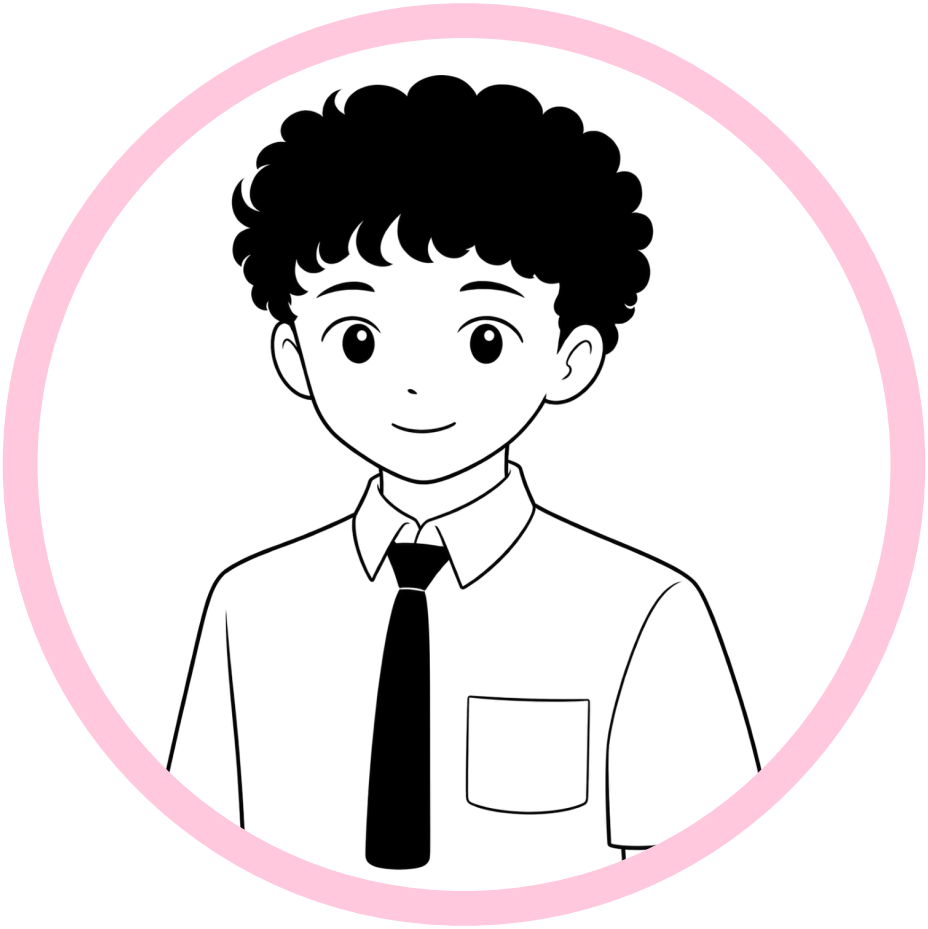
AQA A-LEVEL PSYCHOLOGY · 7182

PSYCHOLOGY



MARK SCHEME

3.2.3 OBSERVATIONAL TECHNIQUES

SCORE	PERCENTAGE
/196	%

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) **[AO1 = 1]**

1 mark for D

1

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

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

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

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

Possible content:

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

Credit other valid points including discussion of ethical issues.

3

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

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

STEMUP.CO.UK

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]

Q3.

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

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

- excitement is a broad construct / complex behaviour / has many aspects
- heart rate is a narrow, biological / physical component / unit / element / factor in overall excitement.

2

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

For **each** outline award marks as follows:

2 marks for a clear and coherent outline with some elaboration.

1 mark for a limited / muddled outline.

Possible content:

- questionnaire measuring attitude – questions assess how the participant feels about going on the roller coaster
- rating scale of own excitement – the participant could rate their perceived level of excitement on a scale of 1 – 10
- observation of non-verbal behaviour whilst waiting in the queue; e.g. tallying appropriate behavioural categories
- interviews before / after the participant has been on the ride asking questions about how they felt / whether they enjoy other exciting occasions / events / how they cope with excitement, etc.

Do not credit measures of biological function.
Just naming a method e.g. questionnaire / interview / observation is not sufficient.

4

[6]

Q4.

STEMUP.CO.UK

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

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

Q5.

[AO2 = 6 AO3 = 6]

Level	Marks	Description
4	10 – 12	Suggestions are generally well detailed and practical, showing sound understanding of observational techniques. **All four** elements are presented appropriately. Justifications are appropriate. The answer is clear and coherent. Specialist terminology is used effectively. Minor detail and/or explanation sometimes lacking.
3	7 – 9	Suggestions are mostly sensible and practical, showing some understanding of observational techniques. **At least three** elements are presented appropriately. There is some appropriate justification. The answer is mostly clear and well organised. Specialist terminology is mostly used effectively.
2	4 – 6	Some suggestions are appropriate but others are impractical or inadequately explained. **At least two** elements are presented appropriately. Justifications are partial or muddled. The answer lacks clarity, accuracy and organisation on occasions.
1	1 – 3	**At least one** element is addressed but knowledge of observational techniques is limited. Justifications are absent. The whole answer lacks clarity, has many inaccuracies and is poorly organised.
	0	No relevant content.

Four elements of design to be credited:

- **Type of observation with justification** – eg covert or overt, naturalistic, participant or nonparticipant and why
- **Operationalised behavioural categories** – detail of at least two specific and observable behaviours to be recorded. This must go beyond the idea of global constructs such as exercising, socialising, use of other facilities

STEMUP.CO.UK

- **Use of time and/or event sampling with justification** – recordings can take place at specified time intervals (time sampling) eg every minute or as the behaviour occurs (event sampling) eg number of times interaction occurs with another gym member. The type(s) of sampling must be appropriate for the behaviours chosen

- **How reliability of the data collection could be assessed**, inter-observer reliability eg using two observers/raters and comparing separate recordings; statistical comparison (correlation) of data from both observers/raters, intra-observer reliability eg checking video recordings.

Note: test-retest is not creditworthy but repeated observation of a video recording is creditworthy.

Q6.

[AO3 = 4]

Award **1 mark** for any four points explained from the following points, to a maximum of 4 marks:

- behavioural categories allow observers to tally observations into pre-arranged groupings
- examples of behavioural categories appropriate in this situation might be ……
- using categories provides clear focus for the researcher
- categorisation enables proposal of a testable hypothesis
- categories allow for more objective / scientific data recording
- use of categories should result in greater reliability
- categories provide data that is easier to quantify / analyse
- contrast with method described in the stem (own interpretation is too subjective / opinion-based).

Credit other valid points.

Q7.

(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.

Q8.

STEMUP.CO.UK

[AO1 = 2]

1 mark for stating that overt observation is where the observer is clearly visible (not hidden from view).

Plus

1 mark for explanation – people being observed know that they are being observed.

Q9.

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

1 mark C

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

1 mark 4/5^{ths}

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

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

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

1 mark for a vague/muddled conclusion.

Conclusion and explanation:

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

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

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

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

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

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

STEMUP.CO.UK

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.

Q10.

[AO1 = 2]

1 mark each for any two of the following points:

- observers / researchers decide on a specific event relevant to the investigation
- relevant event is recorded every time it happens
- in this investigation this may be every time a child in the playground is approached by / talks to / plays with another.

Students may refer to the investigation described in their answer, though this is not required by the question.

Q11.

[AO3 = 4]

Level	Marks	Description
2	3 – 4	Discussion is relevant, well developed and well explained, with focus on improvements to be had by using controlled observation. The answer is generally coherent with effective use of specialist terminology.
1	1 – 2	Discussion is relevant although there is limited explanation / development and / or limited focus on the issue of improvement. Specialist terminology is not always used appropriately. Award one mark for answers consisting of a single point briefly stated or muddled.

STEMUP.CO.UK

	0	No relevant content.

Possible content:

- controlled environment affords the opportunity for control of extraneous variables
- examples of extraneous variables that might be controlled and how / why they could affect the outcome of a study if not controlled
- exclusion of extraneous variables allows for greater inference about cause and effect
- exclusion of extraneous variables means researcher can replicate the observation to check for reliability of the effect.

Credit other relevant discussion points.

Q12.

[AO2 = 12]

Level	Marks	Description
4	10 – 12	Suggestions are generally well detailed and practical, showing sound understanding of observational techniques. **All four** elements are present. There is sufficient information for most aspects of the study to be implemented with success. The answer is clear and coherent. Specialist terminology is used effectively. Minor detail and / or explanation sometimes lacking.
3	7 – 9	Suggestions are mostly sensible and practical, showing some understanding of observational techniques. **At least three** elements are present. Implementation of some aspects is possible. The answer is mostly clear and well organised. Specialist terminology is mostly used effectively.
2	4 – 6	Some suggestions are appropriate but others are impractical or inadequately explained. **At least two** elements are addressed. Implementation would be difficult based on the information given. The answer lacks clarity, accuracy and organisation on occasions.
1	1 – 3	**At least one** element is addressed but knowledge of observational techniques is limited. Implementation would be very difficult. The whole answer lacks clarity, has many inaccuracies and is poorly organised.
	0	No relevant content.

Four elements of design to be credited:

- **The task for the participants** – detail of what the men and women in the study will have to do. This must go beyond ‘give a presentation to an audience’.
- **The behavioural categories to be used and how the data will be recorded**

STEMUP.CO.UK

– detail of specific and observable behaviours to be recorded. This must go beyond the idea of global constructs such as ‘body language’ or ‘gesture’. Also detail of recording method to be used, eg record sheet.

- **How reliability of the data collection might be established**, eg using two observers / raters and comparing separate recordings; statistical comparison of data from both observers / raters.
- **Ethical issues to be considered**, eg specific or more general ethical considerations as applied to this study – protection of welfare, confidentiality and deception, respect or integrity.

Examples of possible tasks:

- presentation of findings from a school project
- presentation on ‘My Hobby’
- presentation on ‘My Holiday’.

Examples of suitable non-verbal behaviours include:

- arm movements
- smiling
- speech hesitations
- pointing etc.

Q13.

(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

STEMUP.CO.UK

Plus

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

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

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

Credit other relevant applications.

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

1 mark for each of two types of validity identified

Plus

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

Content:

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

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

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

1 mark – C Event sampling

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

1 mark for each of the following:

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

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

1 mark for the correct ratio: 3:2

Plus

STEMUP.CO.UK

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

54:36 both divisible by 9 (or 18)

54 ÷ 9 = 6 and 36 ÷ 9 = 4

6:4 can be simplified to 3:2

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

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

1 mark for each of the following:

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

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

1 mark for the correct answer: df = 2

Plus

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

(2 − 1) x (3 − 1) = 1 x 2 = 2

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

1 mark for each of the following:

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

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

Level	Marks	Description
2	3 – 4	The paragraph is clear and coherent, showing sound understanding of the concept of levels of significance and effective application to the context. There is effective use of terminology.
1	1 – 2	The paragraph shows some understanding of the concept

		of levels of significance and/or some relevant application to the context. The answer lacks clarity and coherence. Terminology is either absent or inappropriately used.
	0	No relevant content.

Possible content:

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

Credit other relevant material.

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

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

Plus

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

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

Content:

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

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

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

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

Q14.

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

STEMUP.CO.UK

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 = 4**

Suitable behavioural categories for investigating children's aggressive behaviour could be:–
pushing, hitting, biting, punching, swearing, etc.
Maximum 2 marks – 1 for each suitable behaviour category.
Candidates may suggest recording playground behaviour on CCTV for later analysis by ticking a box when a relevant behaviour is shown by the child. Alternatively the researcher could watch each child's behaviour in the playground and tick the box when each behaviour is shown. In this case where the researcher stands and whether the children know they are being observed would be relevant.
1 mark for a very brief or slightly muddled explanation e.g. use a tally chart
2nd mark for accurate elaboration

(b) **AO3 = 4**

There are no ethical issues named in the specification, so any potentially relevant issues in this research should be credited.
Although the psychologist would not be responsible for the behaviour of the children in the playground he might consider his responsibility if he saw that one of the children was being harmed.
Likely ethical issues include informed consent, right to withdraw, confidentiality or respect. Ways of dealing will depend on the issue selected.
There are different routes to achieving 4 marks depending on the ethical issue selected, but for full marks both the ethical issue and how the psychologist could have dealt with it should be clear.

AO3 Knowledge of research methods
4 marks Accurate and reasonably detailed Accurate and reasonably detailed answer that demonstrates sound understanding of one relevant ethical issue and how the psychologist could have dealt with this issue.
3 marks **Less detailed but generally accurate** answer that demonstrates relevant understanding of one relevant ethical issue and how the psychologist could have dealt with this issue. or **Accurate and reasonably detailed answer** that demonstrates sound understanding of one relevant ethical issue or how the psychologist could have dealt with an issue.
2 marks Basic Basic answer that demonstrates some relevant understanding of one relevant ethical issue and / or how the psychologist could have dealt with an ethical issue, but lacks detail and may be muddled.
1 mark Very brief / flawed

STEMUP.CO.UK

Very brief or flawed answer demonstrating very little understanding of a relevant ethical issue and / or how the psychologist could have dealt with an issue.
0 marks No creditworthy material.

(c) **AO3 = 4**

There are different routes to full marks in this question. Candidates explain one advantage in reasonable detail or more advantages in less detail. Advantages of using an interview rather than a questionnaire could include it would allow the interviewer to clarify questions and answers; it might be easier to see if participants were answering honestly because their reactions could be observed; it is easier to collect detailed qualitative data.

AO3 **Knowledge of research methods**
4 marks Accurate and reasonably detailed Accurate and reasonably detailed answer that demonstrates sound understanding of one or more advantages of using interviews rather than questionnaires in this situation.
3 marks Less detailed but generally accurate Less detailed but generally accurate answer that demonstrates relevant understanding of one or more advantages of using interviews rather than questionnaires in this situation.
2 marks Basic Basic answer that demonstrates some relevant understanding of one or more advantages of using interviews rather than questionnaires in this situation, but lacks detail and may be muddled.
1 mark Very brief / flawed Very brief or flawed answer demonstrating very little understanding of one or more advantages of using interviews rather than questionnaires in this situation.
0 marks No creditworthy material.

Q15.

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,

STEMUP.CO.UK

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 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.

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

STEMUP.CO.UK

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.

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.

STEMUP.CO.UK

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 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.

STEMUP.CO.UK

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

Q16.

AO3 = 2

Credit advantages of different types of observation (eg covert) even if this is not explicit in the answer.
The researcher is able to look at the way people really behave. If people are unaware they are being watched they will not be susceptible to demand characteristics.
1 mark for brief reference to an advantage eg it's real behaviour.
2nd mark for some accurate / effective elaboration. Eg it looks at real behaviour rather that what people say they would do.

Q17.

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

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

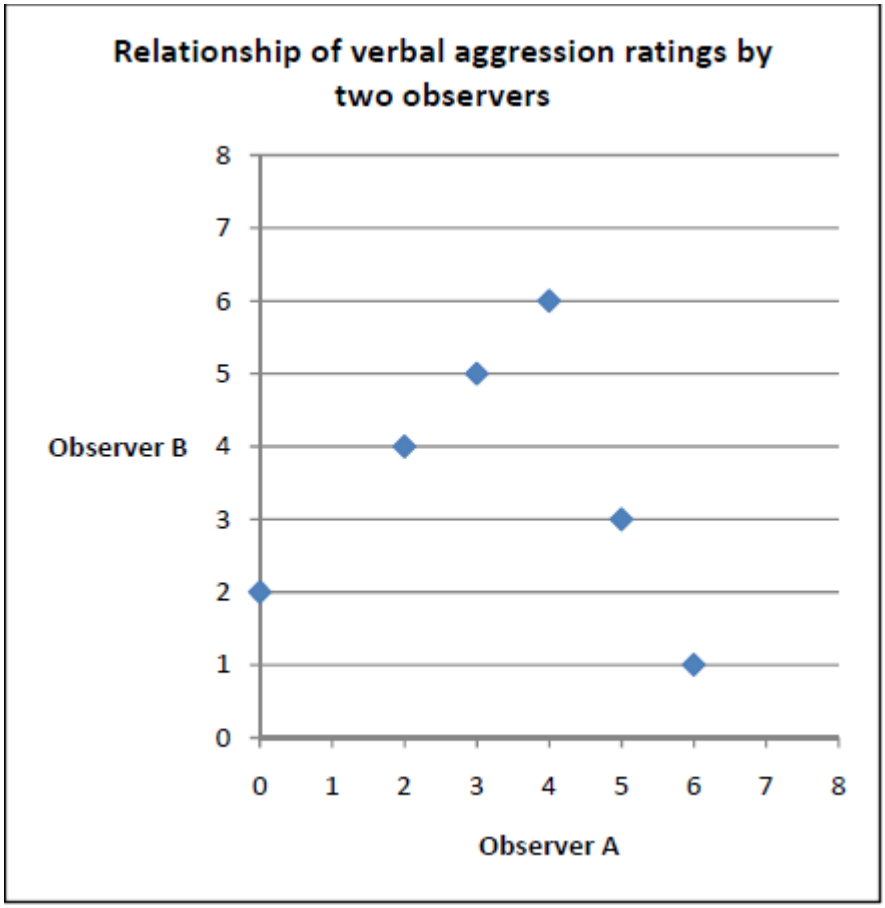
(a) **AO2 / 3 = 4**

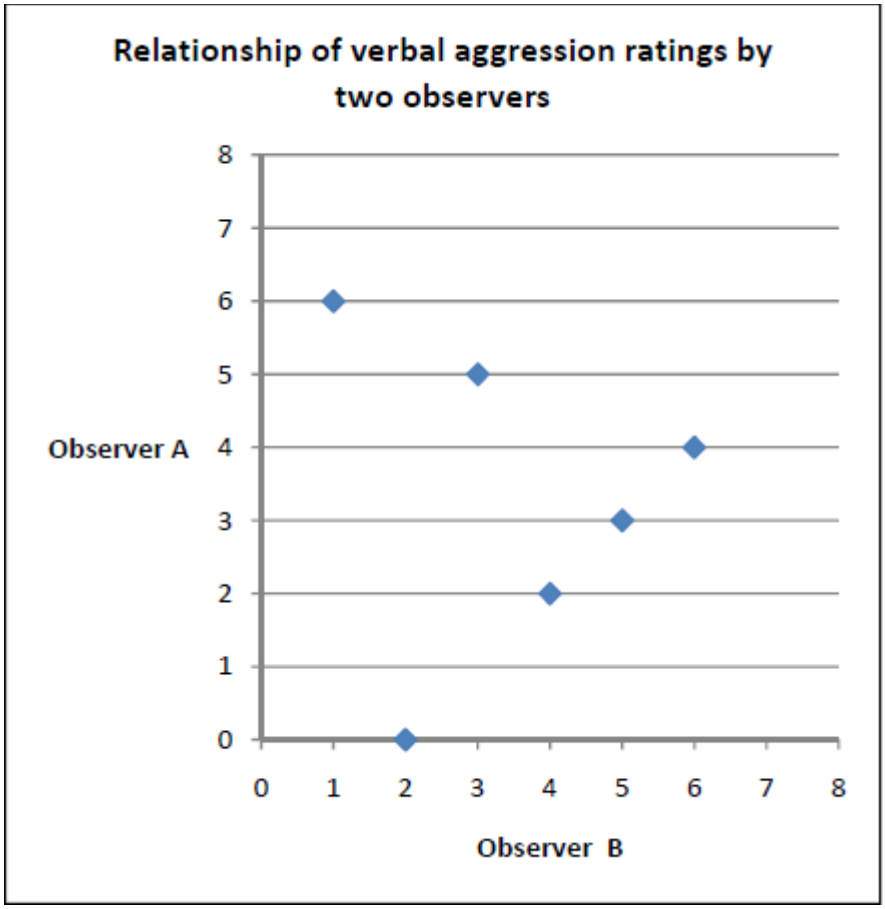
For any credit, candidates must sketch a scattergram.
For full marks, candidates should provide an appropriate title for the scattergram, label each of the axes appropriately and plot the data accurately on the scattergram.

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.

Q18.

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

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

(a) **AO3 = 2**

Behaviour must be operationalised. Suitable behavioural categories could include crying, clinging to mother, smiling, playing independently etc.
One mark for each suitable behavioural category.

(b) **AO3 = 2**

Candidates may refer to time sampling, CCTV and later analysis or ticking a box when the behaviour is shown. Unstructured observation could also be relevant.

One mark for a brief explanation. This could include demonstrating some understanding of the use of behavioural categories, eg draw a table and tick boxes.
A further mark for elaboration, eg drawing the table and / or indicating when the boxes would be ticked.

STEMUP.CO.UK

(c) **AO3 = 2**

One reason for the psychologist carrying out a pilot study would be to check cameras were positioned appropriately. Another would be to check the suitability of the behavioural categories. Alternative relevant reasons should be credited.
One mark for a brief reason eg to check equipment. A further mark for elaboration as above.

STEMUP.CO.UK

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.

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.

Q3.

Performance on part (a) was surprisingly poor because many students focussed on causes of behaviour rather than measurement. Answers had to be contextualised in order to gain credit.

In part (b) most students gained some credit here and there were some excellent straightforward suggestions. High scoring answers often suggested the use of a Likert scale with scale and gave example statements, or observations of behaviour in the queue, with examples of behavioural categories and details of event or time sampling. There were some ingenious suggestions, such as analysing facial expressions from the photographs taken by auto cameras at the top of the roller coaster. Other ideas were just too impractical or implausible to credit. Vague suggestions, such as ‘interview them in the queue and ask how they feel’ gained limited credit, but a few responses were just too brief, for example, simply stating ‘interview’ or ‘questionnaire’. A few biological suggestions were seen.

Q4.

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"*.

Q5.

Students generally did reasonably well on this question. Types of observation were generally understood and justified. Weaker responses sometimes lacked detail/explanations. Behavioural categories were generally poor and operationalisation was rarely evident. Time/event sampling was often stated but for their use was absent or muddled. Confusion was often evident, particularly in relation to time sampling (often mistakenly presented as how long a behaviour lasted). Reliability was often covered well with detailed descriptions of inter-rater reliability given. However, test-retest was frequently suggested inappropriately as a way of assessing reliability and students confused the purpose of assessing reliability with ensuring reliability. Unfortunately, students frequently focussed on additional material, such as ethical issues, sampling of participants, etc, at the expense of covering all the bulleted requirements successfully. Furthermore, some students designed a complete study and a few wrote a whole report, including abstracts, statistical analysis, etc. However, these did not receive credit.

Q14.

(a) Some students clearly understood what behavioural categories are while others had no idea. A number of students were unable to suggest operationalised behaviour categories. Kicking and swearing would be suitable, physical and verbal aggression would be unsuitable. Given the material in the stem, just suggesting children would be observed in the playground, with no further elaboration did not attract credit. A few students did not take account of the stem material and suggested how children could be observed in the home at the age of two.

(b) Most students could identify an ethical issue but some chose issues that were difficult to apply to this scenario. Many ethical issues could be made relevant; the need for informed consent and the need for confidentiality were apt. The effectiveness of the elaboration distinguished between students' marks. Students who chose to write about protection from harm needed to make an appropriate case for this.

(c) Students who responded to this question in the context of the stem material were able to contextualise the comparison between interviews and questionnaires and that helped in producing an appropriate answer. A number of students produced good answers referring to the interviewer's opportunity to clarify questions and answers and the benefits that might accrue from being face to face with the parents. Students needed to explain why parents might be more likely to lie in a questionnaire or in an interview, since a case could be made for either. Some students who did not read the question carefully wrote about the advantages of questionnaires. A few students thought the children, rather than the parents were being interviewed.

Q15.

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

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

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

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

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

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

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

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

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

STEMUP.CO.UK

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

Q16.

There were some very good answers to this question but poor expression in some answers obscured meaning. A number of candidates simply described observational studies rather than offering an advantage.

Q17.

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

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

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

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

Q18.

(a) Many candidates did not understand behavioural categories. Even those candidates who answered the rest of this question well struggled with part (a). Many candidates suggested broad behaviours, eg reunion behaviour, which could not be operationalised.

(b) The requirement to explain how the researcher might record the boy's behaviour was often ignored. Better answers described how behaviour categories could be recorded in a tally chart.

(c) Most candidates appeared to have some idea of why pilot studies are useful. However, a surprising number seemed to think that pilot studies are used to 'give

STEMUP.CO.UK

you the idea for the main study'. Answers sometimes did not contain sufficient explanation to gain the second mark.