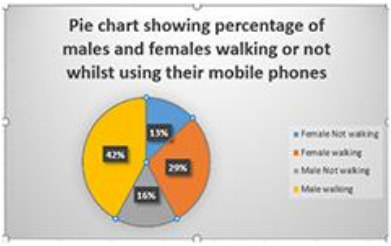
Mark scheme

Question			Answer/Indicative content	Marks	Guidance
1			**Outline one weakness of the method used in Milgram's study into obedience.** As Milgram's study is now generally considered a controlled observation containing elements of the experimental condition, possible weaknesses are likely to apply to lack of external validity, artificiality of task/setting, low construct validity, potential for demand characteristics. <u>Examples of 1-mark answer:</u> The method lacks ecological validity (1). The study was unethical because participants showed signs of extreme stress when asked to administer, what they believed were genuine electric shocks to another person (1). <u>Examples of 2-mark answers:</u> The method lacked ecological validity (1) as sitting in a room giving what are presumed to be genuine electric shocks to another person for getting a question wrong does not reflect a rea-life situation (1). The method lacked external validity (1) because people are rarely expected to electrocute a complete stranger (just because they get a question wrong) as measure of obedience (1). Experiments have low construct validity (1) as the dependent variable – in this case, the voltage to which participants shocked – is a very narrow measure of obedience (1).Other appropriate answer.	2	**2 marks** for a clearly identified and relevant weakness which is appropriately applied to the study i.e. contextualised. **1 mark** for a partial or vague response, not contextualised to Milgram's study; identifying a relevant ethical weakness either explicitly or implicitly through application to the study (as this is not an actual weakness of the method, it is a weakness of how the method was used). **0 marks** – no creditworthy response.
			Total	**2**	

2			**Outline one weakness of the way the sample was organised in Loftus and Palmer's first experiment.** <u>Examples of a 1-mark answer:</u> The way the sample was organised mean the results could lack validity because different participants were used in each condition so individual differences may have affected the results (1). (No context) Different participants were used in each condition so individual differences may have affected results (1). (No context) <u>Examples of 2-mark answers:</u> The results may lack validity because different participants were used in each of the conditions - (smashed, collided, hit, contacted and bumped (1) - so individual differences may have affected the results (1). Different participants were used in the five speed conditions so one may not be comparing participants with the same driving experiences (1) which could have influenced their ability to estimate speed (and therefore the validity of the results) (1).Other appropriate answer – must refer to the way the sample was organised.	2	**2 marks** for a clearly identified and relevant weakness of the way the sample was organised which is appropriately applied to the study i.e. contextualised. **1 mark** for identifying a relevant weakness of how the sample was organised, not contextualised to the study. **0 marks** – no creditworthy response e.g. any references to possible weaknesse
			Total	**2**	
3			Depends on candidates own practical activity **4 mark answer:** Appropriate test named and justified with **two or more** clear reasons in context **3 mark answer:** Appropriate test named and justified with **one** clear reason in context **2 mark answer:** Appropriate test named and justified, but not in context **1 mark answer:** Appropriate test named and weak	Max 4	Context here = the theme of the candidates own chosen practical activity If incorrect test named = zero, regardless of whether any justification is provided or not (and regardless of whether the justification relates to the correct test) Cap at 2 marks if correct test named and reasons given, but one is incorrect

			attempt to justify why (whether in context or not) **OR** correct test named only **0 marks:** No credit worthy information		
			Total	**4**	
4			Strengths could include: easy to analyse data; easy to compare across individuals and conditions Weaknesses could include: don't know reasons why doodling may/may not affect concentration; lack of detail in general Up to 3 marks for each strength / weakness **3 mark answer:** Clear outline of strength/weakness in context **2 mark answer:** Clear outline of strength/weakness but not in context OR Attempted outline of strength/weakness in context **1 mark answer:** Brief and/or weak attempt to outline strength /weakness (whether in context or not) **0 marks:** No credit worthy information	Max 6	- Context – doodling, drawing, scribbling, concentration etc Note: Accept interchanging strengths/weaknesses of qualitative data. For example if a candidate has given a weakness of qualitative data as a strength of now having qualitative data in the study.
			Total	**6**	
5			Answers could include: provides more insight about the reason why doodling did / did not help concentration; can create fixed choice response options that allow ideas why doodling may help concentration or not to be explored more etc **3 mark answer:** Clear outline of strength in context **2 mark answer:** Clear outline of strength but not in context OR	Max 3	- Context – doodling, drawing, scribbling, concentration etc

			Attempted outline of strength in context **1 mark answer:** Brief and/or weak attempt to outline strength (whether in context or not) **0 marks:** No credit worthy information		
			Total	**3**	
6			Nominal as the data is just frequent count of the number of males and females who did or did not use their phone whilst walking **2 mark answer:** Correct data identified and clear justification in context **1 mark answer:** Correct data identified and justified but not in context **OR** Correct data identified and a brief attempt to justify why in context **OR** Correct data identified **0 marks:** No credit worthy information	Max 2	Context = mobile, phone, walk(ing), texting etc
			Total	**2**	
7			Pie chart showing percentage of males and females walking or not whilst using their mobile phones  **1 mark** for correctly calculating what proportion of the circle should represent each of the four behavioural categories. **1 mark** for drawing the sectors in proportional size to the data displayed **1 mark** for clear labelling of each sector of the pie chart **1 mark** is awarded for a clear and appropriate title	Max 4	Context = mobile, phone, walk(ing), texting etc Sectors of the pie chart need only be approximate sizes (examiners do not need to check with protractor or overlay in RM assessor) Calculations can be percentages and/or degrees (decimal places or whole figures) of circle (within labelling of the pie chart, or separately at the side) Females walking 58/200 x 100 = 29% 58/200 x 360 = 104.4 degrees Females not walking 26/200 x 100 = 13% 58/200 x 360 = 46.8 degrees

					Males walking 84/200 x 100 = 42% 84/200 x 360 = 151.2 degrees Males not walking 32/200 x 100 = 16% 32/200 x 360 = 57.6 degrees
			Total	**4**	
8			71% 84 males + 58 females = 142 in total 142/200 * 100 = 71% **3 mark answer:** Correct answer with full workings shown **2 mark answer:** Correct answer with some workings shown **1 mark answer:** Correct answer with no workings shown **0 marks:** No credit worthy information	Max 3	
			Total	**3**	
9	a		Reasons: - test of difference (the study investigated the difference between males/females walking or not whilst using mobile phone) - nominal data collected (frequency count of people using phone or not whilst walking) - independent measures design (males vs females using phone or not whilst walking) **3 mark answer:** Two appropriate reasons provided, both in context **2 mark answer:** Two appropriate reasons provided, but only one, or neither in context **1 mark answer:** One appropriate reason provided but, not in context	Max 3	-Context = mobile, phone, walk(ing), texting etc

			0 marks: No credit worthy information		
	b		The critical value is 3.841 **2 mark answer:** Critical value stated **1 mark answer:** Degrees of freedom correctly identified as 1 but without stating the critical value from the table (or incorrectly stating the critical value) **0 marks:** No credit worthy information	Max 2	If a candidate just circles the correct critical value in the table award 2 marks (as to be able to do this requires knowledge of what df to use) Award 1 mark if df = 1 is just circled on the table (so remember to check table)
	c		$X^2 = 0.2681$, df=1, p>0.05 **OR** in words e.g. this **is** not a significant result (at the 5% probability level) because the calculated value of X^2 had to be equal to or exceed the table critical value (3.841) to be significant, therefore we reject the alternative hypothesis and accept the null hypothesis **2 mark answer:** Correctly written significance statement (calculated value, degrees of freedom and probability level) OR written in words rather than a formal statement **1 mark answer:**Just stating p>0.05 OR weak and/or brief written response **0 marks:** No credit worthy information	Max 2	
			Total	**8**	
10			C	1	
			Total	**1**	
11			B	1	
			Total	**1**	
12			D	1	
			Total	**1**	
13			D	1	

			Total	**1**	
14			C	1	
			Total	**1**	
15			B	1	
			Total	**1**	
16			A	1	
			Total	**1**	
17			B	1	
			Total	**1**	
18			C	1	
			Total	**1**	
19			A	1	
			Total	**1**	
20	a		Possible answers include: pre-existing medical records from GP/hospital/health centre; details from insurance company; details from employer, etc Clear description of how secondary data could be obtained in context Clear description of how secondary data could be obtained but not in context / Attempt at description of how secondary data could be obtained in context Brief and/or weak attempt at description of how secondary data could be obtained (whether in context or not) The candidate has not provided any creditworthy information	Max 3 3 2 1 0	-Context = structure/ weight/BMI, etcDo not credit a definition on its own. However, a definition can add to a correct response. For example a brief/weak attempt could become an attempt by also providing a definition. Secondary data is pre-existing sources that has already been collected but not for the purpose of this research. Can come up with more than one way to collect the data. The data collected could be for the participants in their study (this could be implicit) No credit to the participant/their family weighing themselves. No credit to reference to collecting data from previous psychological studies/articles not involving their participants. **Examiner's Comments** Most candidates did well on this question, clearly outlining how secondary data could be collected. Many candidates referred to accessing GP or hospital records for the participants. Common errors were where a sizeable minority of candidates referred to primary data,

<table>
<tr><td></td><td></td><td></td><td></td><td></td><td>e.g. asking participants to weigh themselves and then give these to these to the researcher. In addition, some candidates suggested that the researcher could access previous research. These candidates were not aware that secondary data must be collected for the participants in the current study and not, different participants from previous research.</td></tr>
<tr><td></td><td>b</td><td></td><td>Likely answers: information already available (more practical); saves time; less personal than asking participants directly, etc
Clear outline of strength in context
Clear outline of strength but not in context | Attempt to outline strength in context
Brief and/or weak attempt to outline strength (whether in context or not)
The candidate has not provided any creditworthy information</td><td>Max 3 3 2 1 0</td><td>-Context = structure/ weight, personality etc

<u>Examiner’s Comments</u>

Many candidates gave well explained responses that were put into the context of the research outlined at the beginning of Section B. Popular responses included saving time, saving embarrassment of being weighed or being more valid as recorded by a professional. If the candidate provided a correct response to 23(a), they were likely to achieve marks for this question as well. However, for those who outlined primary data in 23(a), these candidates often did not achieve marks for 23(b) as the strength they gave was for primary, rather than secondary data. Some of the responses were not contextualised or were very brief. For example, some candidates would state that the data was more valid with a very brief explanation as to why.</td></tr>
<tr><td></td><td></td><td></td><td>Total</td><td>6</td><td></td></tr>
<tr><td>21</td><td></td><td></td><td>Likely answers: the quantitative measurement of ‘extroversion’ does not provide information about why/how the person is like they are, and may have issues related to how the numerical assessment of the variable is implemented/assessed/interpreted, etc.
Clear outline of weakness in context
Clear outline of weakness but not in context | Attempt to outline weakness in context</td><td>Max 3 3 2 1 0</td><td>-Context = structure/ weight, personality etc
Do not credit weaknesses of correlations that do not apply to quantitative data.

<u>Examiner’s Comments</u>

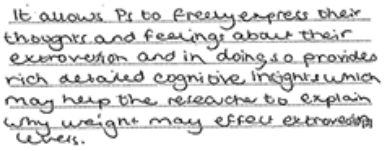
The question was usually answered well by candidates. Many responses outlined weaknesses about not being able to express why someone was extrovert or why they were overweight. Many candidates were able to support their responses with</td></tr>
</table>

			Brief and/or weak attempt to outline weakness (whether in context or not) The candidate has not provided any creditworthy information		good, contextual examples. A few candidates did not put their responses into the context of the research outlined at the start of Section B or gave very brief responses such as stating that quantitative data is not in depth. A common incorrect response was where the candidate evaluated the use of a correlation rather than quantitative data. **AfL** Centres should be encouraged to guide students to look at the marks available for each question. In this instance, for 3 marks merely stating that "quantitative data is not in depth" is not enough detail to gain all the marks available. Candidates should make sure they answer in context when a question includes "in this study".
			Total	**3**	
22	a		Possible examples ... -How would you describe your personality? -In what way do you think weight influences your personality? Open question clearly presented in context Open question clearly presented, but not in context **OR** attempt to present open question in context The candidate has not provided any creditworthy information	Max 2 2 1 0	-Context = structure/ weight, personality etc Allow anything that could affect your weight or personality (e.g. exercise) -Accept open questions related to the assessment of either variable Credit a statement such as 'Describe your personality' (can lead to an open response). **Examiner's Comments** Most candidates wrote an open question that was appropriately contextualised for this study. For example, many candidates asked about why they thought they were an extrovert or why they believed that there is a relationship between extraversion and body type. Some candidates did ask a general question and these received 1 mark as they were not necessarily appropriate for this study in terms of understanding the relationship between extraversion and body type through an open question. A very small minority of

<table>
<tr><td></td><td></td><td></td><td></td><td></td><td>candidates wrote a closed question with response categories and these were not creditworthy.</td></tr>
<tr><td></td><td>b</td><td></td><td>Likely answers: Provides depth/detail/insight; enables the subjective concept of an aspect/type of personality (extroversion) to be investigated, increased validity due to participants being able to express their views/thoughts/feelings, etc
Clear outline of strength in context
Clear outline of strength but not in context | Attempt to outline strength in context
Brief and/or weak attempt to outline strength (whether in context or not)
The candidate has not provided any creditworthy information</td><td>Max 3 3 2 1 0</td><td>-Context = structure/ weight, personality etc

<u>Examiner’s Comments</u>

Candidates performed well on this question. Some of the better responses provided a clear reason as to how having some qualitative data in this study could increase the validity and/or usefulness of the results. Most candidates did achieve some marks on this question. Common reasons for not achieving full marks were that the candidate wrote a very brief response and/or did not contextualised their response in terms of the research study outlined at the beginning of Section B.

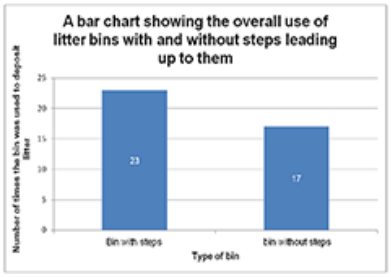
Exemplar 3
 
It allows Ps to freely express their thoughts and feelings about their extroversion and in doing so provides rich detailed cognitive insights which may help the researcher to explain why weight may effect extroversion levels.

Exemplar 3 highlights a good, contextualised outline of a strength that was given full marks. This response also highlights a response that is appropriate for a 3-mark question.</td></tr>
<tr><td></td><td></td><td></td><td>Total</td><td>5</td><td></td></tr>
<tr><td>23</td><td></td><td></td><td>Likely answers: increasing sample size; increasing sample diversity; use of random sampling; improving the ecological validity of the environment of the study; improving the mundane realism of the task, etc
A way the design of the study could increase generalisability of the data collected clearly presented in context
A way the design of the study could increase generalisability of the data collected | Attempt to present a way the design of the study could increase generalisability of</td><td>Max 3 3 2 1 0</td><td>-Context = structure/ weight, personality etc

Generalisability could refer to how representative the sample, the situation and/or task that the participants do.

Credit improvements or generalisability of their original study.

Do not credit definition of generalisability on its own. It must be linked to a feature from the design of the study.</td></tr>
</table>

			clearly presented but not in context / the data collected in context Brief and/or weak attempt to present a way the design of the study could increase generalisability of the data collected (whether in context or not) The candidate has not provided any creditworthy information		**Examiner's Comments** There were some good responses to this question, with most candidates referring to improvements to the sample and how this would help to increase the generalisability of the findings. The vast majority of the responses were contextualised in terms of extraversion and/or body type. Those who did not achieve full marks often did not explain why increasing the sample size would make the study more generalisable. Common errors usually included reference to experimental designs such as independent measures and matched pairs which were not creditworthy.
			Total	**3**	
24			 1 mark is awarded for correctly presenting by value each bar representing the overall use of each type of bin 1 mark is awarded for clear labelling of the x axis 1 mark is awarded for clear labelling of the y axis 1 mark is awarded for units of measurement (total values) on the y axis (or x axis if the bar chart is presented the other way around) 4 features included 3 features included 2 features included 1 feature included The candidate has not provided any creditworthy information	Max 4 4 3 2 1 0	- A title is not necessary, but can add clarity to otherwise unclear labels on axes - Labels on axes must be clear. For example just putting 'category' instead of something like 'type of bin' is unclear (*but remember this can be clarified by a title if provided) - Cap at 3 marks if data presented as a histogram (i.e. no gap between bars) rather than a bar chart Cap at 3 marks if bar chart displays male and females separately (ie. Four bars) – the response has not correctly presented by value each bar representing the overall use of each type of bin. **Examiner's Comments** Most candidates achieved either 3 or 4 marks for this question. They were able to draw an appropriate bar chart using the data in the table. The axes were usually labelled correctly. The most common error was drawing four bars in the table which showed the results for males and females using the bin with steps and the bin without steps. These types of responses were given 3 out of 4 marks available. The question asked for the overall use of

<table>
<tr><td></td><td></td><td></td><td></td><td></td><td>the two different types of bins. Therefore, candidates needed to calculate the overall use for the bins with steps (for both males and females) and the bin without steps (for both males and females) and present these as two bars in their bar chart.</td></tr>
<tr><td></td><td></td><td></td><td>Total</td><td>4</td><td></td></tr>
<tr><td>25</td><td></td><td></td><td>23/40 x 100 = 58%
Correct answer with full workings shown
57.5 calculated with full workings shown | OR correct answer (58) but some of the workings are missing.
Correct answer shown to 2 or 3 significant figures with no or incorrect workings | Some of the correct workings are given. (e.g. 23/40=.58 or .575)
The candidate has not provided any creditworthy information</td><td>Max 3 3 2 1 0</td><td>23/40 x 100 = 57.5 = 58 – 3 marks

23/40x100 = 57.5 – 2 marks

23/40=0.58 – 1 mark

58 – 1 mark
57.5 – 1 mark

% sign not required.

<u>Examiner’s Comments</u>

This question was well answered by the majority of candidates and most showed their workings and correctly presented the finding to two significant figures. The most common error was to present the findings to three significant figures (57.5%) or to not show their full workings. A small minority of candidates gave the incorrect workings and incorrect percentage and therefore achieved no marks.</td></tr>
<tr><td></td><td></td><td></td><td>Total</td><td>3</td><td></td></tr>
<tr><td>26</td><td></td><td></td><td>Conclusions could include: the bin with the steps leading up to it was used more often, suggesting that the steps encouraged people to use the bin more, perhaps because of the novelty value that this afforded etc; Perhaps the bin with steps was used more as people were curious where the steps led to and followed them, then deposited their litter in the bin; Could be a conformity effect of more people using the bin with steps leading other people to copy this behaviour and also use the bin, woman may be more likely to care about the environment as they used the bins more than men, etc.
Accept any other appropriate conclusions here.</td><td>Max 6 3 2 1 0</td><td>-Context = bin/bins, litter, steps

-Clear (explicit) interpretation of findings (not simply stating a finding) is required for top band

3 marks could be obtained by justifying their conclusion

For information -
57.5% use of bin with steps
42.5% use of bin without steps
65% overall use of bin by females
35% overall use of bin by males

64% of the males used bin with steps
54% of the females used bin with steps</td></tr>
</table>

			3 marks for each conclusion Clear, detailed response in context Clear, detailed response but not in context **OR** attempt in context Brief and/or weak outline of a conclusion (whether in context or not) **OR** simply stating a finding The candidate has not provided any creditworthy information		**Examiner's Comments** There were some good, full mark responses to this question where the candidate gave two conclusions that could be obtained from the data collected in this study. Those who gave good responses focused on females using the bins more than males and gave an explanation for why this might be true. This included conclusions such as women being tidier than males or caring about the environment more. Many then gave a conclusion about the bins with steps being used more often than the bin without steps. Some candidates referred to 'nudge theory' and how this was effective in encouraging people to throw their rubbish in the bin. One common error was stating that the 'nudge' of the steps was more effective with females than males. The males showed the greatest increase in bin use when they were in the bin with steps condition compared to the bin without steps condition. A very common response was to present the findings of the study without making any conclusions from this data. These responses achieved poorly on this question.
			Total	**6**	
27	a		Any one reason in context from: nominal (categorical) data obtained; looking for a difference; independent groups (unrelated) One appropriate reason in context One appropriate reason but not in context The candidate has not provided any creditworthy information	Max 2 2 1 0	-Context = bin/bins, litter, steps, male/female **Examiner's Comments** This was very well answered by candidates and often the response was in context of this study. The most common response was to identify that the data was nominal with some candidates identifying that the study used an independent measures design. A significant number of candidates did not put their response in context and could, therefore, only achieve 1 mark for their response. A few candidates left this question blank.
	b		df = 1 Workings ... (R-1) x (C-1)	Max 2 2 1 0	(2-1) x (2-1) = 1 – 2 marks (R-1) x (C-1) = 1 – 2 marks

			(2-1) x (2-1) = 1 Correct answer with workings Correct answer but not workings (or workings incomplete/unclear/incorrect) The candidate has not provided any creditworthy information		**Examiner's Comments** Many candidates were able to calculate the degrees of freedom, and gave workings. Some candidates were successful in achieving 1, but not through the correct workings. Other candidates showed some understanding of the formula by using (r-1) and (c-1) but added these instead of multiplied.
	c		3.841 Correct answer provided The candidate has not provided any creditworthy information	Max 1 1 0	If nothing written but correct answer identified in the table – this is creditworthy. **Examiner's Comments** Many candidates gave the correct response to this question (3.841). There was a follow-on error from 30(b). Those who gave an incorrect response tended to choose either 5.991 or 7.815.
	d		X^2 = 0.4058, df =1, p>0.05 1 mark for each correct feature included ... - calculated value - df - correct indication of significance (i.e. '>') 3 correct features 2 correct features 1 correct feature The candidate has not provided any creditworthy information	Max 3 3 2 1 0	Written out version can receive full credit. Eg The Chi Square calculated value is less than the critical value of 3.841. Therefore the difference is not significant at the 5% probability level. 1 mark for comparing the calculated and critical value. 1 mark for identifying the probability is greater than 5% or is not significant at the 5% level of significance. (95% or 1 in 20 is also acceptable) **OR** state the results are not significant 1 mark for 3.841 or df = 1 **Examiner's Comments** Many of the candidates wrote out the significance statement in full. They were able to identify that the calculated value was less than the critical value and this, therefore, meant the results were not significant. Some gave the incorrect calculated value that they had incorrectly selected in 30(b) which lead to achieving fewer marks. A small number of candidates believed that the calculated value of 0.4058 was

<table>
<tr><td></td><td></td><td></td><td></td><td></td><td>greater than the critical value of 3.841.</td></tr>
<tr><td></td><td>e</td><td></td><td>It informs us that there is no difference in the usage of the two different types of bins. People are not more likely to use the bin with steps leading up to it compared to the one without steps.
Clear response in context
Clear response but not in context | Attempt in context
Brief and/or weak attempt (whether in context or not)
The candidate has not provided any creditworthy information</td><td>Max 3 3 2 1 0</td><td>-Context = bin/bins, steps

For full marks the candidate must refer to the bin with steps and the bin without steps.

The null hypothesis is accepted and/or alternative hypothesis rejected - 1 mark

<u>Examiner’s Comments</u>

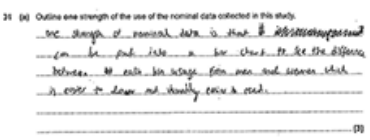
Some candidates did achieve very well for this question and were able to clearly explain what the analysis from the Chi-Square test informed us regarding the use of the two different types of bins in this study. However, many did not achieve full marks due to a lack of clarity. There were several candidates who gave the correct response 30(d) but then responded to this question by saying there was a significant difference in the use of the bins. These types of responses were not creditworthy.</td></tr>
<tr><td></td><td></td><td></td><td>Total</td><td>11</td><td></td></tr>
<tr><td>28</td><td>a</td><td></td><td>Answers could include: easy to collect; easy to analyse/interpret; easy to present in visual (graphical) format; possible to carry out a statistical test; etc
Clear outline of strength in context
Clear outline of strength but not in context | Attempt to outline strength in context
Brief and/or weak attempt to outline strength (whether in context or not)
The candidate has not provided any creditworthy information</td><td>Max 3 3 2 1 0</td><td>-Context = bin/bins, litter, steps, nudge theory etc

(for information) The nominal data is the number of times that the bin with steps and the bin without steps were used
It is also whether the participant is male or female.

For full marks the response must engage with a feature unique to nominal data that leads to the strength (e.g. categories/frequencies)

<u>Examiner’s Comments</u>

Many candidates achieved 2 marks by outlining strengths of quantitative data, rather than engaging with a specific feature of nominal data. As such, relatively few achieved full marks. Most of the responses were contextualised in terms of bin with/without steps.</td></tr>
</table>

<table>
<tr>
<td></td><td></td><td></td><td></td><td></td>
<td>Exemplar 4
 
Exemplar 4 highlights a common response for this question which achieved 2 marks out of a possible 3 marks available. This candidate has correctly identified a strength of nominal data and their response is in context. However, the candidate has not identified a strength that is specific to nominal data as ordinal and interval data could also be put into a bar chart.</td>
</tr>
<tr>
<td></td><td>b</td><td></td>
<td>Answers could include: doesn’t provide reasons for the behaviour observed; easy to miss some behaviours; can be misinterpreted; cannot calculate a mean/median score as participants do not have individual scores, etc
Clear outline of weakness in context
Clear outline of weakness but not in context | Attempt to outline weakness in context
Brief and/or weak attempt to outline weakness (whether in context or not)
The candidate has not provided any creditworthy information</td>
<td>Max 3 3 2 1 0</td>
<td>- Context = bin/bins, litter, steps, nudge theory etc

The nominal data is the number of times that the bin with steps and the bin without steps were used
It is also whether the participant is male or female.

For full marks the response must engage with a feature unique to nominal data that leads to the weakness(e.g. categories/frequencies/discontinuous data etc)

<u>Examiner’s Comments</u>

Like 31(a) many of the candidates did achieve 2 marks by outlining a weakness of quantitative data. Most found it difficult to explain this weakness being due to a feature of nominal data. Stronger responses made a clear contrast with the sophistication that ordinal or interval level data could provide in contrast to nominal level. Most responses were in context.</td>
</tr>
<tr>
<td></td><td></td><td></td><td>Total</td><td>6</td><td></td>
</tr>
<tr>
<td>29</td><td></td><td></td><td>A Type 1 error is a ‘false positive’, meaning that the null hypothesis has been incorrectly rejected (when it is really true). In this study this means</td>
<td>Max 2 2 1 0</td>
<td>-Context = bin/bins, litter, steps

‘false positive’ – 1 mark
‘Incorrectly rejecting the null hypothesis’ – 1 mark</td>
</tr>
</table>

			that there is no real difference in the use of litter bins that have steps up to them compared to those that don't, but it has been claimed that there is a difference. Clear explanation in context Clear explanation but not in context **OR** attempted explanation (whether in context or not) The candidate has not provided any creditworthy information		'Incorrectly accepting the alternate/alternative/experimental hypothesis' – 1 mark **Examiner's Comments** Most candidates did correctly identify a Type 1 error. Common responses included 'false positive', 'falsely rejecting null hypothesis' and 'falsely accepting alternate hypothesis'. Some of the responses were contextualised in terms of the hypothesis for the bin with steps study, but not all.
			Total	**2**	
30			D $r_s = -0.8$	1	**Examiner's Comments** Many answered this correctly.
			Total	**1**	
31			D primar	1	**Examiner's Comments** Nearly all candidates answered this correctly. (99%)
			Total	**1**	
32			B continuous	1	**Examiner's Comments** A number of candidates were unable to give a correct response to this question. C was wrongly given as a response by quite a number of candidates.
			Total	**1**	
33			C the extent to which the findings can be applied to the population	1	**Examiner's Comments** Answered correctly by most candidates. Some candidates chose option A incorrectly.
			Total	**1**	
34			B negatively skewed	1	**Examiner's Comments** A number of candidates were unable to give a correct response to this question. Many candidates chose option D incorrectly.
			Total	**1**	
35				1	**Examiner's Comments**

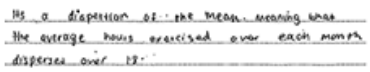
			A Incorrectly accepting the null hypothesis		Answered correctly by most candidates. Some candidates chose option B incorrectly.
			Total	**1**	
36			A when there are a few scores much lower than the rest	1	**Examiner's Comments** A significant number of candidates were unable to give correct response to this question. Incorrect choices by candidates were varied. This would indicate that candidates need to be fully prepared to challenge their understanding of mean, median and mode.
			Total	**1**	
37			A Chi-square	1	**Examiner's Comments** Most candidates answered this question correctly.
			Total	**1**	
38			B/C Chi-square/Mann-Whitney	1	**Examiner's Comments** Most candidates were able to answer this question correctly.
			Total	**1**	
39			C ~	1	**Examiner's Comments** Many candidates answered this question correctly. Some candidates chose option A incorrectly.
			Total	**1**	
40			A criterion	1	**Examiner's Comments** Most candidates answered this question correctly.
			Total	**1**	
41			D self-ratings of aggression (1 to 10) at different times of day (10am to 10pm)	1	**Examiner's Comments** Many candidates answered this question correctly.
			Total	**1**	
42			C $\frac{1}{20}$	1	**Examiner's Comments**

					Most candidates answered this correctly
			Total	**1**	
43			Under the new (2015) BPS guidelines the ethical consideration of 'respect' covers: general respect, informed consent, withdrawal, privacy and confidentiality. Likely answers could include: obtaining parental consent for children under the age of 16; only using children aged 16+ and asking them about their thoughts about morality when they were younger; allowing children to stop participating if they asked to do so, showed any signs of distress; not recording actual names linked to the data collected about morality; only using fictitious accounts in any scenarios used when questioning children about morality Etc 3 marks for each ethical issue addressed Clear outline of how to address ethical issue of 'respect' in context Attempted outline of how to address ethical issue of 'respect' in context **OR** clear outline of how to address ethical issue of 'respect' but not in context Brief and/or weak outline of how to address ethical issue of 'respect' (whether in context or not) The candidate has not provided any creditworthy information	Max 6 3 2 1 0	-Context = morality, morals, good/bad, right/wrong -Responses must relate to the ('new' 2015) BPS guidelines covering 'respect' general respect (respect people's individuality and not show prejudice etc)– i.e.: Informed consent; consent, withdrawal; privacy, confidentiality (anonymity) are all creditworthy. **Examiner's Comments** Candidates who scored well on this question were able to identify two ways 'respect' could be addressed in this study. They could distinguish 'respect' from 'responsibility' and provided context.
			Total	**6**	
44			Likely answers: biased sample (e.g. in terms of its size or diversity or gender or age ranges etc); nature of the questions asked (e.g. if too restricted / limiting – enquiring about only one specific aspect of morality etc) Clear identification of something that could lower generalisability in context Clear identification of something that could lower **OR** attempt to identify something that could lower	Max 2 2 1	-Context = morality, morals, good/bad, right/wrong **Examiner's Comments** Most candidates seemed to have a broad understanding of generalisation but did not link it explicitly enough to the context of the findings of this particular study, therefore capping their marks at 1 out of a possible 2 marks available. **Exemplar 3**

			generalisability, but not in context / generalisability in context or not	0	A small sample size would lower generalisability as population validity would be reduced, for the results collected on partipan participants view on morality Exemplar 3 provides a brief but good enough response to meet the criteria for full marks.
			The candidate has not provided any creditworthy information		
			Total	**2**	
45			Conclusions could include: there was a big variety in the number of hours of exercise taken each month, ranging from 0 to 60 hours suggesting that some people either don't have the time or don't value taking exercise; Ratings of stress varied a lot, from 10 to 100 suggesting that stress affects people in many different ways and some people are able to cope with stress better than others perhaps; the relationship between stress and exercise is mixed, as some people who take a lot of exercise (e.g. person 'j') seem to have a lot of stress in their life, whereas for others (e.g. person 'a') taking a lot of exercise is associated with lower levels of stress. So perhaps there is no significant relationship between exercise and stress, so if there is a benefit this does not apply everyone. Accept any other appropriate conclusions here. **3 marks for each conclusion** Clear, detailed response in context	Max 6 3	-Context = stress and exercise etc -Clear (explicit) interpretation of findings (not simply stating a finding) is required for top band -explicit reference to cause-and-effect is not creditworthy **Examiner's Comments** This question elicited a variety of responses. It required candidates to interpret a table of data and go beyond merely restating an individual finding, as there are 6 marks available. . The most common error was stating that more exercise causes less stress. Some candidates correctly stated one could conclude that there were clear individual differences in the data presented and at best one could conclude there was a weak or no correlation between exercise and stress as another conclusion
			Clear, detailed response but not in context / **OR** attempt in context	2	
			Brief and/or weak attempt to outline a conclusion (whether in context or not) / **OR** simply stating a finding	1	**AfI** Candidates would benefit from being given opportunities to examine raw data tables and practice extrapolating what conclusions could be drawn from them by looking at the nature of the data and any patterns that can be discerned. These skills can also be illustrated through practical work data gathered in preparation for this exam.
			The candidate has not provided any creditworthy information	0	
			Total	**6**	

46			 1 mark is awarded for correctly plotting the data 1 mark is awarded for clear labelling of the x axis 1 mark is awarded for clear labelling of the y axis 1 mark is awarded for units of measurement on both axes All features included 3 features included 2 features included 1 feature included The candidate has not provided any creditworthy information	Max 4 4 3 2 1 0	-* A title is not essential, but can add clarity to otherwise unclear labels on axes -Labels on axes must be clear. For example just putting 'exercise' is unclear (*but remember this can be clarified by a title if provided) co-variables exercise or stress can be on x or y axis. **Examiner's Comments** This was the best answered question on the paper with most candidates presenting a fully labelled and scaled scattergraph with correct plotting of data.
			Total	**4**	
47			Possible things include: participants not wanting to disclose how much/little exercise they engage in (social desirability); participants not wanting to disclose how much stress they experience for fear of embarrassment; problems using a quantitative scale to convey stress experienced (can only report the amount, not type of stress experienced) etc etc. For each thing 3 marks max … Clear outline in context Attempted outline in context / **OR** Clear outline, but not in context Brief and/or weak outline (whether in context or not) The candidate has not provided any creditworthy information	Max 6 3 2 1 0	-Context = stress and exercise etc -Accept as creditworthy the acknowledgement that it would not be valid to try to establish cause-and-effect from the data collected in a correlation study -Accept sample size as something that could affect validity. -accept extraneous variables -subjective interpretation -individual differences **Examiner's Comments** There was good engagement with this question with a range of different issues offered. These popularly included interpretation bias, social desirability, demand characteristics, and extraneous variables. Having presented the 'thing' the biggest weakness was making it specifically link to validity and the nature of the impact it has "in this study".
			Total	**6**	

48			Reasons: -test of correlation/relationship (and the study investigated the relationship between exercise and stress) - at least ordinal data collected (number of hours exercised and ratings of stress) Two appropriate reasons provided, both in context Two appropriate reasons provided, but only one, or neither in context One appropriate reason provided (whether in context or not) The candidate has not provided any creditworthy information	Max 3 3 2 1 0	-Context = stress and exercise etc **Examiner's Comments** There were quite a few non responders to this question. A very common response was to cite correlation but did not provide any link to the nature of the data collected. These responses achieved poorly on this question.
			Total	**3**	
49			This means that there was a weak, negative correlation between the amount of exercise taken each month and the amount of stress experienced. Reference to both the strength (weak) and direction (negative) of the correlation in context Reference to both the strength (weak) and direction (negative) of the correlation, but not in context **OR** Reference to either the strength (weak) OR direction (negative) in context Reference to either the strength (weak) or direction (negative) of the correlation, but neither in context. The candidate has not provided any creditworthy information	Max 3 3 2 1 0	-Context = stress and exercise etc -Any reference to 'difference' as opposed to correlation or relationship between exercise and stress is not creditworthy. **Examiner's Comments** Non responders to Question 26 did same for this question. Those who gave good responses referred to both the direction and the strength of the correlation in context, to gain the full 3 marks available.. One common error was to respond regarding statistical significance.
			Total	**3**	
50	a		18 Correct answer written to two significant figures The candidate has not provided any creditworthy information	Max 1 1 0	**Examiner's Comments** Most candidates presented this correctly, but a minority did not reduce to two significant figures
	b		This informs us that the amount of exercise taken by each participant each month varies quite a lot, with some individuals engaging in a lot more exercise than others. Clear explanation in context	Max 2 2 1	Explanation must be focused on standard deviation. **Examiner's Comments** The concept of standard deviation appears to be poorly understood as the majority of candidates did not

			Clear explanation but not in context **OR** attempted explanation in context or not The candidate has not provided any creditworthy information	0	provide a correct response. Many candidates related it to 18 being the average figure rather than relating it to an indication of spread. Those that did mention spread/dispersion sometimes implied they thought all scores would lie within that range. **Exemplar 4**  Exemplar 4 provides a typical response where the candidate has the idea of spread but is confused as to what standard deviation really informs us in relation to this study. It gained 1 mark as it is an example of a weak attempt.
			Total	**3**	
51			C	1	**Examiner's Comments** A minority were not clear on significance level statements with B or D being given as a response.
			Total	**1**	
52			C	1	**Examiner's Comments** Answered correctly by most candidates.
			Total	**1**	
53			B	1	**Examiner's Comments** A number of candidates were unable to give a correct response to this question.
			Total	**1**	
54			B	1	**Examiner's Comments** A significant number of candidates were unable to give a correct response to this question **Misconception** A significant number of candidates

					answered this question on correctly, offering a range of answers. It is therefore important to that candidates recognise equations and what the symbols within a formula mean. **OCR support** https:/www.ocr.org.uk/Images/260143-inferential-statistics-parametric-and-non-parametric-student-workbook.docx
			Total	**1**	
55			C	1	**Examiner's Comments** A minority miscalculated this and gave 17.5
			Total	**1**	
56	a		Mean stood up = 107/6 = 17.8333333 2 DPs = 17.83	**Max 2**	**Examiner's Comments** Most candidates were able to demonstrate their knowledge of how to calculate the mean, but some did not go on to present their answer correctly to two decimal places.
			Mean correctly calculated and presented to 2 decimal places with workings shown	**2**	
			Mean correctly calculated and presented to 2 decimal places but no workings shown **OR** correct workings but answer not presented to two decimal places / incorrect	**1**	
			The candidate has not provided any creditworthy information	**0**	
	b		Mean sat down = 73/6 = 12.166666 2 SFs= 12	**Max 2**	**Examiner's Comments** Similarly here, most candidates were able to demonstrate their knowledge of how to calculate the mean, but some did not go on to present their answer correctly to two significant figures (and some confused significant figures with decimal places).
			Mean correctly calculated and presented to 2 significant figures with workings shown	**2**	
			Mean correctly calculated and presented to 2 significant figures but no workings shown **OR** correct workings but answer not presented to two significant figures / incorrect	**1**	
			The candidate has not provided any creditworthy information	**0**	
	c		Mean stood up = 107/6 = 17.83 mean %age recall = **17.83/20 x 100 =**	**Max 2**	

			89.15% Mean sat down = 73/6 = 12.17 mean %age recall = **12.17/20 x 100 = 60.85%** Due to variations in the actual figure used for the mean (based on number of decimal places and / or number of significant figures used) acceptable answers can be *anywhere* in the following ranges **Stood up** condition: **89% to 90%** **Sat down** condition: **60% to 61%** Mean percentage correctly calculated for each condition with workings Mean percentage correctly calculated for one condition with workings The candidate has not provided any creditworthy information	**2** **1** **0**	-Due to the error in the wording of this question, candidates who have used the data from the table presented in Section A (Q2) should also receive credit. Mean 'young condition' 26.5/30 x 100 = **88.33**% (acceptable presented to any number of decimal places, or two significant figures) Mean 'old condition' 15/30 x 100 = 50% **<u>Examiner's Comments</u>** Due to the error in the question paper, all candidates were awarded the full 2 marks.
			Total	**6**	
57	a		Step 1: work out the difference of each individual score on the maths test compared to the mean Step 2: square this difference Step 3: find the sum of all the differences squared Step 4: divide the sum of the differences squared by the total number of scores / participants (or total minus 1) Step 5: find the square root One mark for each step of how to calculate the standard deviation up to a maximum of five marks with at least one step to be in context for full marks The candidate has not provided any creditworthy information	**Max 5** **5** **0**	Context = 'stand' / 'standing up', 'maths test', 'concentration' etc -Cap at 4 marks if correct description of how to calculate the standard deviation but not in context -If the steps are presented in the wrong sequence (i.e. that would result in an incorrect calculation of the sd) then only credit up to the last correct step in sequence **<u>Examiner's Comments</u>** This proved to be a challenging question, even though the formula for standard deviation had been provided in a previous question (2b) in Section A. The best responses were characterised by a logical, step-by-step approach of the five key stages involved (subtraction of the mean from each individual score, squaring of this result, summation of these values, division by sample size and finally square rooting). Not all candidates approached the question in this way, with some combining two or more steps in one, which was still creditworthy. For example, saying that you would *divide the total of each individual score minus the mean by the sample size.* However, this did

<table>
<tr><td></td><td></td><td></td><td></td><td></td><td>sometimes get confusing and could be ambiguous. Some candidates also did not seem to appreciate the importance of performing each step in the correct sequence in terms of the effect on the outcome of the overall calculation, and proposed combining / merging some of the steps without an appreciation of the effect of this. For example, some suggested adding all individual scores together first then subtracting the mean from this value, then squaring the result.

Although seemingly difficult and complex at first, the standard deviation is essentially a straightforward concept in that it is simply the mean of by how much each individual score differs from the overall mean score (mean of the mean difference in effect). It is also worth conveying to students at the time of explaining this that it is a very useful statistic that provides something in addition to the mean (knowledge of this had implications for understanding and responding to the next two questions).</td></tr>
<tr><td></td><td>b</td><td></td><td>The standard deviation informs us about the dispersion of scores around the average, so in this study how much variation there was in the typical way a pupil performed depending on whether they were stood up or sat down.

Clear outline of what the standard deviation informs us in context

Attempt whether in context or not OR clear outline of what the standard deviation informs us but not in context

The candidate has not provided any creditworthy information</td><td>Max 2

2

1

0</td><td>Context = ‘stand’ / ‘standing up’, ‘maths test’, ‘concentration’ etc

<u>Examiner’s Comments</u>

Some candidates incorrectly assumed the standard deviation could inform us about the overall performance within each condition in terms of how well or not participants had done, claiming that a high figure for the standard deviation indicated better performance.</td></tr>
<tr><td></td><td>c</td><td></td><td>It informs us that when sat down there is much more variation in performance, with some pupils doing very well on the test and others performing very poorly. When stood up there is much less variation in how pupils perform on the test. Therefore, overall there is more variation in test</td><td>Max 4</td><td>Context = ‘stand’ / ‘standing up’, ‘maths test’, ‘concentration’ etc

-Responses stating / indicating that the higher the standard deviation score the better the performance / concentration ability are not creditworthy</td></tr>
</table>

			performance when sat down compared to standing up with some pupils seeming to benefit from it whereas others not. Clear description of what the findings of the calculation of standard deviation inform us for one or both conditions of the experiment in context Attempt to describe what the findings of the calculation of standard deviation inform us for one or both conditions of the experiment whether in context or not **OR** Clear description of what the findings of the calculation of standard deviation inform us for one or both conditions of the experiment but not in context The candidate has not provided any creditworthy information	**3-4** **1-2** **0**	**Examiner's Comments** The misunderstanding about what the standard deviation informs us evident in some candidates' responses to the previous question was apparent again here, when discussing what the calculation of the standard deviation for each condition informs us. The best responses here acknowledged and reported that the standard deviation for the 'stood-up group' (1.72) indicated that there was less variation in the scores obtained in the maths test in this condition than in the 'sat-down group' where the standard deviation was higher (5.60). They went on to explain how this may suggest that standing up may have had a similar effect in how participants performed compared to sitting down where the scores varied much more, suggesting if affected how well some participants did much more than others.
			Total	**11**	
58	a		The correct test would be the Mann Whitney U test. This is because … 1. It is a test that examines differences between performance in two conditions and the study compared test scores whilst stood up to sitting down) 2. it is a test that is used for independent measures designs where the scores in each condition come from different participants and there where different pupils' scores in the standing up compared to sitting down conditions 3. It is a test that requires ordinal level data which the study had because scores out of 20 in a maths test can be ranked Test correctly identified and at least one justification for its selection referred to in context Test correctly identified but not justified, or not justified in context, or justified incorrectly The candidate has not provided any creditworthy information	**Max 2** **2** **1** **0**	Context = 'stand' / 'standing up', 'maths test', 'concentration' etc **Examiner's Comments** This question revealed the need to be knowledgeable about the criteria for the selection of non-parametric inferential statistical tests. Even those candidates who were correctly able to identify the test sometimes struggled to provide a rationale for their choice (implying that a guess had been made). The best responses were those that correctly identified the test and provided a rationale in context (e.g. explaining that it was an independent measures design, as there were a different group of participants in the 'stood-up' condition compared to the 'sat-down' condition).

	b		The data would be ranked by considering all the scores from each condition ('stood up' and 'sat down') **together as one group**, assigning numbers to denote position in an ordered sequence. The lowest score would receive rank 1, the next score rank 2 and so on (or awarding the highest score rank 1 and so on, providing consistency is maintained) Clear outline of how to rank data in context Attempt to outline how to rank the data whether in context or not **OR** clear outline of how to rank data but not in context The candidate has not provided any creditworthy information	**Max 2** **2** **1** **0**	Context = 'stand' / 'standing up', 'maths test', 'concentration' etc **Examiner's Comments** This proved to be a challenging question. Many candidates were aware of the basic principle of ranking in that it involves assigning numbers to denote position in an ordered sequence. However, not many acknowledged that for the Mann Whitney U test this involved ranking the data collectively as one whole group. It was impressive that some candidates clearly did understand this and some even providing the 9 correct ranks for the data (not that this was necessary to achieve full marks here). This demonstrates the need and importance of not only making sure students are aware of the criteria for the use of the inferential statistical tests, but actually practice having a go at performing some calculations using them.
			Total	**4**	
59	a		The correct test would be the Chi square test. This is because … 1. It is a test that examines differences between different types of student and whether they opened the box or not 2. It is a test that is used for independent measures designs where the values in each condition come from different participants and there where different types of student (science, arts, humanities) 3. It is a test that uses nominal level data and the frequency counts of number of students who did or did not open the box is nominal Test correctly identified and two justifications for its selection referred to with one in context Test correctly identified and two justifications for its selection referred to but neither in context Test correctly identified but with no justification (or incorrect justifications offered) The candidate has not provided any creditworthy information	**Max 3** **3** **2** **1** **0**	-Context = student(s) (science, arts, humanities), did open, did not open box

<table>
<tr>
<td></td>
<td>b</td>
<td></td>
<td>
Expected frequencies (E) for use in a chi square test are calculated separately for each cell using the formula row total x column total / overall total. For example, in this study the E value for the cell ‘Science students who did open the box’ would be …
60 x 62 / 150 = 24.8. The table below shows the calculated E values (in brackets) for each cell (although these are not required, some students may include them as part of their explanation of how to calculate them).
<table>
<tr><th>Type of student</th><th>DID open the box</th><th>Did NOT open the box</th></tr>
<tr><td>Science</td><td>39 (E=24.8)</td><td>21 (E=35.2)</td></tr>
<tr><td>Arts</td><td>15 (E=24.8)</td><td>45 (E=35.2)</td></tr>
<tr><td>Humanities</td><td>8 (E=12.4)</td><td>22 (E=17.6)</td></tr>
</table>
Clear explanation of how to calculate the expected frequencies for use in a Chi square test
Attempts to explain how to calculate the expected frequencies in context OR clear explanation of how to calculate the expected frequencies but not in context
The candidate has not provided any creditworthy information
</td>
<td>
Max 2

2

1

0
</td>
<td>
-Context = student(s) (science, arts, humanities), did open, did not open box

Calculations (if correct) for the actual E values, if provided, are also creditworthy.
</td>
</tr>
<tr>
<td></td>
<td>c</td>
<td></td>
<td>
The symbol ‘>’ means ‘greater than. In this study it would mean the probability of the null hypothesis being true (i.e. that there is no difference between different types of students and whether they open the box or not) is greater than 5% so the null hypothesis is accepted and the alternative hypothesis is rejected (the findings are not (statistically) significant).
Clear explanation of what the symbol ’>’ means in context
Attempts to explain what the symbol ’>’ means OR clear explanation of what the
</td>
<td>
Max 2

2

1

0
</td>
<td>
-Context = student(s) (science, arts, humanities), did open, did not open box

-Also accept reference to interpretation of the symbol ‘003C’ in relation to other levels of probability in addition to 5% (e.g. 2% or 1% etc)
</td>
</tr>
</table>

			in context in context / symbol'>' means but not in context The candidate has not provided any creditworthy information		
			Total	**7**	
60	a		B	1	**Examiner's Comments** Mostly correct responses
	b		C	1	**Examiner's Comments** Mostly correct responses
	c		C	1	**Examiner's Comments** Mostly correct responses
			Total	**3**	
61	a		157/20 = 7.85 The mean is 7.85, so to two significant figures = 7.9 Mean correctly stated to two significant figures with all workings shown Mean correctly stated to two significant figures but with no workings shown / **OR** workings shown but mean not written to two significant figures Mean only stated but not to two significant figures The candidate has not provided any creditworthy information	**Max 3** **3** **2** **1** **0**	-Remember to check table of data presented in the question as some candidates may have written their answer here 2 mark = 157/20 = 7.85 (as not to two SFs) **Examiner's Comments** Most candidates were able to calculate the mean correctly, but some lost a mark by not presenting the final answer to two significant figures.
	b		1,1,2,3,3,3,3,3,3,**4,5**,5,5,5,5,6,6,7,7,9 The median is 4.5 Median correctly stated with workings shown Median correctly stated but with no workings shown / **OR** workings shown without answer being stated The candidate has not provided any creditworthy information	**Max 2** **2** **1** **0**	-Remember to check table of data presented in the question as some candidates may have written their answer here **Examiner's Comments** Most candidates were able to calculate the median correctly
	c		1,1,2,**3,3,3,3,3,3**,4,5,5,5,5,5,5,6,6,7,7,9 The mode is 3 Mode correctly stated The candidate has not provided any creditworthy information	**Max 1** **1** **0**	-Remember to check table of data presented in the question as some candidates may have written their answer here **Examiner's Comments** Most candidates were able to identify the mode correctly

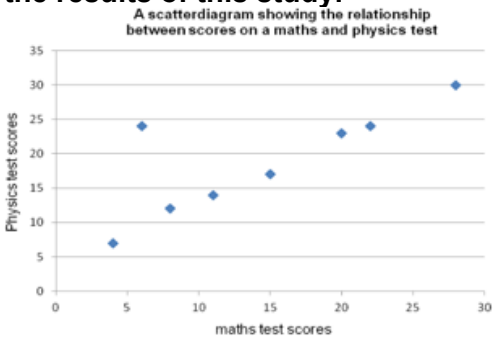
			Total	**6**	
62	a		Males range = 5 (10-5) Females range = 8 (9-1) a * Also accept answers from the alternative calculation for the range that includes +1 Males range = 6 (10-5(+1)) Females range = 9 (9-1(+1)) * Combined males & females (10-1) = 9 (or +1 = 10)	**Max 4**	If not clear what refers to male and female (i.e. if just figures presented) cap 2 -* If a candidate has calculated the range overall (across males and females combined) correctly, and shown workings can be awarded 4 marks. For example, If some lack of clarity in either the labelling of what the figure presented refers to and / or the calculations cap at 2 **Examiner's Comments** Most candidates were able to calculate the range for the males and the females correctly, and show workings. It was acceptable here to use the +1 variation of the process for determining the range. Some candidates also interpreted the question as requiring the overall range (aggregated across all the males and females' ratings collectively), which was also creditworthy.
			For each calculation of the range (i.e. for males and females) …	**2**	
			Range correctly stated with correct workings shown	**1**	
			Range correctly stated but with no workings shown **OR** correct workings shown without answer being stated or answer wrongly stated		
			The candidate has not provided any creditworthy information	**0**	
	b		For example… Females (range = 9) vary more than males (range = 6) in how much they think the importance of physical appearance is for being in a relationship. Males are more consistent, suggesting men are of the same or similar opinion that appearance does matter in terms of forming relationships than females. Accept any other appropriate conclusion.	**Max 3**	-Context = love (and things 'love related') -Note: a conclusion is an interpretation of a finding (here the interpretation of the range). It is not creditworthy just to state findings again on their own -Nothing creditworthy for conclusions based on other descriptive statistics (e.g. the mean), or incorrectly interpreting the range as an indicator average rating (e.g. claiming that females regard appearance as more important than males)
			Clear outline of conclusion in context	**3**	
			Attempt in context	**2**	
			Attempt but not in context	**1**	
			The candidate has not provided any creditworthy information	**0**	1 mark example *There is more variation in females than males scores* 2 mark example *There is more variation in females than males scores about physical appearance* Example 3 marks *Males are more consistent,*

					suggesting men are of the same or similar opinion that appearance does matter in terms of forming relationships than females. **Examiner's Comments** Overall, this question was poorly answered and showed a quite widespread fundamental lack of understanding in how to interpret the range. Many candidates confused the range with the mean, making comments that related more to the mean when attempting to reach a conclusion. Even those who demonstrated awareness of what the range really is often discussed things in general (ie stating there was more variation in the ratings given by females), rather than discussing things in context of the research scenario provided. This highlights the importance of covering even the more basic statistical concepts in sufficient detail and in an applied context so that candidates appreciate more fully the information that can be gained from the use of such statistics.
			Total	**7**	
63	a		The answer presented as the X^2 value will vary slightly depending on how many decimal places are used at different stages of the calculation. Therefore … accept the overall X^2 answer as anything between 5.0 and 5.02 (e.g. 5.0, 5.01278772, 5.012, 5.02) and to any number of decimal places **Step-by-step calculations for Chi square test** Cell | O-E | $(O-E)^2$ | $(O-E)^2 \setminus E$ | $\sum (O-E)^2 \setminus E$ a | 5 – 8.5 = –3.5 | $(–3.5)^2$ = 12.25 | 12.25 / 8.5 = 1.441 | 1.441 b | 15 – 11.5 = 3.5 | $(3.5)^2$ = 12.25 | 12.25 / 11.5 1.065 | 1.065 c | 12 – 8.5 = 3.5 | $(3.5)^2$ = 12.25 | 12.25 / 8.5 1.441 | 1.441 d | 8 – 11.5 = –3.5 | $(–3.5)^2$ = 12.25 | 12.25 / 11.5 1.065 | 1.065 **5.012**	**Max 5**	Rationale for allocation of marks … As the expected frequencies are already provided, there are essentially four steps left to perform to reach the correct calculation. As there are 5 marks in total to be awarded, the breakdown is one mark for each step correctly performed, with the extra mark for showing all relevant workings. * Please note where candidates have calculated X^2 cell-by-cell this is also creditworthy -Alternative ways of using the formula, providing it arrives at the correct cell value (expressed as a decimal or even fraction) and overall answer is creditworthy (e.g. some students who learn about X^2 in different subjects may do this) **Examiner's Comments**

			Correct calculation with all workings shown Correct calculation of the sum of $(O-E)^2/E$ but with some or all workings missing Correct calculation of each individual $(O-E)^2/E$ value Correct calculation of each individual $(O-E)^2$ value Correct calculation of each individual (O-E) value The candidate has not provided any creditworthy information * **OR** 1 mark for each set of calculations done correctly for each cell ($(O-E)^2/E$) Plus 1 mark for overall correct answer (X^2 value)	**5** **4** **3** **2** **1** **0**	It was pleasing that many candidates were able to correctly calculate Chi square. Candidates who used variations on the formula provided (eg conversion in to fractions) were credited, provided each individual cell value was correct and the overall Chi square value. It was also acceptable to work with and present the overall answer to any number of decimal places. However, it was noticeable, that some candidates made errors when choosing to present to a certain number of decimal places in how they rounded up or down (which has implications for the overall Chi square value arrived at).
	b		The critical value is 3.841 Critical value correctly stated Degrees of freedom correctly identified as 1 but without stating the critical value from the table (or incorrectly stating the critical value) The candidate has not provided any creditworthy information	**Max 2** **2** **1** **0**	-If a candidate just circles the correct critical value in the table award 2 marks (as to be able to do this requires knowledge of what df to use) -Award 1 mark if df = 1 is just circled on the table (so remember to check table) **<u>Examiner's Comments</u>** This question highlighted the importance of practising using the inferential statistical formula on the specification to become fully conversant with the various steps and stages involved in determining significance. It was clear that some candidates had done this and then found this a relatively straightforward question to answer. However, others struggled and did not seem to know much / anything about the concept of degrees of freedom, which was essentially in being able to arrive at the correct answer to this question.
	c		$X^2 = 5.01$, df=1, $p<0.05$ **OR** In words e.g. this is a significant result (at the 5% probability level) because the calculated value of X^2 had to be equal	**Max 2**	Example 2 marks (written version) *As the calculated value is higher than the critical it is significant* -Example 1 mark response *The null is rejected,* and / or *the alternative hypothesis is accepted*

			to or exceed the table critical value (3.841) to be significant Correctly written significance statement (calculated value, degrees of freedom and probability level) **OR** written in words rather than a formal statement just stating p<0.05 **OR** weak and / or brief written response The candidate has not provided any creditworthy information	**2** **1** **0**	**Examiner's Comments** Some candidates in response to this question provided the full, formal statement of statistical significance that would appear in the write-up of a practical report when citing the outcomes of the Chi square test (X^2 = 5.01, df=1, p<0.05). Others interpreted the question as asking for a more general, informal presentation of the outcome of the Chi square test. Both types of response were credited. However, some candidates omitted this question altogether showing the need to cover this when explaining what happens after the outcome of an inferential statistical test has been performed.
			Total	**9**	
64	a		**Outline how each of the following ethical considerations could have been dealt with in this study. (a) Confidentiality** A maximum of 2 marks for how the consideration could have been dealt with. 1 mark for a brief or basic response, or 2 marks for a more developed response which is also suitably contextualised, e.g. 'test scores are kept anonymous (1), for example by making sure that numbers are used to match scores not names (1)'. Other appropriate responses should be credited.	2 AO2 c/g	Examiners need to check for overlap – for example, candidates may address confidentiality or consent under protection of participants but are likely to have been credited for this already.
	b		**Outline how each of the following ethical considerations could have been dealt with in this study. (b) Informed consent** A maximum of 2 marks for how the consideration could have been dealt with. 1 mark for a brief or basic response, or 2 marks for a more developed response which is also suitably contextualised, e.g. 'participants are fully briefed (1) so they can agree to being tested on the two abilities (1)'.	2 AO2 c/g	Examiners need to check for overlap – for example, candidates may address confidentiality or consent under protection of participants but are likely to have been credited for this already.

			Other appropriate responses should be credited.		
	c		**Outline how each of the following ethical considerations could have been dealt with in this study. (c) Protection of participant** A maximum of 2 marks for how the consideration could have been dealt with. 1 mark for a brief or basic response, or 2 marks for a more developed response which is also suitably contextualised, e.g. 'participants are told of their right to withdraw (1) at the beginning and reminded of this throughout the language ability and physical coordination tests (1)'. Other appropriate responses should be credited.	2 AO2 c/g	Examiners need to check for overlap – for example, candidates may address confidentiality or consent under protection of participants but are likely to have been credited for this already.
			Total	**6**	
65			C	1	
			Total	**1**	
66			C	1	
			Total	**1**	
67			D	1	
			Total	**1**	
68			A	1	
			Total	**1**	
69			B	1	
			Total	**1**	
70			B	1	
			Total	**1**	
71			A	1	
			Total	**1**	
72			A	1	
			Total	**1**	
73			**Identify two findings from the data presented in this table.**	4	-Context = GCSE, maths, physics Max 2 x AO2 marks for each finding

			For example: overall test scores higher in physics test; highest overall score was 30 in physics test; lowest score was 4 in maths test; larger range of scores in maths test (4 to 28); mode is 24 for physics scores etc **3-4 marks:** Clear identification of two findings **1-2 marks:** Attempt to identify two findings OR clear identification of one finding **0 marks**: No creditworthy response.		identified
			Total	**4**	
74			**Draw a scatter diagram displaying the results of this study.**  1 mark is awarded for correctly plotting the data 1 mark is awarded for including units of measurement on both axes 1 mark is awarded for clear labelling of each axis 1 mark is awarded for a clear and appropriate title **4 marks:** All features identified **3 marks**: 3 features identified **2 marks**: 2 features identified **1 mark:** 1 feature identified **0 marks**: No creditworthy response.	4	
			Total	**4**	
75	a		**Calculate the range for each test taken.** Maths test: 28 - 4 = 24 OR (28 - 4) + 1 = 25 Physics test: 30 - 7 = 23 OR (30 - 7) + 1 = 24	2	-Accept either pure range calculation (highest value minus lowest), or the measurement error calculation (highest value – lowest value + 1) 1 x AO1 mark for knowledge and understanding of what the range is and how to calculate it

			2 marks: correct calculation of range for both the maths and physics tests data **1 mark:** correct calculation of the range for either the maths test or physics test data **0 marks**: No creditworthy response.		1 x AO2 mark for calculation of the range with data presented
	b		**What conclusion can be reached by interpreting the range for each test?** Possible answers could include: The range for both the maths and physics tests are very similar (24 and 23) indicating that the variation in individual performances in the tests is very similar for both the maths and physics test (suggesting people who are good at maths are also good at physics and vice versa); the range for the maths scores is quite large (24) indicating there is a big variation in individuals maths ability, with some scoring very high and others very low on the test; the range for the physics scores is quite large (23) indicating there is a big variation in individuals maths ability, with some scoring very high and others very low on the test. **3-4 marks:** Clear conclusion in context **1-2 marks:** Clear conclusion but not in context OR attempted conclusion in context **0 marks**: No creditworthy response.	4	-Context = GCSE, maths, physics 2 x AO2 marks for evaluation of what range indicates in this study 2 x AO3 marks for application of what the range is in this study
	c		**Suggest one advantage of using standard deviation instead of the range to analyse the data from each test.** Advantages include: more accurate as involves all individual scores, unlike the range that only considers the highest and lowest values; less affected by extreme values; more likely to produce a more	3	-Context = GCSE, maths, physics 1 x AO1 mark for knowledge and understanding of what standard deviation is 2 x AO3 marks for evaluating the use of standard deviation in this study

			representative figure. **3 marks**: Clear and detailed outline of advantage in context **2 marks:** Clear brief outline of advantage, but in context **OR** clear and detailed outline of advantage, but not in context **1 mark:** Attempt to outline advantage, but lacks clarity / detail (whether in context or not) **0 marks**: No creditworthy response.		
			Total	**9**	
76	a		**The psychologist used the Spearman's ranked correlation coefficient test to analyse the data from this study. Explain why this was an appropriate test to use.** Rationale for use of the test could include: because it is a test of correlation and the researchers were investigating the relationship between scores in a maths test and physics test; because ordinal level data (maths and physics test scores) was obtained and the Spearman's test requires this level of data **2 marks:** Clear explanation of why the Spearman's ranked correlation coefficient test was appropriate **1 mark:** Attempt to explain why the Spearman's ranked correlation coefficient test was appropriate **0 marks**: No creditworthy response.	2	-Context = GCSE, maths, physics 1 x AO1 mark for knowledge and understanding of what the Spearman's ranked correlation coefficient test is 1 x AO2 mark for application of knowledge about the criteria for using the Spearman's ranked correlation coefficient test
	b		**Explain how the data would be ranked for use in this test.** Test scores would be assigned a number, using rank 1 for the lowest score, rank 2 for the next score and so on for both the maths test and the physics test (but each set of scores ranked separately to each other). **2 marks:** Clear explanation of how to rank the data **1 mark:** Attempt to explain how to	2	-Context = GCSE, maths, physics 1 x AO1 mark for knowledge and understanding of what ranking data involves 1 x AO2 mark for application of knowledge of how to rank the data in this study for use in the Spearman's ranked correlation coefficient test NB Rank 1 can be assigned to the highest value score rather than the lowest providing the rest of the ranks are assigned consistently with this

			rank the data **0 marks**: No creditworthy response.		9i.e. the second highest score is assign rank 2 and so on
	c		**The inferential test result produced a calculated value of +0.7083. Explain what this means.** The + sign indicates it is a positive correlation meaning that the higher the maths scores the higher the physics scores. The value of +0.7083 indicates a strong positive correlation between the maths and physics scores **3 marks:** Clear and detailed explanation in context **2 marks:** Clear and detailed explanation but not in context **1 mark**: Attempted explanation whether in context or not **0 marks**: No creditworthy response.	3	-Context = GCSE, maths, physics 1 x AO1 mark for knowledge of what a positive correlation is 2 x AO2 marks for application of knowledge and understanding of what a strong positive correlation is in this study
			Total	**7**	
77	a		**What does the term 'critical value' refer to?** A critical value is a figure in a table of critical values that the answer from an inferential statistical test is compared with to check if the findings are statistically significant or not at a given level of probability **1 mark:** Critical value defined correctly **0 marks**: No creditworthy response.	1	
	b		**How would the critical value be obtained in this study?** By using a table of critical values for the Spearman's ranked correlation coefficient test and using the number of participants that data from each test (maths and physics) was obtained from (8) to look up and locate the correct critical value	2	-Context = GCSE, maths, physics

<table>
<tr><td></td><td></td><td></td><td>2 marks: Clear explanation of how the critical value would be obtained

1 mark: Attempt to explain how the critical value would be obtained
0 marks: No creditworthy response.</td><td></td><td></td></tr>
<tr><td></td><td></td><td></td><td>Total</td><td>3</td><td></td></tr>
<tr><td>78</td><td></td><td></td><td>D</td><td>1</td><td></td></tr>
<tr><td></td><td></td><td></td><td>Total</td><td>1</td><td></td></tr>
<tr><td>79</td><td></td><td></td><td>B</td><td>1</td><td></td></tr>
<tr><td></td><td></td><td></td><td>Total</td><td>1</td><td></td></tr>
<tr><td>80</td><td></td><td></td><td>B</td><td>1</td><td></td></tr>
<tr><td></td><td></td><td></td><td>Total</td><td>1</td><td></td></tr>
<tr><td>81</td><td></td><td></td><td>In general, population validity is a form of external validity that refers to the extent to which the findings from research can be applied to other people who did not take part directly in the research.

Examples of points that could be made here include: size of the sample (40) and how this affects generalising the findings to others; balance of males and females (20 males, 20 females in the sample); potential bias in the sample due to the use of the self-selected sampling method used (so only those interested in, or currently in, or out of love may have volunteered to take part etc). etc

Clear evaluation with two or more points with some context (for 6 marks two of the points must be in context)

Clear evaluation with two or more points made but not in context | OR one clear evaluation point in context | OR two points, one in context one not

Attempt to evaluate population validity (whether in context or not)

The candidate has not provided any creditworthy information</td><td>Max 6

5-6

3-4

1-2

0</td><td>-Context = love (and things ‘love related’)

-Accept positive and / or negative evaluation points as creditworthy

-Do not accept as creditworthy any general evaluation points related to the methodology used in general (e.g. use of the self-report method etc)

<u>Examiner’s Comments</u>

Most candidates were able to evaluate the population validity of the research presented using two or more relevant points. However, some used points that were either inappropriate, or not true. For example, referring to demand characteristics affecting how honestly participants would complete the questionnaire, and claiming that random sampling had been used (when there was no reference to this in the research scenario presented). The best responses made a relevant point first (eg overall sample size) then went on to discuss the implications of this in the context of the research undertaken.</td></tr>
<tr><td></td><td></td><td></td><td>Total</td><td>6</td><td></td></tr>
</table>

82			A	**1**	**Examiner's Comments** This question required a good understanding of the process of conducting inferential statistical tests and shows the importance of covering this fully when preparing candidates for the exam. Many struggled with this question
			Total	**1**	
83			A	**1**	**Examiner's Comments** Responses to this question revealed that some candidates did not know what secondary data referred to.
			Total	**1**	
84			B	**1**	**Examiner's Comments** The term 'significant result' should have been familiar to candidates having covered what 'probability and significance' refers to in the context of inferential statistics, but some candidates struggled with this question.
			Total	**1**	
85			A	**1**	**Examiner's Comments** Mostly correct responses here
			Total	**1**	
86			A	**1**	**Examiner's Comments** In order to do well on this question candidates needed to be aware of what was involved in different types and levels of data, which clearly some were not.
			Total	**1**	

87			D	**1**	**Examiner's Comments** This question revealed that some candidates were not familiar with the process of conducting inferential statistical tests and understanding their outcomes. It reveals the need to prepare candidates by getting them to actually undertake some calculations using these tests in order to become more conversant with them and understand what the outcomes mean.
			Total	**1**	
88			A	**1**	**Examiner's Comments** Mostly correct responses here
			Total	**1**	
89			B	**1**	**Examiner's Comments** It was pleasing that there were many correct responses to this question requiring a knowledge of standard deviation and how the preliminary step in this calculation is the variance, but also an understanding the mathematical concepts of squares and square roots.
			Total	**1**	
90			A	**1**	**Examiner's Comments** Both these questions reveal the need to be aware of research methods in the context of the core studies, and recognise how delivery of component 2 can facilitate and enhance learning of the content of component 1.
			Total	**1**	
91			Possible answers could include: clarity of behavioural categories; whether behavioural categories discussed / agreed beforehand;	**Max 6**	-Context = journey (including modes of transport – e.g. car, bus, plan train etc), and / or any relevant related behaviours from the candidates suggested behavioural categories)

			where observers are positioned; number of observers etc For each thing referred to that may influence inter-rater reliability Clear, detailed response in context Clear, detailed response but not in context **OR** attempt in context Brief and / or weak attempt to describe something that may influence inter-rater reliability (whether in context or not) The candidate has not provided any creditworthy information	**3** **2** **1** **0**	-Accept answers that relate to things that would make inter-rater reliability low or high (or both) -Remember, inter-rater reliability is not the extent to which if the study was repeated the same findings would be obtained. It is concerned with the ability of two or more observers to look out for and record the same behaviours in the same way. -Nothing creditworthy for ref to the influence of situational variables, such as the time of day, length or journey or the number or types of people on the train in different carriages etc **Examiner's Comments** Candidates should be reminded that when the question rubric specifically asks for two things, only the first two things presented can receive credit. Responses here typically included reference to problems that different individual observers may have in interpreting behavioural categories in the same way as each other, the benefit of using of standardised pre-determined behavioural categories within a structured observation and the training of observers in the use and interpretation of such categories. There is still some confusion amongst candidates between test-retest reliability in general and inter-rater reliability, with some incorrectly making reference to the difficulty of obtaining the same findings if the study was to be repeated on a different day at a different time or on a different mode of transport. Inter-rater reliability is about the ability of two or more observers to look out for and record the same behaviours in the same way. It would not be expected (indeed highly unusual) if the same findings were to be obtained if the study was to be repeated.
			Total	**6**	
92			Under the new (2015) BPS (British psychological Society) ethical considerations 'responsibility' in general refers to the general care of participants. More specifically it includes: **protection of participants**	**Max 6**	-Context = journey (including modes of transport – e.g. car, bus, plan train etc), and / or any relevant related behaviours from the candidates suggested behavioural categories) -Both points could be about (different)

			(ensuring participants are not harmed (mentally or physically); and **debrief** (informing participants about how and why the research was done afterwards (and sharing findings, need for deception if any used etc) Clear, detailed description in context Clear, detailed description but not in context **OR** attempt in context Brief and / or weak attempt to describe how to address the ethical consideration of 'responsibility' (whether in context or not) The candidate has not provided any creditworthy information	**3** **2** **1** **0**	aspects of the same thing – e.g. two points about protection for harm in some way -Responses referring to ethical considerations in general (e.g. deception, or keeping data anonymous etc) without explicit reference to how this could relate to 'responsibility' (e.g. protection of participants in some way) are not creditworthy -Some example 1 mark responses … *Debrief participants* *Ask if pps happy for their data to be used* -If a candidate just identifies 'protection' and / or and 'debrief' then award 1 mark **Examiner's Comments** Many candidates struggled with this question and responded with reference to ethical issues in general. The best responses were characterised by discussing how participants could be protected from harm (eg not made to feel anxious by making it obvious they were being observed) and providing details of an appropriate debrief (eg explaining to participants at the end that they had been observed, what the purpose of this was for and providing an opportunity to address any concerns they may have, such as worrying about realising they had been watched etc). It was not enough just to refer to harm and the use of a debrief by name only. This question revealed which candidates were familiar with the 'new' BPS guidelines on ethical considerations in research and the different categories now used.
			Total	**6**	
93			Answer = **3:1** First, the number of people represented by 37.5% (who regarded 'personality' as most important for love) and 12.5% (who regarded 'wealth' as most important for love)	**Max 4**	-Zero if ratio presented the wrong way round (1:3) **Examiner's Comments** Most candidates were able to successfully work out the respective

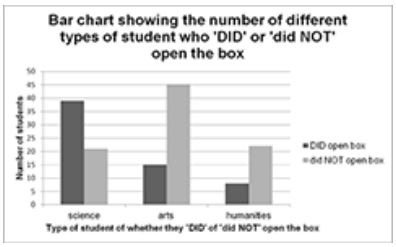
			needs to be calculated. This requires a knowledge of percentages. 37.5% of 40 = 37.5/100 x 40 = 15 12.5% of 40 = 12.5/100 x40 = 5 So the ratio is 15:5, which can be simplified to 3:1	**4**	ratios in response to this questions and show some appropriate workings.
			Correctly stated ratio with all workings shown clearly	**3**	
			Correctly stated ratio but with some or all workings missing **OR** correctly stated ratio, but not simplified (i.e. left at 15:5)	**2** **1**	
			Some correct workings shown	**0**	
			Some correct workings shown but also with some that are incorrect		
			The candidate has not provided any creditworthy information		
			Total	**4**	
94			D	**1**	**Examiner's Comments** Mostly correct responses here
			Total	**1**	
95			B	**1**	**Examiner's Comments** There was a varied response to this question, indicating there is still some confusion amongst some candidates about the different types of reliability.
			Total	**1**	
96			Possible answers could include: type of students in the sample (e.g. if from different faculties etc); number of students in the sample; instruction of what not to do (e.g. if changed from simply being told 'not to open a box' etc); location of where the study was conducted; presence of other people in the room at the same time etc	**Max 6**	-Context = student(s) (science, arts, humanities), did open, did not open box
			For each thing referred to that may affect reliability …	**3**	
			Clear, detailed response in context	**2**	
			Clear, detailed response but not in context **OR** attempt in context	**1**	

			Attempt, whether in context or not **OR** simply stating a finding The candidate has not provided any creditworthy information	**0**	
			Total	**6**	
97	a		B	**1**	
	b		D	**1**	
			Total	**2**	
98			C	**1**	
			Total	**1**	
99			A	**1**	
			Total	**1**	
100			A	**1**	
			Total	**1**	
101			B	**1**	
			Total	**1**	
102			A	**1**	
			Total	**1**	
103			A	**1**	
			Total	**1**	
104			D	**1**	
			Total	**1**	
105	a		C	**1**	
	b		B	**1**	
			Total	**2**	
106			A	**1**	
			Total	**1**	
107	a		Enables more statistical analyses, enables use in correlation analysis, can assess degree / extent of sleep loss of concentration etc Clear description of strength in context Attempt to describe strength in context **OR** Clear description of	**Max 3** **3** **2** **1**	-Context = ‘sleep’ and ‘concentration’

			strength but not in context	**0**	
			Attempt to describe strength (whether in context or not)		
			The candidate has not provided any creditworthy information		
	b		Lack of understanding of reasons / meanings, e.g. for why / how concentration is influenced by sleep etc	**Max 3**	-Context = 'sleep' and 'concentration'.
			Clear description of weakness in context	**3**	
			Attempt to describe weakness in context **OR** Clear description of weakness but not in context	**2**	
			Attempt to describe strength (whether in context or not)	**1**	
			The candidate has not provided any creditworthy information	**0**	
			Total	**6**	
108			For example, through the use of a self-report about the experiences of carrying out the task assessing concentration, or questions about the quality of sleep had the previous night etc (in a written questionnaire or interview)	**Max 3**	-Context = 'sleep' and 'concentration'
			Clear outline how qualitative data could be obtained in context	**3**	
			Attempt to outline how qualitative data could be obtained **OR** Clear outline of how qualitative data could be obtained	**2**	
			Attempt to outline how qualitative data could be obtained (whether in context or not)	**1**	
			The candidate has not provided any creditworthy information	**0**	
			Total	**3**	
109	a		Could consider here the extent to which the sample used is representative or not (population validity) or the extent to which the task set to assess concentration relates to everyday life examples of	**Max 3**	-Context = 'sleep' and 'concentration' **External validity** is the extent to which the findings can be applied (or generalized) to other people **(population validity)** or situations /

			how we concentrate (ecological validity) etc. More generally, the extent to which the research can be applied beyond the confines of the parameters of the research situation / context. Clear outline in context Attempt to outline how the study could be considered to be high in external validity in context **OR** Clear outline of how the study could be considered to be high in external validity but not in context Attempt to outline how the study could be considered to be high in external validity (whether in context or not) The candidate has not provided any creditworthy information	**3** **2** **1** **0**	circumstances (**ecological validity** – the extent to which the research relates to everyday life). Having a representative sample and conducting the research in real life settings helps external validity.
	b		Could consider here the extent to which the sample used is representative or not (population validity) or the extent to which the task set to assess concentration relates to everyday life examples of how we concentrate (ecological validity) etc. More generally, the extent to which the research can be applied beyond the confines of the parameters of the research situation / context. Clear outline in context Attempt to outline how the study could be considered to be low in external validity in context **OR** Clear outline of how the study could be considered to be low in external validity but not in context Attempt to outline how the study could be considered to be low in external validity (whether in context or not) The candidate has not provided any creditworthy information	**Max 3** **3** **2** **1** **0**	-Context = ‘sleep’ and ‘concentration’ **External validity** is the extent to which the findings can be applied (or generalized) to other people (**population validity**) or situations / circumstances (**ecological validity** – the extent to which the research relates to everyday life). Having a representative sample and conducting the research in real life settings helps external validity.
			Total	**6**	
110			Demand characteristic refers to any change in participants’ behaviour / response as a consequence of some	**Max 2**	-Context = ‘sleep’ and ‘concentration’

			aspect of the realisation / knowledge that they are being studied / monitored. Examples here(depending on how the candidate has suggested planning and conducting the research in response to earlier questions) could include such things as giving inaccurate details about the amount of sleep they have had or deliberately performing better or worse in the concentration task, or simply performing worse because of anxiety about doing well etc. Clear explanation in context Attempt to explain what the term demand characteristic refers to in context **OR** clear explanation but not in context The candidate has not provided any creditworthy information	**2** **1** **0**	
			Total	**2**	
111			Percentage of science students who DID open the box 39/60 x 100 = 65% Percentage of science students who DID NOT open the box 21/60 x 100 = 35% For each calculation … Correct calculation / answer with workings shown Correct answer without workings The candidate has not provided any creditworthy information	**Max 4** **2** **1** **0**	
			Total	**4**	
112			45:15 Which can be simplified to 3:1 For each calculation … Ratio correctly calculated and presented in its simplest form (3:1) Ratio correctly calculated but not presented in its simplest form (45:15, or 9:3) The candidate has not provided any creditworthy information	**Max 2** **2** **1** **0**	
			Total	**2**	

113			For example (but please accept other, appropriate versions) …  All features included 3 features included 2 features included 1 feature included The candidate has not provided any creditworthy information	**Max 4** **4** **3** **2** **1** **0**	-Context = student(s) (science, arts, humanities), did open, did not open box
			Total	**4**	
114			Conclusions could include: overall more students did not open the box, so were perhaps obedient (perhaps being a student means you are more likely to follow instructions / orders?); Science students were the most curious and opened the box, perhaps because science students are more inquisitive and like to find things out; Arts students were the least likely to open the box, perhaps because they are used to doing what they are told? etc For each conclusion … Clear, detailed response in context Clear, detailed response but not in context **OR** attempt in context Attempt, whether in context or not **OR** simply stating a finding The candidate has not provided any creditworthy information	**Max 6** **3** **2** **1** **0**	-Context = student(s) (science, arts, humanities), did open, did not open box -For max marks must be a conclusion
			Total	**6**	
115			Possible answers could be based on the ethical issue of consent for real purpose of the study, privacy (observing without knowledge), and withdrawal (as true purpose of being in the waiting room was not disclosed) etc	**Max 3** **3**	-Context = student(s) (science, arts, humanities), did open, did not open box

			Clear outline of ethical issue in context Attempt to outline ethical issue in context **OR** Clear outline of ethical issue but not in context Attempt to outline ethical issue (whether in context or not) The candidate has not provided any creditworthy information	**2** **1** **0**	
			Total	**3**	
116			**Which is an example of interval level data?** 1 mark for A – the mass, in grams, of the brain of an individual with schizophrenia.	**1** **AO2 f (m)**	
			Total	**1**	
117			**Identify THREE ethical issues that would need to be considered when carrying out this observation.** 1 mark for each feasible ethical issue (even if brief) e.g. deception, (lack of) right to withdraw, need for debriefing, confidentiality, (lack of) consent, etc. Candidates can outline the issue (in the context of the study) without explicitly naming it e.g. respect people's privacy when they are changing). Other appropriate responses should be credited.	**3** **AO2 c**	The issue does not have to be specific to covert observations but should apply to the method. This means most issues are creditworthy but guard against those that are not e.g. use of non–human animals. Be careful not to credit issues that overlap or make similar points e.g. causing distress and causing discomfort.
			Total	**3**	
118	a		**Identify TWO findings from the bar chart.** 1 mark for recognising that the image was mainly perceived as a monkey or teapot compared to anything else. 1 mark for recognising that the image was perceived as much as a monkey as a teapot. Other appropriate responses should be credited.	**2** **AO3 2a (m)**	
	b		**Explain why a bar chart is appropriate for presenting this data.**	**2** **AO3 2b (m)**	

			1 mark for explaining that data is discrete / categorical/nominal. 1 mark for explaining that the graph allows for comparison / looking for a difference. Other appropriate responses should be credited.		
	c		**Calculate the percentage number of times that the image was identified as neither a monkey nor a teapot. Show your workings.** 1 AO1 mark for 10% (or 10) 1 AO2 mark for workings, i.e. 2/(9+9+2)=0.1 Other appropriate responses should be credited.	**2** **1 AO1 1b (m)** **1 AO2 h (m)**	
			Total	**6**	
119	a		**Identify and simplify the ratio of the number of participants who perceived a monkey in the first condition and the number who perceived a monkey in the second condition.** 1 AO1 mark for identifying the ratio 15:5 1 AO2 mark for simplifying the ratio to 3:1	**2** **1 AO1 lb (m)** **1 AO2 h (m)**	The ratio can be identified through description i.e. the ratio is 15 participants to 5 participants but the simplified ratio must be in standard format.
	b		**Identify and simplify the ratio of the number of participants who perceived a teapot in the first condition and the number who perceived a monkey in the second condition.** 1 AO1 mark for identifying the ratio 10:12 1 AO2 mark for simplifying the ratio to 5:6	**2** **1 AO1 lb (m)** **1 AO2 h (m)**	The ratio can be identified through description i.e. the ratio is 10 participants to 12 participants but the simplified ratio must be in standard format.
			Total	**4**	
120			**The psychologist analysed the data using the Chi Squared test. Give TWO reasons for this choice of test with reference to the study.** Up to 2 marks for any of the following reasons;	**4** **2 AO1 1b (m)** **2 AO2 b (m)**	Do not credit parametric assumptions as data is only nominal in the first place.

			• experiment investigated a difference or association • design was unrelated • data was nominal. A further 2 marks for relating the chosen criteria to features of the study (1 mark for each criteria) e.g. 'the experiment investigated a difference (1) between perception of an image depending on how it had been primed (1)'. e.g. 'the design was unrelated (1) as each condition contained a different set of participants (1).' e.g. 'the data was nominal (1) as responses were categorised as either 'monkey' or 'teapot' (1).' Other appropriate responses should be credited.		
			Total	**4**	
121			**Explain how the psychologist would determine the appropriate degrees of freedom (df) for this test.** 1 AO1 mark for stating that degrees of freedom can be calculated as (rows-1)(column-1). 1 AO2 mark for stating that in this study, there is 1 degree of freedom.	**2** **1 AO1 1b (m)** **1 AO2 h (m)**	
			Total	**2**	
122			**Using the above critical values, explain whether the psychologist has found a significant difference or not.** 1 mark for explaining the results are significant at p=0.05 1 mark for explaining why, i.e. observed value is greater than critical value. 1 mark for explaining that 0.05 is the accepted level of significance for analysis. 1 mark for explaining that the results were not significant at the other more stringent levels of significance.	**4** **AO3 1b (m)**	

			Total	**4**	
123			(From the scatter diagram) **Which is the best estimate of the correlation coefficient for the above data?** 1 mark for C – 0.3.	**1** **AO2 h (m)**	
			Total	**1**	
124			**What is a weakness of using a mode as a measure of central tendency?** 1 mark for D – it relies on a score occurring more than once.	**1** **AO1 1b (m)**	
			Total	**1**	
125			**Which is a requirement of a parametric test?** 1 mark for D – standard deviations are not significantly different.	**1** **AO1 1b (m)**	
			Total	**1**	
126			**In Bandura's (1961) Bobo doll study, the participants were pre-tested to assess their aggression levels. What was the main purpose of his procedure?** 1 mark for A – to allow for a matched pairs design.	**1** **AO1 1b**	
			Total	**1**	
127			**Look at the following data set from a condition where participants were timed (in seconds) completing a task in a crisis situation.** **{36 45 51 67 54 19 50 45 27 76 54 45}** **What is the range of this data set?** 1 mark for C – 58.	**1** **AO2b (m)**	
			Total	**1**	
128			**What is meant by the term 'socially desirable responses' in psychological research?**	**1**	

			1 mark for D – responses which the participants think they ought to give even if they are not true.	**AO1 1a (r)**	
			Total	**1**	
129			**Which is an example of qualitative data?** 1 mark for A – the diary entries of six patients suffering from schizophrenia.	**1** **AO2 e (m)**	
			Total	**1**	
130			**What is the probability of a significant result occurring by chance where the significance level is p≤0.025?** 1 mark for A – 2.5% or less.	**1** **AO1 1b (m)**	
			Total	**1**	
131			**Which inferential test should a researcher use to decide whether a correlation is significant?** 1 mark for C – Spearman's Rho test.	**1** **AO1 1b (m)**	
			Total	**1**	
132			Advantage: can use more descriptive statistics (e.g. able to work out differences in performance in when standing up compared to sitting down) Disadvantage: doesn't inform us why standing up or sitting down affects performance in test scores Accept any other creditworthy advantage or disadvantage 2 marks for advantage, 2 marks for weakness Clear outline of advantage / disadvantage in context Attempt to outline advantage / disadvantage whether in context or not **OR** clear outline of advantage / disadvantage but not in context The candidate has not provided any creditworthy information	**Max 4** **2** **1** **0**	Context = 'stand' / 'standing up', 'maths test', 'concentration' etc **Examiner's Comments** Most candidates were able to provide appropriate examples of relevant advantages and disadvantages in response to this question (with many stating that quantitative data is easier / more flexible to analyse but does not inform us about reasons why the data occurred as it did). However, some candidates lost marks by not contextualising their answer.
			Total	**4**	
133	a		D	**1**	**Examiner's Comments**

					Mostly correct responses
	b		A	**1**	**Examiner's Comments** Mostly correct responses, but occasionally some candidates chose option C (12) indicating they would use the total sample size (rather than sample size per condition)
	c		B	**1**	**Examiner's Comments** Mostly correct responses
	d		B	**1**	**Examiner's Comments** Some candidates struggled with this question and it shows the importance of preparing to answer questions requiring knowledge of the criteria for the selection of an inferential statistical test.
			Total	**4**	
134			B	**1**	**Examiner's Comments** Mostly correct responses, although some did choose option D (>>) instead
			Total	**1**	
135			B	**1**	**Examiner's Comments** Many candidates struggled with this question and it shows the need to be knowledgeable about the many different types of reliability. One suggestion is to incorporate this in to the teaching of the core studies and the practical activities undertaken.
			Total	**1**	
136	a		A	**1**	**Examiner's Comments**

					Some candidates incorrectly chose option C (positively skewed)
	b		A	**1**	**Examiner's Comments** Responses to the previous linked question influenced performance here, with candidates who chose the wrong option for the previous question usually selecting a wrong choice option here also.
			Total	**2**	
137			C	**1**	**Examiner's Comments** Mostly correct responses
			Total	**1**	
138			B	**1**	**Examiner's Comments** Some candidates incorrectly chose option A (25), perhaps because 25 represents a quarter of 100
			Total	**1**	
139	a		D	**1**	**Examiner's Comments** Mostly correct responses
	b		C	**1**	**Examiner's Comments** Mostly correct responses
			Total	**2**	
140			B	**1**	**Examiner's Comments** Mostly correct responses
			Total	**1**	
141			B	**1**	**Examiner's Comments**

					This proved to be quite challenging for candidates and once again shows the importance of the need to be knowledgeable about the many different types of reliability.
			Total	**1**	
142			**2 marks for each finding** Answers could include: maths scores were generally higher when stood up than sat down; there was more variation in the maths scores when sat down; the maximum score was 20 obtained by both someone in the standing up condition and the sitting down condition etc etc. Finding clearly identified in context Attempt to identify finding The candidate has not provided any creditworthy information	**Max 4** **2** **1** **0**	Context = 'stand' / 'standing up', 'maths test', 'concentration' etc Creditworthy descriptive statistics Stood up: Sat down: Mean = 17.83 Mean = 12.17 Median = 18 Median = 13 Mode = 18 Range = 20-4 = 16 Range = 20-15 = 5 (or +1 =17) (or +1 = 6) **Examiner's Comments** Most candidates were able to select and present two findings from the data table provided.
			Total	**4**	
143			D	1	**Examiner's Comments** There was some confusion here about how to interpret the symbols
			Total	**1**	
144			C	1	**Examiner's Comments** This question revealed that some candidates were not familiar with the concept of internal reliability and how it can be checked
			Total	**1**	
145			A	1	**Examiner's Comments** Mostly correct responses
			Total	**1**	
146			C	1	**Examiner's Comments** Responses to this question indicated some candidates were not familiar with the term 'descriptive statistic'
			Total	**1**	
147			C	1	**Examiner's Comments** This question demonstrated the need

<table>
<tr><td></td><td></td><td></td><td></td><td></td><td>to be able to identify and interpret graphical displays</td></tr>
<tr><td></td><td></td><td></td><td>Total</td><td>1</td><td></td></tr>
<tr><td>148</td><td></td><td></td><td>A</td><td>1</td><td><u>Examiner’s Comments</u>
Responses to this question indicated some candidates were not familiar with the term ‘measure of dispersion’</td></tr>
<tr><td></td><td></td><td></td><td>Total</td><td>1</td><td></td></tr>
<tr><td>149</td><td></td><td></td><td>Likely answers: the ability of two or more observers to look out for and record the behaviour of couples (mimicking each other’s behaviour) in the same way.

Clear explanation of what inter-rater reliability refers to in context

Attempt to explain what inter-rater reliability refers to in context OR Clear explanation of what inter-rater reliability refers to but not in context

Brief and / or weak attempt to explain what inter-rater reliability refers to (whether in context or not)

The candidate has not provided any creditworthy information</td><td>Max 3

3

2

1

0</td><td>-Context = bar / drink, couples
Also accept (appropriate) behavioural categories as context?

-Inter-rater reliability is not simply about being able to replicate the study / findings

-Reference to clarity about what the behavioural categories are and how to interpret them is creditworthy to demonstrate understanding of what inter-rater reliability refers to

-Cap at 1 mark if simply describing a way to check inter-rater reliability

<u>Examiner’s Comments</u>
Many candidates struggled to explain the concept of inter-rater reliability, withmany simply stating that it involved repeating the study and getting the same findings.
Higher scoring candidates on this question used examples to elaborate on their answer and responded in context.</td></tr>
<tr><td></td><td></td><td></td><td>Total</td><td>3</td><td></td></tr>
<tr><td>150</td><td></td><td></td><td>= 3:2
Examples of workings …
120:80, 60:40, 3:2
or simply stating divide each side by 40</td><td>Max 2

2

1</td><td>Zero if answer not simplified

<u>Examiner’s Comments</u>
Most candidates demonstrated a good understanding of ratios, but</td></tr>
</table>

			Correct answer with some appropriate workings clearly shown Correct answer but no workings (or incorrect workings) **OR** some appropriate workings shown but no final answer stated The candidate has not provided any creditworthy information	**0**	occasionally some did not provide any workings.
			Total	**2**	