Possible answer: Using a repeated measures design (1) plus explanation of how this could be done or why this would improve the study.

STEMUP.CO.UK

Credit answers that refer to matched pairs / need for random allocation.

Can credit (ii) in respect of incorrect answer to (i).

(f) (i) **[AO3 = 1]**

One mark for correct knowledge of the term.

Possible answer: A variable other than the independent variable or an additional / or another variable that might have an effect on the dependent variable, only accept the word ‘results’ for DV (1).

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

One mark for stating that if EVs are not controlled for, then the results may be confounded (1) the researcher does not know what is causing the effect (1) or other suitable expansion eg effect on reliability or validity.

Do not accept - results will not be accurate.

Credit answers that refer to the study by way of illustration.

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

Instruction must be written verbatim for more than 1 mark.

Essential points	- reference to presence of audience - reference to shooting (hoops) throwing (hoops)
Optional	- where to stand - ethics - introduction of self

4 marks	Both essential and at least one optional point addressed clearly such that completion of the task in the experimental condition would be easily possible. Information should be clear, relevant, sensible and logically structured. Must be verbatim.
3 marks	Both essential points are addressed such that completion of the task in the experimental condition would be relatively easy. There may be deficiencies in clarity, some irrelevance, illogical sequencing or inappropriate content. Must be verbatim.
2 marks	Any two points are addressed. There may be omissions / irrelevancies / muddle such that completion of the task would be very difficult. Must be verbatim.
1 marks	There must be at least one relevant point (optional or essential). Information may be unclear / inappropriate / irrelevant / muddled such that completion of the task would be very difficult.
0 marks	No relevant information. Completion of the task would not be possible.

STEMUP.CO.UK

NB 2 - 4 marks = Verbatim Instructions

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) **AO3 = 2**

0 marks for a non-directional or correlational hypothesis.
The DV in this experiment is number of pictures correctly identified. Hypotheses where the DV is incorrect (eg number of participants who identified 10 pictures) = 0 marks.
1 mark if not fully operationalised, eg Participants who used the memory improvement strategy did better.
2 marks Participants who use a memory improvement strategy will correctly identify more pictures / objects than participants who do not use a memory improvement strategy.

(b) **AO3 = 1**

In an independent groups design a different group of participants is used in each condition.

1 mark = Different participants / people in each condition / group

Different / separate groups

Random allocation to groups / conditions.

0 marks = Different / separate conditions

Independent participants / people

Different experiments.

(c) **AO3 = 2 + 2**

Strength

The participants are naÏve because they take part in only one condition, so are less likely to show demand characteristics. There are no order effects such as practice or fatigue because participants take part in one condition.

Limitation

Individual variation, because there are different participants in each condition. More participants are needed than if a repeated measures design was used.

In each case 1 mark for very brief or slightly muddled strength or limitation, 2nd mark for appropriate elaboration of explanation.
0 marks for simply stating there are different participants in each condition.

(d) **AO3 = 3**

STEMUP.CO.UK

A pilot study is used to check aspects of the research such as whether participants understand standardised instructions, whether timings are adequate etc. It allows the researcher to try out the study with a few participants so that adjustments can be made before the main study, so saving time and money.

1 mark for a very brief explanation. Further marks for appropriate elaboration or identification of other reasons. Eg

To check it works. 1 mark
To check the standardised instructions are clear. 2 marks
To check the standardised instructions are clear enough for the participants to understand what they are required to do in the experiment. 3 marks

This question requires an explanation of why a pilot study was used, so a description of what a pilot study is (small scale study carried out before the main research) is not credit-worthy on its own. Candidates do not have to refer to a specific aspect of this experiment.

However, to gain full marks the answer must be relevant, so reference to checking sound levels for example would not be relevant.

(e) **AO3 = 2**

The standard deviation (spread of scores) is larger in the condition with the memory improvement strategy.

Candidates who use the word 'range' to suggest spread should be credited.

1 mark The standard deviation is larger in the condition with the memory improvement strategy.
2 marks The data shows the dispersion or spread of scores is larger in the condition with the memory improvement strategy.

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) **AO3 = 1**

10 children.

(b) **AO3 = 2**

Number of aggressive acts shown during a 10-minute observation. Aggression score on a questionnaire completed by parents.

1 mark for brief or slightly muddled answer, eg ask children's mothers how aggressive they are.
2nd mark for accurate elaboration, eg ask children's mothers how aggressive they are on a scale from 1 to 10.

(c) **AO3 = 3**

STEMUP.CO.UK

This is a correlational study, not an experiment, and correlation does not prove a causal relationship. Children who are already aggressive may be put into day care for longer, or another variable, such as parents' divorce, may contribute to high scores for both time in day care and aggression.

1 mark eg correlation doesn't prove cause or simply stating that aggression may be a result of other factors.
2 further marks for accurate elaboration as above.
Credit any explanation that focuses on causal inference.
An alternative approach is to argue that some research contradicts this claim. 1 mark for there is contradictory evidence and further marks for accurate reference to relevant research.
0 marks for reference to media exaggeration or individual children.

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.

AO3 = 4

- The two averages are very similar, suggesting that both therapies are as good as each other.

- The range of each group is very different. This suggests that for some people Therapy A was very beneficial, but for others it had little benefit. For Therapy B, there was a much smaller range, suggesting that it has a similar effect on improvement for all the patients.

4 marks Effective interpretation of data Effective interpretation that demonstrates sound knowledge of what the data shows, with reference to both the average and the range.
3 marks Reasonable interpretation of data Reasonable interpretation of what the data shows; or effective interpretation of either the average or the range.
2 marks Basic interpretation of data Basic interpretation of what the data shows.
1 mark Rudimentary interpretation of data Rudimentary, muddled interpretation of the data, demonstrating very limited knowledge. Or reference to, for example, larger range/higher average/similar range.

STEMUP.CO.UK

0 marks
No creditworthy material.

Q28.

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.

(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

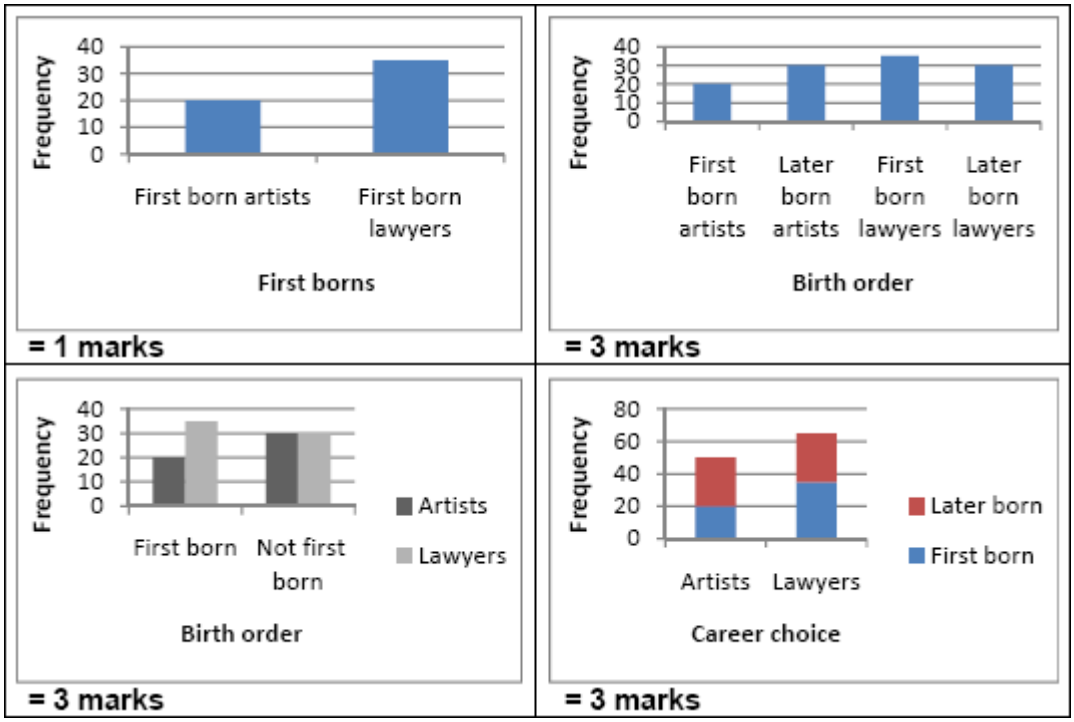
	Artists	Lawyers	Totals
First born	20	35	55
Not first born	30	30	60
Totals	50	65	115

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

STEMUP.CO.UK

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.



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.

Q29.

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 = 1**

Independent groups. Accept “independent” but not “individual”. Unrelated and between groups / subjects are also credit-worthy.

(b) **AO3 = 2**

There may be differences between the groups as there are different participants in each condition.
More participants are required than for a repeated measures design.
1 mark for very brief or muddled answer eg individual differences or needs more participants.
2 marks as above.

(c) **AO3 = 2**

The graph shows the cognitive interview is effective. There were more correct statements made after the cognitive interview than after the traditional interview.
There was no difference in the number of incorrect statements made.
1 mark for a very brief or muddled statement. Eg It shows it’s effective.
2 marks for some elaboration with reference to either correct statements, incorrect statements or both.

(d) **AO2 = 2**

The answer should clearly relate to one or more of the main techniques used in a cognitive interview:

Context reinstatement
Recall from a changed perspective
Recall in reverse order
Report everything.
The main additional features of the enhanced cognitive interview:
Encourage to relax and speak slowly
Offer comments to help clarify their statements
Adapt questions to suit the understanding of individual witnesses.

1 mark for simple identification of a relevant cognitive technique, or a very brief suggestion eg “tell me everything you saw.”
Further mark for application or elaboration. Eg “Please tell me everything you can remember about the robbery from the film you have just seen”, or “Report all the

STEMUP.CO.UK

details you can remember even if they don't seem very relevant." 2 marks.
Answers which could not relate to the film or robbery should be restricted to a maximum of 1 mark.

(e) **AO3 = 4**

Investigator effects occur when the researcher's behaviour or characteristics influence the research in some way. This includes the way the presence of the researcher may influence the participants.
0 marks for an incorrect answer or one which simply re-states the words eg "how the investigator effects research".
1 mark very brief or muddled answer eg the researcher's influence.
Maximum 2 marks for an accurate understanding of investigator effects but no reference to this experiment.
Candidates whose explanation of possible investigator effects in this experiment shows understanding of the term can be awarded 4 marks, even though there is no separate definition.
Eg The psychologist may expect the cognitive interview to be more effective than the standard interview. This may be unconsciously communicated to the participants though mannerisms such as smiling or frowning. (4 marks). There are different routes to 4 marks.
Candidates could be stronger on definitions or stronger on application.

Q30.

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.

AO3 = 4

Possible conclusions:

- Some therapies are more successful for some disorders than other therapies.
- For some disorders there is little difference between the different types of therapy.
- For disorder A, biological therapies are much more effective than any other therapy.
- Psychoanalytic therapies are not very effective, except for disorder C, but even here, there is little difference between the three types of therapy.

Findings should only be credited if they are explicitly linked to a conclusion.
Candidates could offer several conclusions or only a few, there is a breadth / depth trade-off.

STEMUP.CO.UK

4 marks Accurate and reasonably detailed Accurate and reasonably detailed answer that demonstrates sound interpretation and understanding of what the bar chart shows about the effectiveness of different types of therapies. There is appropriate selection of material to address the question.
3 marks Less detailed but generally accurate Less detailed but generally accurate answer that demonstrates interpretation and understanding. There is some evidence of selection of material to address the question.
2 marks Basic Basic answer that demonstrates some relevant interpretation and understanding but lacks detail and may be muddled. There is little evidence of selection of material to address the question.
1 mark Very brief / flawed or inappropriate Very brief or flawed answer demonstrating very little understanding. Selection and presentation of information is largely or wholly inappropriate.
0 marks No creditworthy material.

Q31.

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**

0 marks for a directional hypothesis.
1 mark if not operationalised, eg “Age affects memory.” “There will be a difference between the two conditions.”
2 marks for eg “There will be a difference in how many numbers are correctly recalled by children and adults.” “Children and adults have different short-term memory spans.” Or “The capacity of short-term memory is different for adults and children.

Candidates may write a hypothesis where the IV is how many numbers are in the list and the DV is the number of participants who can recall that digit span.
Eg As numbers in the list increase, recall changes. 1 mark.
As the number of random numbers in the list increases, the number of participants recalling the list correctly, changes. 2 marks.

(b) **AO3 = 2**

The experiment uses adults in one condition and children in the other so it would be impossible to use a repeated design unless the researchers waited for the children to grow into adults.
Given the nature of this experiment, demand characteristics and order effects are inappropriate.
1 mark for a brief explanation. A further mark for elaboration. Eg Can compare the

STEMUP.CO.UK

two different groups to see who is better. 0 marks (because this relates to all experimental designs).
They needed to have different people in each condition. 1 mark.
They needed to have different people in each condition based on age. 2 marks.
They needed to have children in one group and adults in the other. 2 marks.

(c) **AO3 = 2**

Children 6
Adults 7
1 mark for each correct answer.

(d) **AO3 = 3**

The frequency distribution shows that there is a difference in results between the two age groups.
Adults recalled more digits than children. However, the difference is small and some children recalled more digits than some adults. Candidates might refer to the modal scores being different while the range is the same.
Any credit-worthy material should be credited.
1 mark for a very brief answer eg identifying there is a difference between adults and children and / or adults score more than children. Further marks for more detail as above.

(e) **AO2 = 2**

Other research has suggested the capacity of short-term memory is 7 + / – 2. The results do support this as the range is from 5-9.
1 mark for a brief or muddled explanation eg capacity is 5-9 / other research has similar findings.
2nd mark for elaboration as above.
Candidates will be credited for reference to research such as Jacobs which found STM increases with age. However, reference to such research is not a requirement.

Q32.

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.

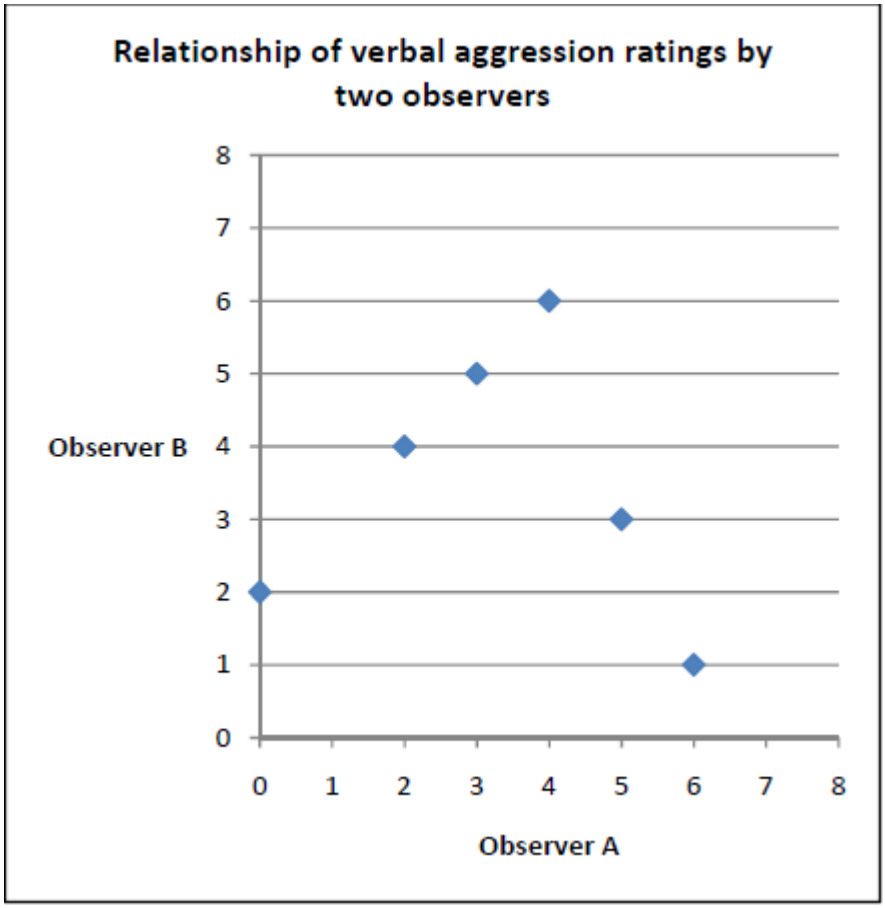
(a) **AO2 / 3 = 4**

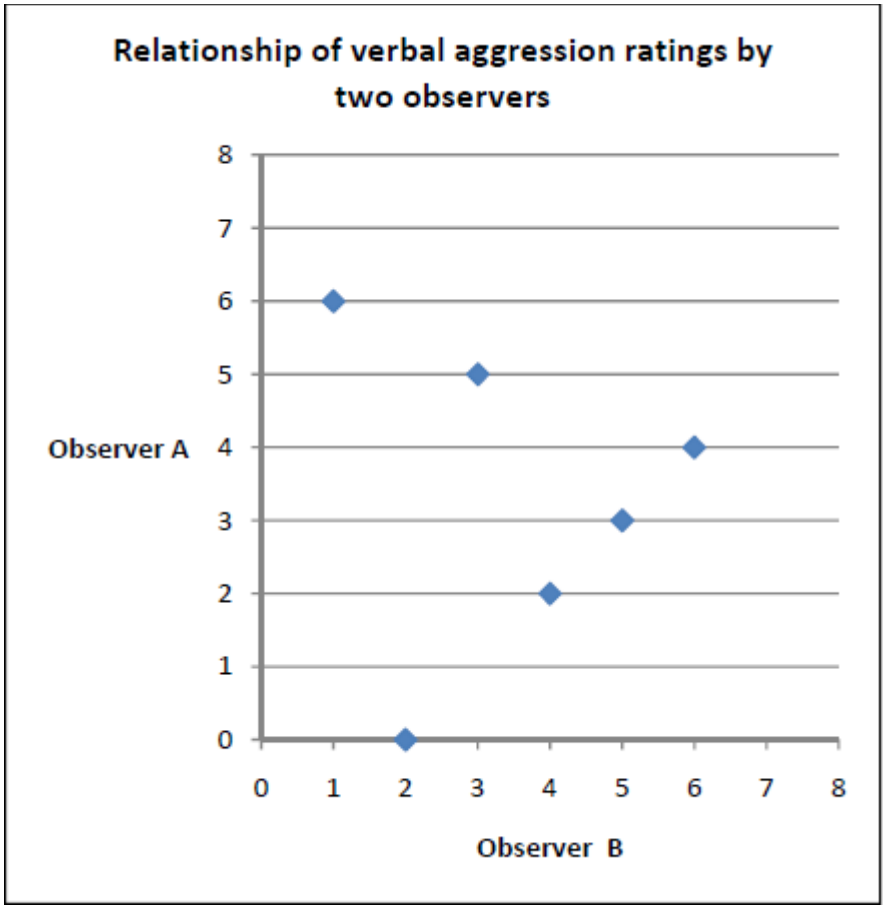
For any credit, candidates must sketch a scattergram.
For full marks, candidates should provide an appropriate title for the scattergram, label each of the axes appropriately and plot the data accurately on the scattergram.

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**

STEMUP.CO.UK

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.

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.

Q33.

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.

AO3 = 2

This graph shows a fairly strong negative correlation between stress and white blood cell activity / the immune system. As the stress increases the immune functioning decreases. The following can all receive a mark: direction, strength, and a description of their relationship.

Q34.

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.

STEMUP.CO.UK

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.

AO3 = 4

Conclusions can include: there are two factors that influence conformity, the ambiguity of the task and the size of the majority. A large majority is most influential with an ambiguous task, but still exerts pressure even when the task is easy. However, a small majority has less effect and the type of task does not seem to be an important variable.

4 marks Accurate and reasonably detailed Accurate and reasonably detailed answer that demonstrates sound knowledge and understanding of what the bar chart shows about conformity. There is appropriate selection of material to address the question.
3 marks Less detail but generally accurate Less detailed but generally accurate answer that demonstrates knowledge and understanding. There is some evidence of material to address the question.
2 marks Basic Basic answer that demonstrates some relevant knowledge and understanding but lacks detail and may be muddled. There is little evidence of selection of material to address the question.
1 mark Very brief/flawed of inappropriate Very brief or flawed answer demonstrating very little knowledge. Selection and presentation of information is largely or wholly inappropriate.
0 marks No creditworthy material.

Q35.

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

STEMUP.CO.UK

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 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
- 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.

STEMUP.CO.UK

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.

Q10.

This was a data response question, asking students to interpret the data in the table. The focus here must be on these results and not on prior knowledge. Answers that went beyond the data and discussed the findings of Bickman's study, or Milgram's findings did not gain credit. Better answers included comment on both what the confederate was wearing (smart versus casual) and the task involved (easy versus difficult). Some students appeared to think that the terms obedience and conformity are interchangeable.

Q11.

This was the first time on this paper that students had been required to sketch a graph and it was done very well indeed, in spite of the fact that many did not have a ruler or pencil. The main weakness was in correctly labelling the y-axis.

Q12.

(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

STEMUP.CO.UK

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.

(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.

Q13.

(a) This was generally answered well although a surprising number of students failed to include both elements of ‘order’ and ‘middle’ in their answers. Some students just presented the partial formula of n(+1) / 2 with no explanation of what this meant.

(b) Students failed to notice that the DV was an ‘estimation’ not the median score.

(c) Students are still poor at writing clear hypotheses with both conditions of the IV present and a measurable DV. Many answers were aims and the expression ‘more likely’ was often used.

(d) Students were still confused about the term ‘experimental design’, often thinking it referred to conducting a field or laboratory experiment. Others did not realise that in both the first and second parts of the investigation the researcher was looking for a difference between the performances of men and women meaning that on both occasions the experimental design was independent groups.

(e) Overall, there was a lack of clarity in some responses to this question. Unless students made it clear exactly how the researcher would end up with a sample of 15 men and 15 women from the factory, they could not gain full credit. References to random number generators / computers often did not describe how the names would become numbers.

(f) This was answered well with the majority of students gaining at least one mark.

STEMUP.CO.UK

(g) This was answered reasonably well with many students gaining at least 2 marks.

(h) Unfortunately, once again, students did not recognise that if asked about debriefing participants then the focus is on telling people everything about the study in which they have just taken part. Too many answers merely concentrated on ethical points or added information which would not normally be available at debriefing, such as, details of all the results. Also, some accounts stated that participants could withdraw from the study, even though they had finished participating.

(i) Students often limited themselves to one mark because they failed to develop the reason they proposed as an explanation for why researchers conduct pilot studies. In general, it is accepted that the purpose of a pilot study is to identify possible flaws so that they can be eliminated / to ensure the data collected is appropriate / to ensure time is not wasted.

Q14.

Some students failed to read the stem carefully and found it difficult to answer the question. It was insufficient to identify a technique as 'mental re-instatement'. Students were required to suggest the context was being mentally re-instated. Where students chose to write about mental re-instatement of the context, they needed to remember that they should refer to the context of watching a film of a violent crime. Many focussed on creating mental context within the film rather than of watching the film. Students who chose 're-instate the context' sometimes muddled their instructions to give instructions to 'recall everything'. Generally those who chose to focus on 'recall from a changed perspective', or 'recall in reverse order' found it easier to write instructions for the participants. A few students did not write instructions in direct speech and were limited to a maximum of one mark for the instructions.

Where students did not score full marks for this question, they had usually failed to refer to all of the figures in the table, or to draw any conclusion from the figures.

Q15.

Most students were able to state correctly that the correlation was a positive one; as scores on a stress questionnaire increased, so did the number of days off work through illness. Credit was also given to any reference to strength (moderate or strong) as well as to the flattening out of the graph.

While most students were able to consider the main weakness of correlations, few could correctly outline a strength. In fact many contradicted themselves by saying that a strength was the ability to show cause and effect on a graph and the weakness was that a causal relationship can never be assumed. Better answers referred to the strength as being the ability to study the relationship between variables that occur naturally; or to measure things that cannot be manipulated experimentally.

Q16.

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

STEMUP.CO.UK

(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.

Q17.

Better students were able to extract the relevant information from the table and use it effectively. They considered the baseline of 65% (no confederates) and then compared it to the other two conditions, 92.5% and 10%, which showed the power of confederates. They were also able to comment that in fact the disobedient confederates seemed to have more power than the obedient ones, perhaps by providing role models or allies.

Since this question only asked about the confederates, reference to the third condition (experimenter in different room) was not credit-worthy. This illustrates the need for students to read the question carefully and select and shape their answer accordingly.

A significant number of students confused conformity and obedience and used these terms interchangeably. They seemed to forget that this data referred to Milgram's experiment into obedience and seemed to think that the confederates were a majority.

Q18.

(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.

STEMUP.CO.UK

(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.

STEMUP.CO.UK

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.

(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.

Q19.

This was a straightforward question for which many students gained at least one of the two marks for referring to numerical data. However, not all students provided a suitably clear example of quantitative data collected by a psychologist who studied conditioning. Those who managed to gain full marks often did so with, for example, reference to the work of Pavlov and Skinner, by referring to the amount of saliva or time taken to respond.

Q20.

Almost all students gained at least one mark. Those who gained full credit did so by appropriately referring to the nomothetic / scientific nature of the behaviourist approach and disadvantages of qualitative data. Some students gave good and relevant justifying commentary for comparisons with quantitative data collection. Students who only achieved one mark, generally did so because they gave two vague or very brief reasons or a generic evaluation of qualitative data not linked to investigations carried out by behaviourists. A minority of students gave one reason only.

Q21.

(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.

STEMUP.CO.UK

(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 were able to write confidently and in an informed manner.

Q22.

(a) Attention was drawn to the requirement for the variables to be operationalised in this answer. Some responses met this requirement effectively, other responses were more vague. Inevitably there were some who mixed up the IV and DV.

(b) Although almost all students recognised that the mean score for aggression was higher in those who started day care before the age of two than after the age of two, fewer students pointed to the magnitude of the difference being small.

(c) There was a wide range of incorrect answers to this question. Clearly a number of students did not recognise the term “dispersion”.

(d) Most students were able to draw a bar chart. Those who drew separated bars to represent those who started day care before or after the age of two created a better visual impact than those who chose to join the two bars. Students who did not score full marks usually failed to label the axes fully eg labelled the ‘y’ axis as mean score, rather than mean aggression score.

(e) In a majority of responses, hypotheses were both appropriate and directional. Some students failed to operationalise part of the hypothesis and so did not score full marks. A few responses were written in the form of a correlational hypothesis, which was not appropriate.

Q23.

Students seem to struggle with the interpretation of graphs and many do not seem to understand the difference between what data show ie findings and what can be concluded from the data. The graph shows a negative correlation, which is moderate-strong; this shows that the more weeks of treatment the lower the depression score. Students could also comment on the apparent “plateau” or change in direction in the middle of treatment.

Very few students considered the strength of the relationship and weaker students just read the data off the graph.

Q24.

(a) Students found this question very difficult. Many of the answers given were untestable statements because they did not contain two conditions and / or the DV was not operational. Answers were often in the following format – *There will be a better performance in the audience condition.*

(b) The majority of students were able to describe the pattern of the results but failed to use this as an explanation for the conclusion that could be drawn – the presence of an audience affected / increased the number of accurate shots through the hoop.

(c) Titles for the graph were generally good. The y axis sometimes did not contain reference to 'mean' number of successful shots. Plotting was accurate. Many students were able to present clear and accurate graphical displays.

(d) Students needed to be very clear about the strength and then explain why that was a strength of random sampling. Many students merely wrote a definition of random sampling.

(e) (i) Although students were able to identify a limitation of using independent groups – such as there are participant variables, rather than explaining why this is a limitation, they then went on to describe what is meant by participant variables.

(ii) Answers to this question often lacked information about overcoming the limitation.
Students could suggest that using matched pairs or repeated measures might be appropriate, but could not expand on what a researcher would actually do.

(f) (i) This was not well answered. There were many very vague definitions of extraneous variables such as, something that cannot be controlled / something that has to be controlled.

(ii) This was quite well answered although only a few students referred to 'confounding' in their answer.

(g) This question required students to imagine themselves as the researcher about to read out to participants in the 'audience' condition the instructions for participation. Some students produced very clear information in verbatim form. There were some students who focussed only on ethical issues.

Q25.

(a) Many candidates produced an operationalised directional hypothesis. Sometimes the dependent variable in the hypothesis was not operationalised, resulting in less than full marks.

(b) Most responses were appropriate with candidates showing a good understanding of independent groups design.

(c) Strength: many responses focused appropriately on the absence of different types of order effects or the limitation of demand characteristics relative to a repeated measures design. Answers which elaborated a relevant strength gained full marks.

STEMUP.CO.UK

Limitation: Good answers often referred to examples of individual differences between the groups affecting reliability. Appropriate elaboration was often evident.

Lack of clarity of expression sometimes obscured the point being made.

(d) A substantial number of candidates were able to explain the purpose of a pilot study, usually focusing on usefulness for checking and amending aspects of the procedure before the main study. The opportunity to observe and address unexpected ethical issues was also a creditworthy point. Candidates were particularly successful when they used an example to illustrate their answer for example to check the clarity of the pictures. Answers which merely outlined what a pilot study is were not credited.

(e) Responses to this question divided clearly between those who understood and could apply the concept of standard deviation and those who did not understand the concept and hence could not interpret the numbers in the table. A common wrong answer was to claim that the memory improvement group did 'better', presumably because 2.8 is higher than 0.29.

Q26.

(b) Candidates were required to identify and elaborate an appropriate methodology for measuring aggression. Many were able to do this.

Unethical answers such as 'provoke the children and observe their response', did not gain credit.

(c) By identifying the lack of causal inference in correlation and suggesting other relevant factors which may affect children's aggression, some candidates obtained full marks. Another appropriate route was to provide evidence from research to create a counterclaim. Trying to answer by using the scattergram was not a successful strategy as the question asks about causal inference and graph shows only the relationship between factors.

Q27.

It was encouraging to see that this cohort of candidates was able to go beyond simply describing the findings. They were able to make suggestions about what they showed. For example, that both therapies showed some improvement, as there were no scores of zero; that in fact neither showed much improvement as the average was only 6.

However, it was also clear from the responses that a minority of candidates had no real understanding of what range tells us about data.

Q28.

(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

STEMUP.CO.UK

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

Q29.

(a) Most candidates accurately reported independent groups / measures, unrelated or between subjects / groups design. Candidates seemed better prepared to identify the experimental design than in previous examination series, although there were some who incorrectly identified the type of experiment (eg laboratory experiment).

(b) If candidates got (a) wrong (laboratory experiment) then their limitation was usually wrong, so there were some incorrect references to low ecological validity. When candidates selected a correct limitation eg participant variables they did not always elaborate it well enough for two marks. Examples of individual differences were most plausible when they focussed on an appropriate characteristic for this study eg memory differences.

(c) Candidates who looked at the results on the graph and commented appropriately on the pattern of correct and incorrect statements, scored full marks. It is advisable for candidates to read the labels on the axes of a graph. A number of candidates failed to do this and so reported inaccurately on the average number of *participants* who had made statements.

(d) Answers generally showed good understanding of the cognitive interview approach. Candidates who just named a process, eg *report everything*, did not score full marks. Some elaboration or application to the stem was necessary to obtain a second mark.

(e) Responses to the term *investigator effects* were varied. A few candidates wrote excellent responses identifying cues the investigator might produce in the context of the study, which might lead to the investigator's expectations being fulfilled. Quite a number of candidates inappropriately discussed leading questions and often cited Loftus's work.

Q30.

Drawing conclusions from data seems to be a topic that candidates find hard. Perhaps teachers might emphasise the difference between a conclusion and a finding (although findings can be used to support a conclusion). Also they need to highlight to candidates

STEMUP.CO.UK

that these types of questions are straightforward and are not trying to trick them. Often very able candidates seemed to think that they had to go beyond the data and make links to theory etc.

Q31.

(a) It was disappointing to see such poor understanding of how to write a non-directional hypothesis for an experiment. Common pitfalls were to write a directional hypothesis, a null hypothesis or a correlational hypothesis. Some candidates did not appear to understand the conventions of hypothesis writing at all and simply wrote a statement such as: 'They wanted to test memory span' or 'The capacity of STM is 7+ or – 2'.

(b) Some candidates clearly understood why the researchers had chosen an independent groups design and were able to offer a full answer worth 2 marks. However, this question was not answered well by many candidates. Candidates needed to apply their knowledge to answer this question but, unfortunately, many candidates simply gave a rote–learned advantage of an independent design without thinking about the context of this particular experiment where the problem of order effects or demand characteristics did not apply.

(c) Another straightforward question generally answered correctly.

(d) There were some full-mark answers where candidates demonstrated good understanding and were able to comment on several aspects of the graph. However, many candidates were only able to make a single point ie that adults generally had better recall than children and then 'padded out' their answer with repetition of the same point. A common pitfall was for candidates to state that 6 was the maximum number of digits recalled by children (or 7 by adults) instead of realising that this was the modal score. This was surprising as they had often successfully identified the modes in (c). Few candidates seemed to understand how to measure the range by looking at the information on the graph. It appears that candidates could benefit from more practical experience of interpreting graphs

(e) Most candidates answered this well and were able to refer to earlier research by Miller or Jacobs which had demonstrated the limited capacity of STM (ie 7+/– 2 items or a range between 5 and 9 items). A minority confused capacity with duration and tried to answer in terms of the Peterson and Peterson study.

Q32.

(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

STEMUP.CO.UK

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’.

Q33.

Most candidates offered accurate and detailed answers for this question.

Q34.

This question required candidates to draw conclusions from data, but many could not help going beyond the graph and interpreting the findings in terms of conformity research (Asch, NSI, ISI etc). The question did not require an explanation, but only what the chart showed. Many candidates did not gain many marks because they adopted the wrong approach.

Q35.

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 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

STEMUP.CO.UK

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.

STEMUP.CO.UK

Q1.

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**

Identification of the mean or median.
There are no repeated scores in either list, so the mode would not be appropriate.
Justification for the mean could be that it used all of the available data, that it is a powerful / sensitive measure or that it is suitable for use with interval / ratio data.
Justification for using the median is that it is relatively unaffected by outlying scores.

(b) **AO3 = 2**

The specification names random, opportunity and volunteer sampling. Answers must relate to volunteer sampling which involves participants selecting themselves.

1 mark:
very brief suggestion, eg put up a notice.
2 marks:
some elaboration that could apply to the scenario, eg advertise on the staff room notice board, asking teachers to sign a list.

(c) **AO3 = 1**

Extraneous variables are anything other than the independent variable that could affect the dependent variable. In this study they could include participant differences, environmental variables such as temperature or noise, and experimenter variables including the way in which the research is conducted.

(d) **AO3 = 2**

The control must relate to the variable in (c).
1 mark for a brief suggestion. Second mark for some elaboration.
For example, control for participant differences by making sure both groups are similar (1 mark).
Control for participant differences by randomly assigning the participants to conditions (2 marks).

Q2.

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 = 1**

STEMUP.CO.UK

Identification of a positive correlation, +, +ve.

(b) **AO3 = 4**

The strength and weakness must be appropriate for investigating day care.

1 mark for stating a relevant strength / weakness. A further mark for some elaboration which would apply to day care.

For example, a strength would be it is an ethical way of collecting data (1 mark) because there is no manipulation of time the child spends in day care (2 marks).
A weakness you can't infer cause and effect (1 mark) because you can't be sure that time in day care causes the child to be disobedient (2 marks).

Q3.

(a) **AO3 = 3**

Candidates may point out that the % of secure attachment in all three countries is very similar, but that insecure attachments vary. Country one has the lowest % of insecure-avoidant but the highest of insecure resistant. Country three has the lowest % of insecure-resistant but the highest of insecure-avoidant.

One mark for a brief outline of one point. Two further marks for accurate elaboration of one point in detail or more than one point more briefly.

(b) **AO3 = 3**

Candidates may refer to limitations of the strange situation as a way of identifying attachment type. Alternatively they may focus on the differences in number of studies in each country.
They could also criticise the use of meta analysis.
One mark for a brief outline of a relevant criticism. Two further marks for elaboration.
For example, the findings for country two come from 18 different studies. We can't be sure that all of the studies were carried out in the same way (second mark). It is possible that the 'Strange Situation' or ways of categorising types of attachment were different in the different studies (third mark).
If candidates give more than one criticism, the best should be credited.

STEMUP.CO.UK

Q1.

(a) Some candidates failed to recognise the term ‘central tendency’. Others identified a correct measure but struggled to identify why the measure they had chosen should be used. A number of candidates explained how to calculate the measure or simply defined it, which could not receive credit. Candidates who had been well prepared for this question gave clear and concise answers such as “The median, because it is unaffected by outliers.”

(b) Sampling method was generally well understood and many candidates were able to apply the method to teachers in the school. Candidates who had not read the stimulus material carefully enough suggested ways to obtain a volunteer sample which would not have been appropriate for this specific example. Others confused volunteer with opportunity or random sampling.

(c)(d) A number of candidates had not read the study closely enough and selected possible extraneous variables to do with the word lists. Good answers often cited individual differences or environmental variables in question (c) and were then able to explain how to control this in question (d), eg matched pairs or keeping noise, temperature constant / controlled.

Q2.

(a) Most candidates were able to identify this as a positive correlation. There were very few incorrect answers.

(b) Candidates found it harder to identify a strength than a weakness. Good answers often explained the advantage when it would be unethical to manipulate variables. For the weakness most candidates focused on causality issues.

Q3.

(a) This was a straightforward question and many candidates gained full marks. Some candidates inappropriately used their knowledge of cross-cultural research to write about different parenting styles in various countries instead of simply reading the table as the question required. It would be worth pointing out to candidates that when a question says, “Outline what the table shows”, that is exactly what is required.

(b) Most candidates were able to identify a criticism although they often failed to explain the criticism for full marks. Some candidates ignored the requirement to ‘explain one criticism’ and, instead, identified two or three criticisms without elaborating any of them. Other common errors were explaining a criticism of the strange situation with no reference to cultural variations, or explaining a criticism of investigations of cultural variation with no reference to the strange situation.

STEMUP.CO.UK