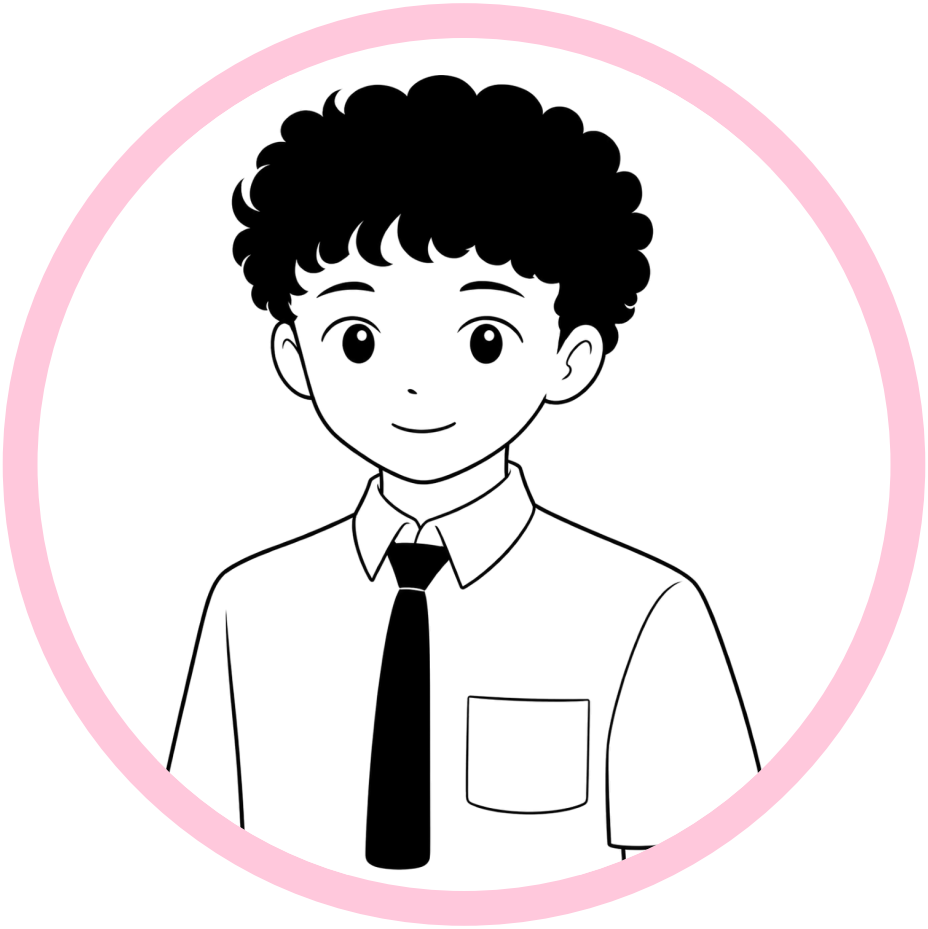
AQA A-LEVEL PSYCHOLOGY · 7182

PSYCHOLOGY



MARK SCHEME

3.2.3.1 VALIDITY

SCORE	PERCENTAGE
/171	%

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 = 6]**

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

Possible content:

Outline of what the experiment entails:

- an explanation of the general purpose of the research
- they would need to watch an assigned horror film every night before going to bed for 7 nights and would need to watch an assigned romantic comedy every night before going to bed for 7 nights
- a daily requirement to truthfully respond to a text message asking whether they had experienced a nightmare
- the two-week duration of the experiment.

Ethical guidelines:

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

Format and style of consent form:

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

Credit any other relevant information.

6

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

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

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

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

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

Possible content:

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

3

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

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

1 mark for a limited or muddled explanation.

Possible explanations:

- 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.

STEMUP.CO.UK

Accept alternative wording.

2

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

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

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

OR

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

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

Note: Credit alternative methods of randomly splitting the sample

3

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

Mean:

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

Plus

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

Standard deviation:

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

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.

STEMUP.CO.UK

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

Accept any other valid answer.

2

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

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

Plus

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

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

Possible content:

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

Accept alternative answers.

Note: Independent groups design would not be an appropriate modification in this study; however candidates can make the case for a matched pairs design, which could be credit worthy.

3

[25]

Q2.

[AO2 = 3]

3 marks for a clear, coherent and fully elaborated explanation
2 marks for an explanation with some elaboration
1 mark for a limited or muddled explanation

Content
An observation would be more appropriate because a self-report method like a questionnaire would lead to socially desirable answers/lying/self-delusion/misremembering, so would not reflect what really happens at

STEMUP.CO.UK

the gym, so the data would lack validity.
(Accept alternative wording.)

Note: responses which are inappropriate for the scenario are not creditworthy.

Q3.

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

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

STEMUP.CO.UK

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

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

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

Plus

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

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

OR

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

- 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).

Q4.

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

1 mark appears to support the nature side of the debate.

Plus

1 mark because the concordance rate is stronger in the identical twins where there is greater genetic relatedness (or nurture must also play a role – not 100% concordance).

Full credit can be awarded to answers which argue for mathematical ability being partly due to nurture as both percentage concordance rates are less than degree of genetic relatedness.

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

Level	Marks	Description
2	3 – 4	Answer focuses clearly on concurrent validity. How a correlational test would be used to determine the relationship between the two sets of scores is clearly described with reference to calculation of a correlation coefficient and need for a significant positive correlation.
1	1 – 2	Answer focuses on validity. How a correlational test would be used to determine the relationship between the two sets of scores is partly described. The answer lacks accuracy and detail.
	0	No relevant content.

Content:

- concurrent validity would involve correlating the results on the maths test with results for the same group of people on an established maths reasoning test
- A Spearman's rho or Pearson's r test should be used for the two sets of test results
- if the mathematical ability test is valid then there should be a significant positive correlation between the two sets of test scores at the 0.05 level.

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

STEMUP.CO.UK

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.

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]**

STEMUP.CO.UK

1 mark – C Event sampling

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

1 mark for each of the following:

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

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

1 mark for the correct ratio: 3:2

Plus

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

54:36 both divisible by 9 (or 18)

54 ÷ 9 = 6 and 36 ÷ 9 = 4

6:4 can be simplified to 3:2

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

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

1 mark for each of the following:

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

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

1 mark for the correct answer: df = 2

Plus

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

(2 − 1) x (3 − 1) = 1 x 2 = 2

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

1 mark for each of the following:

- Yes, the result is significant

STEMUP.CO.UK

- 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.

- 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:

STEMUP.CO.UK

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.

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

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

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

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

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

One mark for answers either:

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

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

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

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

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

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

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

1 mark for a knowledge of a way (not just naming a type of validity) and 2nd

STEMUP.CO.UK

mark for explaining how this would be implemented in this case. Most likely answers will address face validity or concurrent validity, but accept any other way such as construct validity, content validity, criterion validity and predictive validity.

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

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

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

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

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

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

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

- clearly identified independent and dependent variables and at least one extraneous variable identified and control suggested;
- the experimental design – independent groups, repeated measures or matched pairs;
- detail of sample;
- materials required for carrying out the research, eg task for assessing levels of recall, timing device if needed;
- sufficient procedural details to carry out a replication (might include standard instructions, ethics, etc.)

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

Mark bands

8 – 7 marks	**Very good answers** All 5 points well addressed and some sound justification. Answer shows sound knowledge and understanding and an ability to

STEMUP.CO.UK

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

Q7.

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

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

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

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

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.

STEMUP.CO.UK

- The rating may be invalid as the students selected songs which varied in difficulty so the tunefulness reflected the difficulty of the song not the students' ability.

- Operationalising musical ability as tuneful singing is a very narrow measure. Someone can have musical ability such as playing an instrument which would not be reflected by this measure.

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

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

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

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

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

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

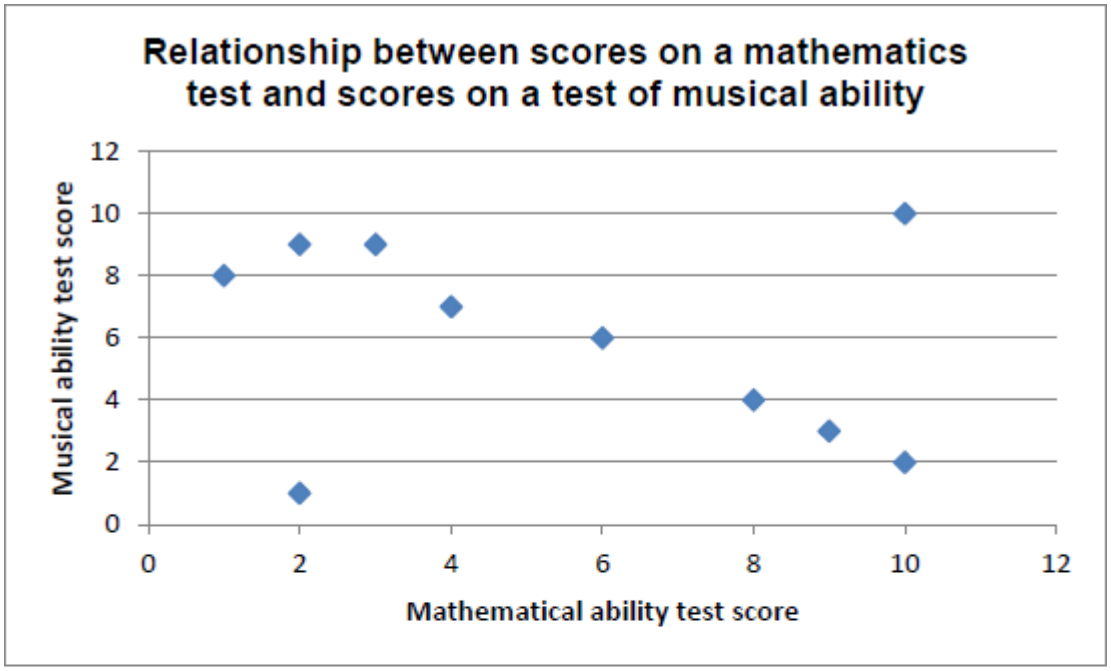
No credit for definition of a random sample.
1 mark for a brief or muddled reason (it is not biased).
2 marks for a reason that clearly points to an advantage of random sampling. This could be achieved through a comparison with another method (it is less likely to be biased than a volunteer sample).

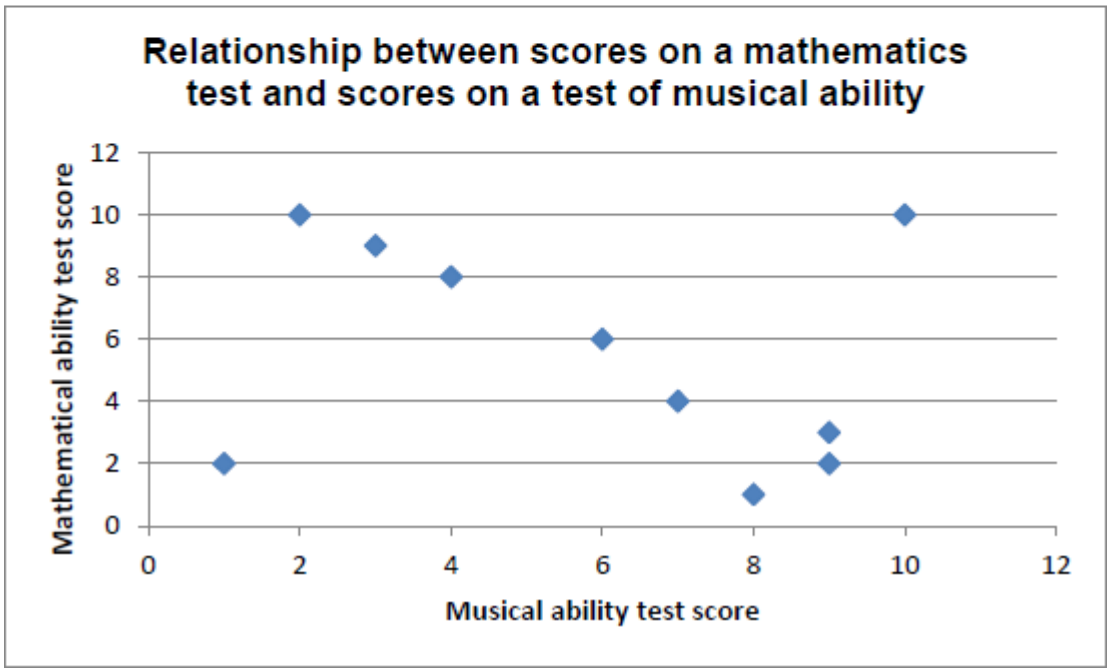
(e) **AO2 / AO3 = 3**

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

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

STEMUP.CO.UK





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

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

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

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

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

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

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

Design – 1 mark

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

Sampling – 2 marks

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

Procedure and assessment of musical ability – 4 marks

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

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

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

Debrief – 3 marks

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

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

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

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

Musical ability scores:

Participant number	Left handed	Right handed
1		

STEMUP.CO.UK

2		
3		

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

Q8.

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

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

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

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

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

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

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

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

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

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

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

STEMUP.CO.UK

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Q9.

STEMUP.CO.UK

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

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

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**

They wanted to clarify some of the issues raised by previous research where some studies had shown that red facilitated tasks and other studies had shown the opposite. They believed that one way to reconcile these different findings was to look at particular cognitive tasks eg ones which required attention to detail and to compare them with tasks which tap into very different skills eg creativity and thus to narrow down the benefits of providing red backgrounds.

One mark for a brief answer eg 'they wanted to investigate the effects of colour on performance.' One further mark for elaboration, in relation to colour and / or performance.

(b) **AO1 = 2**

Candidates need to show understanding of reporting conventions. The introduction is an important part of the report that provides background information on theories and studies relevant to the investigation. One mark for a brief explanation of the purpose eg 'It provides background information', and one further mark for elaboration or for other detail such as reviewing methodological issues or how the current aims / hypothesis were derived.

(c) **AO1 = 1**

In this question, candidates are not required to relate validity to this particular study so a general definition of validity is acceptable. Definitions of specific types of validity (eg population validity) can also gain credit.
Validity refers to how well a test or a piece of research measures what it says it measures = 1 mark.
Answers such as 'truth' or 'whether it is true' 'legitimacy' or 'accuracy' = 0 marks.

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

The Canadian researchers who actually undertook this study suggested the following possible practical applications:

- to help decide what colour to pick for an educational facility
- to help decide what colour enhances persuasion in a consumption context
- to help decided what colour enhances creativity in a new product design process.

STEMUP.CO.UK

Any plausible practical applications are credit-worthy.
1 mark for identifying an application and 1 further mark for elaboration.
'You could use particular colours for pages in textbooks' = 1 mark.
'Red might be used in textbooks covering analytical subjects like maths' = 2 marks.

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

If the researchers had judged the toys themselves, they might have been biased in favour of their hypothesis. There are no objective criteria for what makes a toy either practical or original.
Independent judges would be able to decide between themselves on a set of criteria and then apply them to the toys made by the participants.

Some candidates might interpret 'independent judges' in this question to mean judges who do not confer with one another. In this case, an acceptable answer would be that they could not conform with one another when making their judgement.

One mark for a brief explanation, eg to avoid experimenter bias, and one further mark for elaboration, eg if the researchers judged the toys themselves.
An answer explaining the value of rating the toys should be credited.

(f) **AO2 / AO3 = 5**

Candidates need to use the details in the description of the study to write an appropriate set of instructions for potential participants.

The instructions should be clear and succinct. They must:

- explain the procedures of this study relevant to participants
- include a check of understanding of instructions.

They should also use language appropriate for a formal document and be as straightforward and courteous as possible.

This is not a consent form so explicit references to ethical considerations are not necessary for full marks. However, it is perfectly acceptable to include comments such as 'you are free to withdraw from the study at any time.'

AO3 Mark Bands
5 marks Effective The **standardised instructions** provide accurate detail of the procedure and go beyond the information given in the question eg provide details of time allowed.
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

STEMUP.CO.UK

procedure.
0 marks No creditworthy material is presented.

Q10.

AO3 = 3

Candidates are likely to refer to the fact that in real life settings research has high validity because the findings can be generalised to other similar situations. It is therefore more likely to be relevant eg to eyewitness testimony in court cases. There are often real consequences / emotional impact in real life which do not occur in laboratory investigations.
In a laboratory participants may show demand characteristics because they know they are in an experiment. This is less likely in real world settings.
Answers which refer to advantages of laboratory research or disadvantages of real world research are not relevant and should not receive credit.
1 mark for a brief explanation eg higher ecological validity.
Further marks for some elaboration as above.

STEMUP.CO.UK

Q1.

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.

Q2.

This was a well-answered question with almost 90% of students achieving two or three marks. Most students knew disadvantages of questionnaires but lacked an understanding of observations, often stating that they were better as the *"researcher could see for themselves"*. Weaker students merely paraphrased the stem, stating *"people's perception of how they spend their time in the gym is often not very accurate"*.

Q6.

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

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

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

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

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

of variables in such a study.

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

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

Q7.

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

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

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

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

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

(e) This question required students to draw a scatter graph to display the data. About

STEMUP.CO.UK

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

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

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

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

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

(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.

Q8.

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

(b) Many students were able to make correct use of the table and draw an appropriate conclusion about the statistical significance of the T value. Stronger students were able to present this information well, with some even correctly stating that the results

were not significant at the 0.02 level and explaining why. A few less successful students were confused about the critical and calculated values of T.

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

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

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

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

Q9.

(a) This question was answered well. Most candidates provided a detailed aim that was awarded 2 marks. A minority of candidates provided a more general aim that was credited with just 1 mark, such as "to investigate the effect of colour on performance in cognitive tasks". For both marks some elaboration (either related to colour or performance) was required.

(b) This question was problematic for a lot of candidates. Many candidates confused the introduction with the abstract or the method sections and received no credit. Some recognised the inclusion of aims / hypotheses in the introduction but did not achieve 2 marks because they did not make reference to background information. A few impressive answers showed real understanding and referred to "contextualising the research".

(c) Although this question was worth only 1 mark, many candidates produced lengthy answers.

Some distinguished between specific types of validity such as external validity or population validity. A small number of candidates became confused between validity and reliability and provided a definition of the test-re-test method. Just over half of candidates gained the mark.

(d) There were some lovely, imaginative responses to this question which was answered well in general. The majority of candidates achieved 2 marks by including an example of how colour could be used in a real world setting. The most popular answer was use of colour in classroom walls or on textbook pages to aid learning in particular subjects.

(e) The majority of answers to this question demonstrated an understanding that independent judges were required to reduce bias, and in doing so the majority achieved two marks.

(f) Few students achieved full marks on this question, providing little additional information to that included in the question stem. Candidates were too focused on providing details of ethics (which was not required), at the expense of standardised instructions. Some candidates also made an error in their instructions by stating that participants would be given 40 shapes, 20 red and 20 blue, when in fact participants would only be given one colour of 20 shapes. A further common error was writing that participants would be given a limited time to make the toy. Writing a limited time is not a clear standardised instruction and stronger candidates wrote exactly what the time limit would be. Very few candidates checked if participants had any questions at the end of the instructions. Candidates who had conducted research were at an advantage here and produced answers of a higher quality.

Q10.

This question was generally answered well. Most candidates focused on the higher ecological validity provided by real-world studies. Weaker answers simply made this point without any elaboration. Better answers explained why a real-world setting would provide more ecological validity – usually in terms of heightened anxiety and / or consequentiality associated with real-life events. The best answers made reference to EWT studies rather than simply stating the general advantages of real-life studies over laboratory studies.

STEMUP.CO.UK