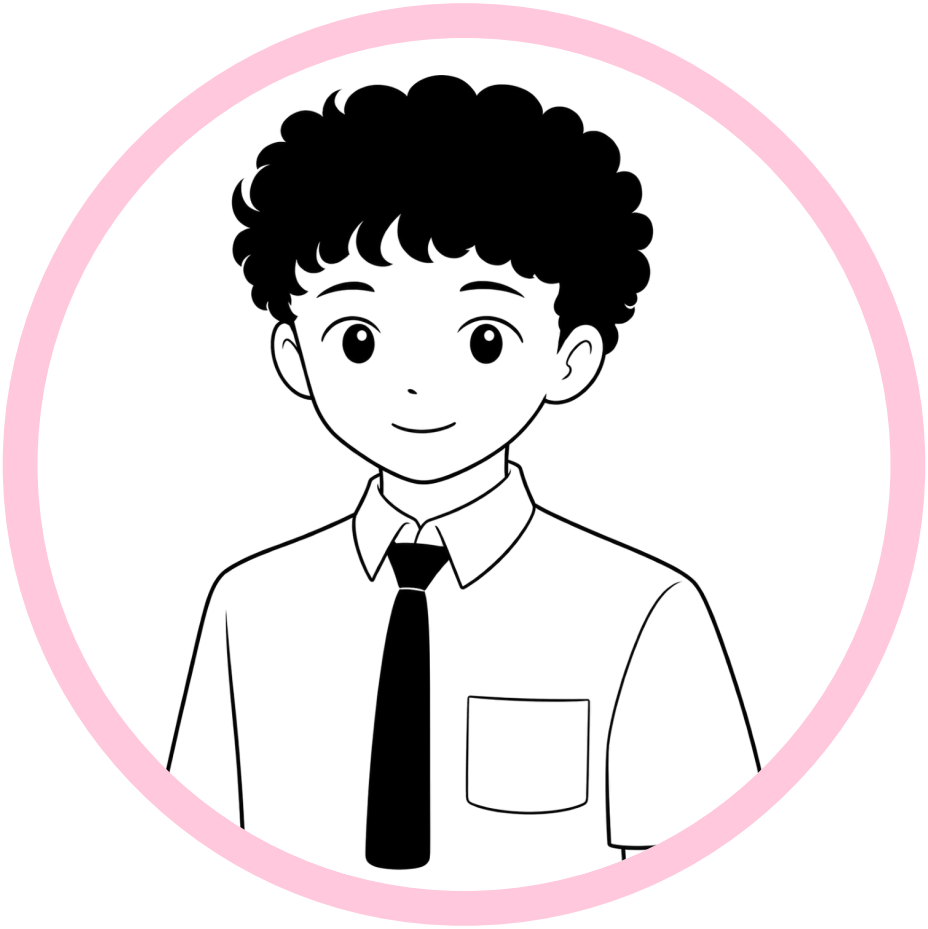
OCR A-LEVEL PSYCHOLOGY · H567

PSYCHOLOGY



EXAM QUESTION PACK

1.3 DATA RECORDING, ANALYSIS AND PRESENTATION

SCORE	PERCENTAGE
/850	%

STEMUP

These questions belong to OCR, not to us.

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

What we do not own

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

Where to find the originals

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

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

1. Which measure of dispersion only uses the lowest and highest values in a data set?

A range
B standard deviation
C variance
D none of the above

Your answer ☐ **[1]**

2(a). A psychologist wanted to see how age might affect memory. Group A were aged 20-30 years and Group B were aged 60-70 years. All participants were shown the same 20 objects on a screen for 10 seconds and then asked to immediately recall as many of the objects as they could.?

Group A 20-30 years	Number of items recalled	Group B 60-70 years	Number of items recalled
1	17	1	14
2	13	2	12
3	17	3	5
4	11	4	13
5	16	5	14
6	12	6	13
7	19	7	13

Which of these is the mean score of the 20-30 year old participants?

A 12
B 15
C 16
D 17

Your answer ☐ **[1]**

(b). Which of these results from the table above would be classed as anomalous data?

A 5
B 13
C 17
D 19

Your answer ☐ **[1]**

3.

Slipping up can be good

Psychologists have investigated many ways that could improve how children learn. However, most of these have concentrated on studying the effect of different teaching styles, rather than focusing on the child themself. A psychologist taking a more child-focused approach wants to study if being more relaxed affects concentration levels. They want to investigate if young children can concentrate better when wearing comfortable slippers on their feet compared to wearing shoes. The study is to be conducted in one large primary school with 240 children on the register.

Outline **one** way you could uphold the ethical consideration of respect in this study.

[2]

4.

Slipping up can be good

Psychologists have investigated many ways that could improve how children learn. However, most of these have concentrated on studying the effect of different teaching styles, rather than focusing on the child themself. A psychologist taking a more child-focused approach wants to study if being more relaxed affects concentration levels. They want to investigate if young children can concentrate better when wearing comfortable slippers on their feet compared to wearing shoes. The study is to be conducted in one large primary school with 240 children on the register.

Explain **two** factors that could affect the external validity of this study.

1

2

[6]

5.

Sounds familiar

Memory can be influenced by many different things. However, there are techniques which we can use to improve our memory. One such technique involves reading aloud the information we want to remember. A psychologist investigated this by giving participants a set of 30 words to try and remember. Six participants studied the words in silence. A different group of six participants were instructed to read the words aloud when trying to learn them. The data collected is presented in the table below.

Number of words correctly recalled (max. 30)					
Reading aloud group			**Silent study group**		
Participant	**Score**	**Rank**	**Participant**	**Score**	**Rank**
a	24	9.5	a	14	2
b	27	11	b	16	3
c	21	6	c	12	1
d	20	5	d	29	12
e	23	8	e	17	4
f	22	7	f	24	9.5
$n_1 = 6$			$n_2 = 6$		

Outline **one** conclusion that can be made from the raw data presented in this table.

[3]

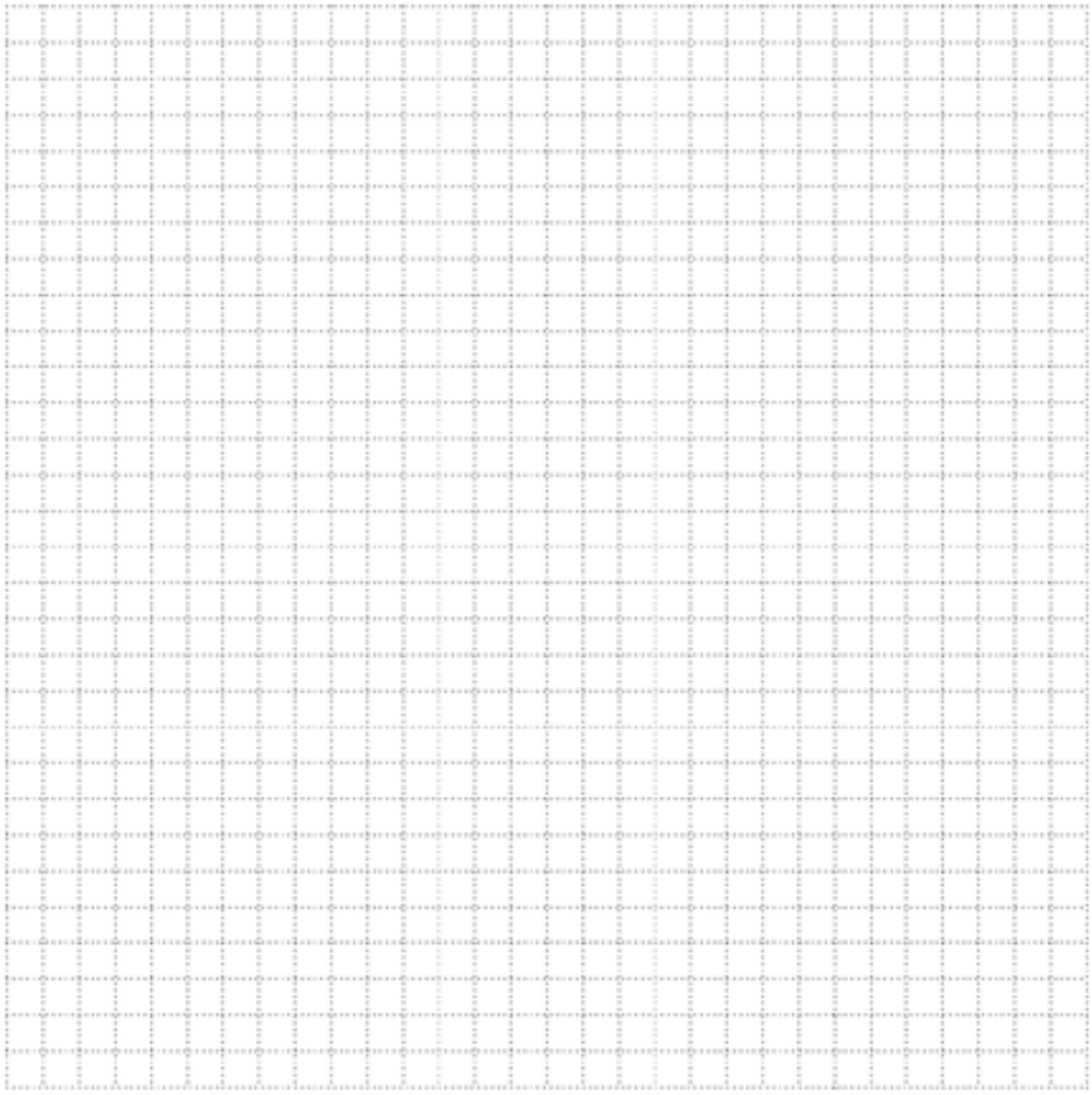
6.

Sounds familiar

Memory can be influenced by many different things. However, there are techniques which we can use to improve our memory. One such technique involves reading aloud the information we want to remember. A psychologist investigated this by giving participants a set of 30 words to try and remember. Six participants studied the words in silence. A different group of six participants were instructed to read the words aloud when trying to learn them. The data collected is presented in the table below.

Number of words correctly recalled (max. 30)					
Reading aloud group			**Silent study group**		
Participant	**Score**	**Rank**	**Participant**	**Score**	**Rank**
a	24	9.5	a	14	2
b	27	11	b	16	3
c	21	6	c	12	1
d	20	5	d	29	12
e	23	8	e	17	4
f	22	7	f	24	9.5
$n_1 = 6$			$n_2 = 6$		

Draw a fully labelled bar chart showing the mean number of words recalled in each condition. Plot the means to two significant figures.



[4]

7.

Sounds familiar

Memory can be influenced by many different things. However, there are techniques which we can use to improve our memory. One such technique involves reading aloud the information we want to remember. A psychologist investigated this by giving participants a set of 30 words to try and remember. Six participants studied the words in silence. A different group of six participants were instructed to read the words aloud when trying to learn them. The data collected is presented in the table below.

Number of words correctly recalled (max. 30)					
Reading aloud group			**Silent study group**		
Participant	**Score**	**Rank**	**Participant**	**Score**	**Rank**
a	24	9.5	a	14	2
b	27	11	b	16	3
c	21	6	c	12	1
d	20	5	d	29	12
e	23	8	e	17	4
f	22	7	f	24	9.5
$n_1 = 6$			$n_2 = 6$		

Give **one** reason why the Mann-Whitney U test is the appropriate inferential test to use to analyse the data from this study.

[2]

8.

Sounds familiar

Memory can be influenced by many different things. However, there are techniques which we can use to improve our memory. One such technique involves reading aloud the information we want to remember. A psychologist investigated this by giving participants a set of 30 words to try and remember. Six participants studied the words in silence. A different group of six participants were instructed to read the words aloud when trying to learn them. The data collected is presented in the table below.

Number of words correctly recalled (max. 30)					
Reading aloud group			**Silent study group**		
Participant	**Score**	**Rank**	**Participant**	**Score**	**Rank**
a	24	9.5	a	14	2
b	27	11	b	16	3
c	21	6	c	12	1
d	20	5	d	29	12
e	23	8	e	17	4
f	22	7	f	24	9.5
$n_1 = 6$			$n_2 = 6$		

Before using the formula for the Mann-Whitney U test, the data obtained must be ranked. In the results, two participants have the same score of 24. Explain how this is dealt with when ranking the data.

[3]

9(a).

Sounds familiar

Memory can be influenced by many different things. However, there are techniques which we can use to improve our memory. One such technique involves reading aloud the information we want to remember. A psychologist investigated this by giving participants a set of 30 words to try and remember. Six participants studied the words in silence. A different group of six participants were instructed to read the words aloud when trying to learn them. The data collected is presented in the table below.

Number of words correctly recalled (max. 30)					
Reading aloud group			**Silent study group**		
Participant	**Score**	**Rank**	**Participant**	**Score**	**Rank**
a	24	9.5	a	14	2
b	27	11	b	16	3
c	21	6	c	12	1
d	20	5	d	29	12
e	23	8	e	17	4
f	22	7	f	24	9.5
$n_1 = 6$			$n_2 = 6$		

Calculate the U value for the Mann-Whitney U test for the data collected in this study. Show your workings. You may use the formula presented below.

U = the smaller of U_1 and U_2

Where U_1 is ...

$$U_1 = R_1 - \frac{n_1(n_1 + 1)}{2}$$

and U_2 is ...

$$U_2 = R_2 - \frac{n_2(n_2 + 1)}{2}$$

[5]

(b). How is the critical value used to determine if the findings are statistically significant?

[1]

10.

Sounds familiar

Memory can be influenced by many different things. However, there are techniques which we can use to improve our memory. One such technique involves reading aloud the information we want to remember. A psychologist investigated this by giving participants a set of 30 words to try and remember. Six participants studied the words in silence. A different group of six participants were instructed to read the words aloud when trying to learn them. The data collected is presented in the table below.

Number of words correctly recalled (max. 30)					
Reading aloud group			**Silent study group**		
Participant	**Score**	**Rank**	**Participant**	**Score**	**Rank**
a	24	9.5	a	14	2
b	27	11	b	16	3
c	21	6	c	12	1
d	20	5	d	29	12
e	23	8	e	17	4
f	22	7	f	24	9.5
$n_1 = 6$			$n_2 = 6$		

Explain what $p > 0.05$ would mean if it appeared as part of the significance statement when reporting the findings from this study.

[3]

11(a).

Sounds familiar

Memory can be influenced by many different things. However, there are techniques which we can use to improve our memory. One such technique involves reading aloud the information we want to remember. A psychologist investigated this by giving participants a set of 30 words to try and remember. Six participants studied the words in silence. A different group of six participants were instructed to read the words aloud when trying to learn them. The data collected is presented in the table below.

Number of words correctly recalled (max. 30)					
Reading aloud group			**Silent study group**		
Participant	**Score**	**Rank**	**Participant**	**Score**	**Rank**
a	24	9.5	a	14	2
b	27	11	b	16	3
c	21	6	c	12	1
d	20	5	d	29	12
e	23	8	e	17	4
f	22	7	f	24	9.5
$n_1 = 6$			$n_2 = 6$		

Explain **one** strength of using quantitative data in this study.

[3]

(b). Explain **one** weakness of using quantitative data in this study.

[3]

12.

Sounds familiar

Memory can be influenced by many different things. However, there are techniques which we can use to improve our memory. One such technique involves reading aloud the information we want to remember. A psychologist investigated this by giving participants a set of 30 words to try and remember. Six participants studied the words in silence. A different group of six participants were instructed to read the words aloud when trying to learn them. The data collected is presented in the table below.

Number of words correctly recalled (max. 30)					
Reading aloud group			**Silent study group**		
Participant	**Score**	**Rank**	**Participant**	**Score**	**Rank**
a	24	9.5	a	14	2
b	27	11	b	16	3
c	21	6	c	12	1
d	20	5	d	29	12
e	23	8	e	17	4
f	22	7	f	24	9.5
$n_1 = 6$			$n_2 = 6$		

This study used an independent measures design.

Outline **two** ways that an independent measures design could affect the validity of this study.

1

2

[4]

13. Which type of reliability is checked by a correlation between the data of two researchers?

A inter-rater
B replicability
C split-half
D test-retest

Your answer ☐ **[1]**

14. Which of these indicates there is less than a 2.5% chance of the null hypothesis being true?

A $p < 2.50$
B $p < 0.25$
C $p < 0.025$
D $p > 0.0025$

Your answer ☐ **[1]**

15. Which of the following is the best measure of central tendency to use if there is an extreme value in the data set?

A both the mean and the mode
B mean
C median
D mode

Your answer ☐ **[1]**

16. If the standard deviation of a data set is 9, what is the variance?

A 3
B 4.5
C 9
D 81

Your answer ☐ **[1]**

17. Which of these is true of the calculated value (observed value) in a statistical test of significance??

A it is always a whole number
B it is based on the data collected
C it is in a table of values
D it will always be a number between 1 and -1

Your answer ☐ **[1]**

18. What is 57.638 written to two significant figures?

A 57
B 57.63
C 57.64
D 58

Your answer ☐ **[1]**

19. Which of these examples of research would collect nominal data?

A a correlation study investigating the relationship between a person's height (in centimetres) and their level of confidence (on a scale of 0-to-20)
B a self-report study that asks people to rate how happy they are on a scale of 0-to-100
C an experiment that records blood pressure as a measure of anxiety to two different stressors
D an observation study which records whether customers say 'thank you' or not to the checkout assistant in a shop

Your answer ☐ **[1]**

20. Which of these is **not** a type of reliability?

A inter-rater
B peer-review
C split-half
D test-retest

Your answer ☐ **[1]**

21. Which term refers to when participants change their behaviour to present themselves in a more favourable way?

A social acceptance
B social desirability
C social inclusion
D social referencing

Your answer ☐ **[1]**

22. In Freud's (1909) study of Little Hans the data was obtained through information contained in letters written by Hans' father. Which type of data is this?

A binary
B nominal
C primary
D secondary

Your answer ☐ **[1]**

23. What is 28.6359 written to two significant figures?

A 28
B 28.64
A 28.7
D 29

Your answer ☐ **[1]**

24. If a psychologist plans to conduct a study which makes it easy for participants to stop taking part at any stage if they want to, which ethical consideration is being addressed?

A consent
B debrief
C deception
D right to withdraw

Your answer ☐ **[1]**

25.

Count your blessings

As many as 1 in 4 people experience mental health issues at some time in their life and psychologists are always looking for new ways to help people cope with such matters. One interesting idea is that by writing down things that we should be grateful for (such as having good friends, a nice sunny day and our physical well-being) we can help improve our mental health.
Psychologists want to investigate this further by using the self-report method, asking people about what kind of things they feel grateful for and how this makes them feel.

Identify **one** reason why the study might have low population validity.

---

[1]

26.

Count your blessings

As many as 1 in 4 people experience mental health issues at some time in their life and psychologists are always looking for new ways to help people cope with such matters. One interesting idea is that by writing down things that we should be grateful for (such as having good friends, a nice sunny day and our physical well-being) we can help improve our mental health.
Psychologists want to investigate this further by using the self-report method, asking people about what kind of things they feel grateful for and how this makes them feel.

Outline how you would address **two** ethical considerations in the design of this study.

1

2

---

[6]

27.

The eyes have it

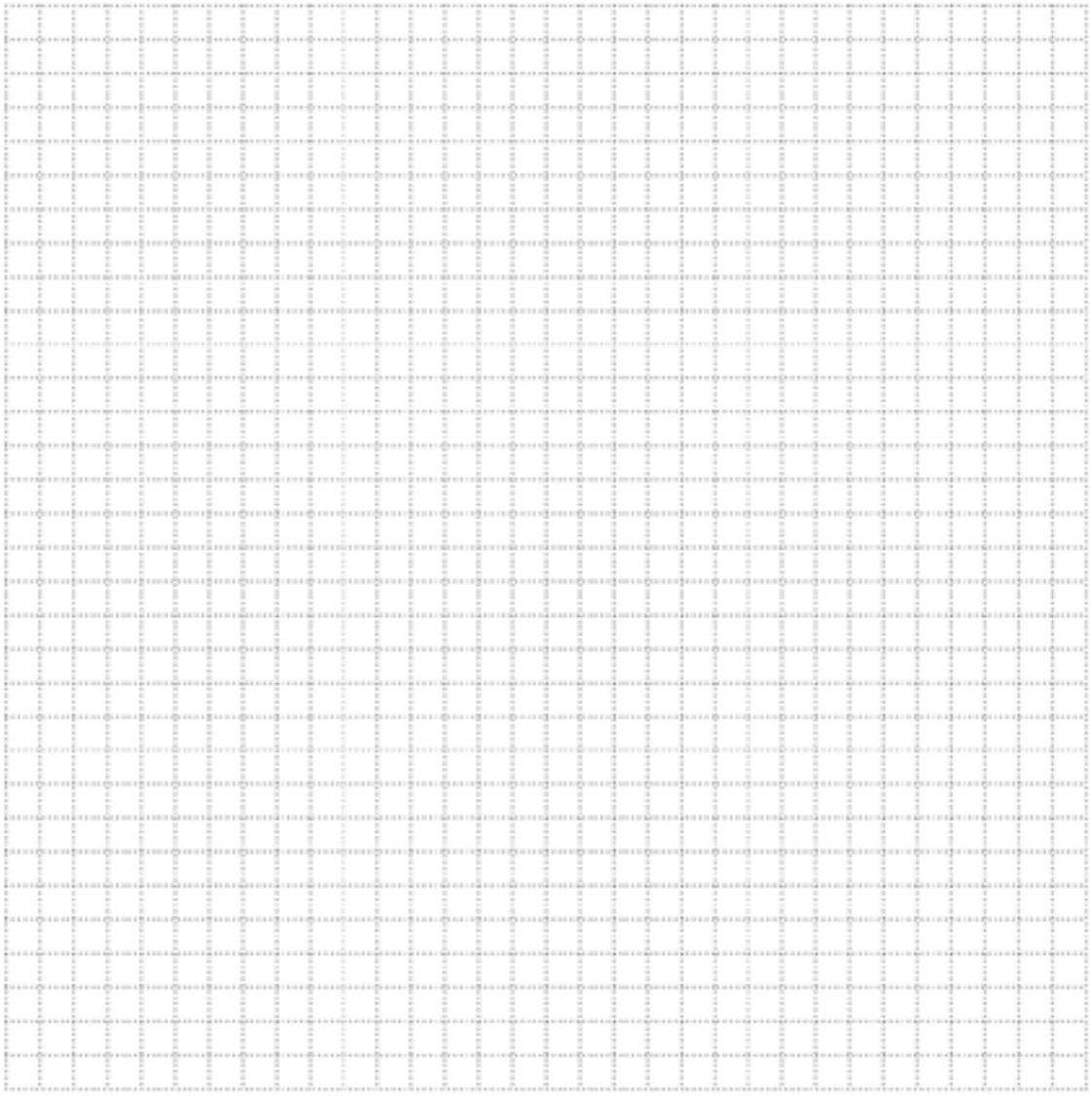
Most people agree that puppy dogs are cute, but why is this?

A psychologist investigated the possibility that the eyes play a big part. They conducted a repeated measures design experiment using a photo of the same puppy but with the eyes made to look different sizes (big or small). First participants were shown the photo with the big eyes, then a week later the same participants were shown the photo with the small eyes. On each occasion participants were asked to rate how cute the puppy was on a scale of 1 ('not very cute') to 20 ('extremely cute').

The data collected is shown in the table below.

Ratings of how cute a puppy is (1 = 'not very cute', 20 = 'extremely cute')		
Participant	**Condition (a) with big eyes**	**Condition (b) with small eyes**
a	18	7
b	14	3
c	12	14
d	17	20
e	19	9
f	16	10

Draw a fully labelled bar chart displaying the mean rating for how cute the puppy was perceived to be in each condition of this study.



[4]

28.

The eyes have it

Most people agree that puppy dogs are cute, but why is this?

A psychologist investigated the possibility that the eyes play a big part. They conducted a repeated measures design experiment using a photo of the same puppy but with the eyes made to look different sizes (big or small). First participants were shown the photo with the big eyes, then a week later the same participants were shown the photo with the small eyes. On each occasion participants were asked to rate how cute the puppy was on a scale of 1 (‘not very cute’) to 20 (‘extremely cute’).

The data collected is shown in the table below.

Ratings of how cute a puppy is (1 = 'not very cute', 20 = 'extremely cute')		
Participant	**Condition (a) with big eyes**	**Condition (b) with small eyes**
a	18	7
b	14	3
c	12	14
d	17	20
e	19	9
f	16	10

Outline **two** conclusions that can be made from the data collected in this study.

1 ----------

2 ----------

[6]

29.

The eyes have it

Most people agree that puppy dogs are cute, but why is this?

A psychologist investigated the possibility that the eyes play a big part. They conducted a repeated measures design experiment using a photo of the same puppy but with the eyes made to look different sizes (big or small). First participants were shown the photo with the big eyes, then a week later the same participants were shown the photo with the small eyes. On each occasion participants were asked to rate how cute the puppy was on a scale of 1 ('not very cute') to 20 ('extremely cute').

The data collected is shown in the table below.

Ratings of how cute a puppy is (1 = 'not very cute', 20 = 'extremely cute')		
Participant	**Condition (a) with big eyes**	**Condition (b) with small eyes**
a	18	7
b	14	3
c	12	14
d	17	20
e	19	9
f	16	10

Outline **one** strength and **one** weakness of the use of quantitative data in this study.

Strength

Weakness

---

[6]

30(a).

The eyes have it

Most people agree that puppy dogs are cute, but why is this?

A psychologist investigated the possibility that the eyes play a big part. They conducted a repeated measures design experiment using a photo of the same puppy but with the eyes made to look different sizes (big or small). First participants were shown the photo with the big eyes, then a week later the same participants were shown the photo with the small eyes. On each occasion participants were asked to rate how cute the puppy was on a scale of 1 ('not very cute') to 20 ('extremely cute').

The data collected is shown in the table below.

Ratings of how cute a puppy is (1 = 'not very cute', 20 = 'extremely cute')		
Participant	**Condition (a) with big eyes**	**Condition (b) with small eyes**
a	18	7
b	14	3
c	12	14
d	17	20
e	19	9
f	16	10

The inferential statistical test used to analyse the data collected in this study is the Wilcoxon Signed Ranks test. Explain how data is ranked for this test in this study.

[3]

(b). Using the Wilcoxon Signed Ranks test, calculate the value of T for the data collected in this study by completing the table below.

Formula / process:
T = the sum of the ranks of the differences with the least frequent sign

Ratings of how cute a puppy is (1 = 'not very cute', 20 = 'extremely cute')					
Participant	**Condition (a) with big eyes**	**Condition (b) with small eyes**	**Difference (a) - (b)**	**Ranks of differences**	**T value** (sum of the ranks of the differences with the least frequent sign)
a	18	7			
b	14	3			
c	12	14			
d	17	20			
e	19	9			
f	16	10			

[3]

31.

The eyes have it

Most people agree that puppy dogs are cute, but why is this?

A psychologist investigated the possibility that the eyes play a big part. They conducted a repeated measures design experiment using a photo of the same puppy but with the eyes made to look different sizes (big or small). First participants were shown the photo with the big eyes, then a week later the same participants were shown the photo with the small eyes. On each occasion participants were asked to rate how cute the puppy was on a scale of 1 ('not very cute') to 20 ('extremely cute').

The data collected is shown in the table below.

Ratings of how cute a puppy is (1 = 'not very cute', 20 = 'extremely cute')		
Participant	**Condition (a) with big eyes**	**Condition (b) with small eyes**
a	18	7
b	14	3
c	12	14
d	17	20
e	19	9
f	16	10

Outline **one** reason why the reliability of the data collected in this study could be low.

---

[3]

32. What is criterion validity?

A whether the measurement of a variable can be applied to the target population
B whether the measurement of a variable can be replicated
C whether the measurement of a variable can predict outcomes for similar behaviour in the future
D whether the measurement of a variable relates to everyday life

Your answer ☐ **[1]**

33. The table below shows the findings from a sample of 50 people asked about their favourite colour. What percentage of people said their favourite colour was blue?

Colour	Number of people who said this was their favourite colour
Red	17
Blue	11
Green	9
Yellow	7
Pink	6

A 2.2%
B 5%
C 11%
D 22%

Your answer ☐ **[1]**

34. Which type of statistical analysis requires data to come from conditions with similar variances?

A descriptive statistical tests
B measures of dispersion statistical tests
C non-parametric statistical tests
D parametric statistical tests

Your answer ☐ **[1]**

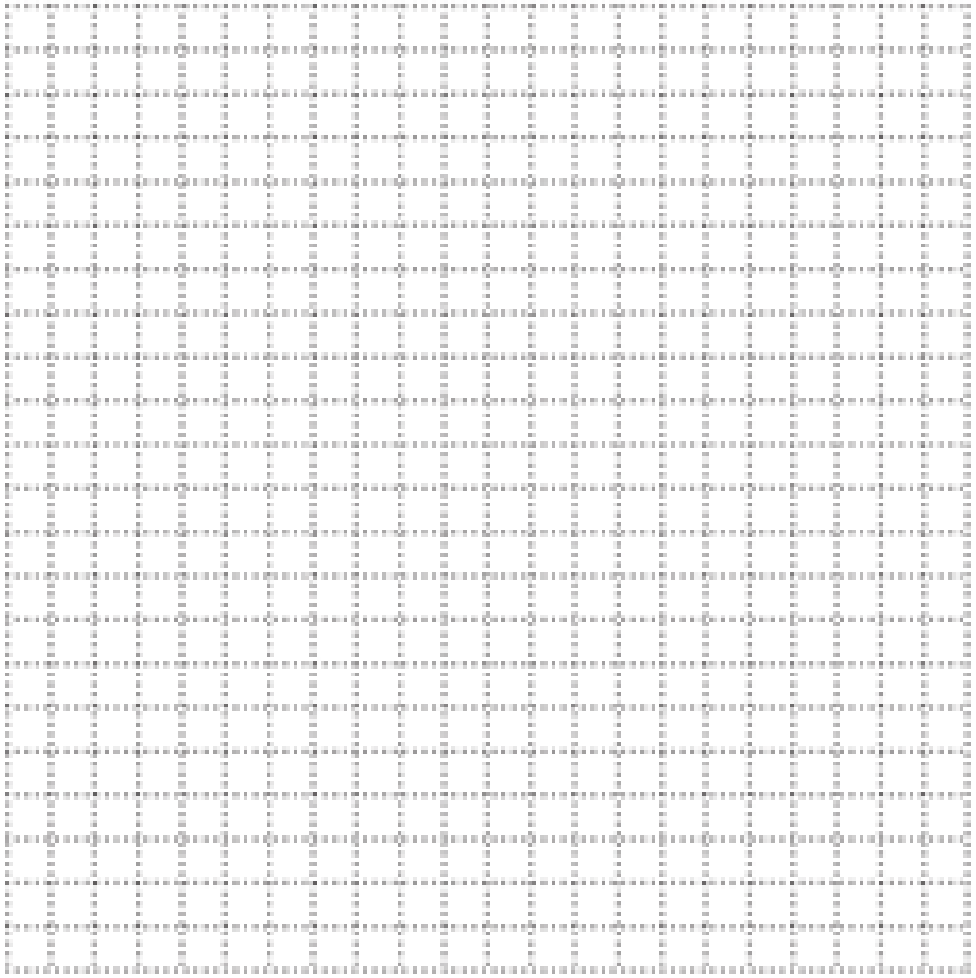
35(a).

A friendly name?

Research suggests many things can influence how friendly a person is thought to be. Personality is an obvious one and sense of humour is another. However, there are also less obvious things, such as just knowing a person's name. A psychologist investigated this using an independent measures design experiment. They compared the ratings of friendliness given to shop assistants who wore a name badge to those who did not. Some of the data collected is presented in the table below.

Ratings of friendliness (0 to 20) given to shop assistants who were wearing a name badge compared to shop assistants who were not (0 = 'not friendly at all' to 20 = 'extremely friendly')			
Wearing name badge		**Not wearing name badge**	
Rating	**Rank**	**Rating**	**Rank**
18	17	13	11
14	12	2	1
10	8	6	5
17	15	4	3
16	14	8	7
18	17	18	17
5	4	7	6
11	9	15	13
20	20	12	10
19	19	3	2

Draw a fully labelled bar chart showing the mean rating of friendliness in each condition in this study.



[4]

(b). Outline **one** conclusion that can be reached from the information in this bar chart.

[4]

36.

A friendly name?

Research suggests many things can influence how friendly a person is thought to be. Personality is an obvious one and sense of humour is another. However, there are also less obvious things, such as just knowing a person's name. A psychologist investigated this using an independent measures design experiment. They compared the ratings of friendliness given to shop assistants who wore a name badge to those who did not. Some of the data collected is presented in the table below.

Ratings of friendliness (0 to 20) given to shop assistants who were wearing a name badge compared to shop assistants who were not (0 = 'not friendly at all' to 20 = 'extremely friendly')

Wearing name badge		Not wearing name badge	
Rating	**Rank**	**Rating**	**Rank**
18	17	13	11
14	12	2	1
10	8	6	5
17	15	4	3
16	14	8	7
18	17	18	17
5	4	7	6
11	9	15	13
20	20	12	10
19	19	3	2

Explain why it may have been better to use the median to calculate the measure of central tendency in this study rather than the mean.

[2]

37(a).

A friendly name?

Research suggests many things can influence how friendly a person is thought to be. Personality is an obvious one and sense of humour is another. However, there are also less obvious things, such as just knowing a person's name. A psychologist investigated this using an independent measures design experiment. They compared the ratings of friendliness given to shop assistants who wore a name badge to those who did not. Some of the data collected is presented in the table below.

Ratings of friendliness (0 to 20) given to shop assistants who were wearing a name badge compared to shop assistants who were not (0 = 'not friendly at all' to 20 = 'extremely friendly')

Wearing name badge		Not wearing name badge	
Rating	**Rank**	**Rating**	**Rank**
18	17	13	11
14	12	2	1
10	8	6	5
17	15	4	3
16	14	8	7
18	17	18	17
5	4	7	6
11	9	15	13
20	20	12	10
19	19	3	2

Calculate the U value for the data collected in this study using the Mann-Whitney U test formula provided. Show your workings.

U = the smaller of U_1 and U_2

Where U_1 is ...

$$U_1 = R_1 - \frac{n_1(n_1 + 1)}{2}$$

and U_2 is ...

$$U_2 = R_2 - \frac{n_2(n_2 + 1)}{2}$$

[5]

(b). The table below shows critical values at the 5% level of probability for the Mann-Whitney U test. Using the table, state the critical value **and** explain how you found this.

N_a \ N_b	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
5	2	3	5	6	7	8	9	11	12	13	14	15	17	18	19	20
6		5	6	8	10	11	13	14	16	17	19	21	22	24	25	27
7			8	10	12	14	16	18	20	22	24	26	28	30	32	34
8				13	15	17	19	22	24	26	29	31	34	36	38	41
9					17	20	23	26	28	31	34	37	39	42	45	48
10						23	26	29	33	36	39	42	45	48	52	55
11							30	33	37	40	44	47	51	55	58	62
12								37	41	45	49	53	57	61	65	69
13									45	50	54	59	63	67	72	76
14										55	59	64	67	74	78	83
15											64	70	75	80	85	90
16												75	81	86	92	98
17													87	93	99	105
18														99	106	112
19															113	119
20																127

[2]

(c). Write the significance statement for the analysis performed on this data.

[2]

38.

A friendly name?

Research suggests many things can influence how friendly a person is thought to be. Personality is an obvious one and sense of humour is another. However, there are also less obvious things, such as just knowing a person's name. A psychologist investigated this using an independent measures design experiment. They compared the ratings of friendliness given to shop assistants who wore a name badge to those who did not. Some of the data collected is presented in the table below.

Ratings of friendliness (0 to 20) given to shop assistants who were wearing a name badge compared to shop assistants who were not (0 = 'not friendly at all' to 20 = 'extremely friendly')

Wearing name badge		**Not wearing name badge**	
Rating	**Rank**	**Rating**	**Rank**
18	17	13	11
14	12	2	1
10	8	6	5
17	15	4	3
16	14	8	7
18	17	18	17
5	4	7	6
11	9	15	13
20	20	12	10
19	19	3	2

No qualitative data was collected in this study. Explain why this is a weakness of this study.

[3]

39.

A friendly name?

Research suggests many things can influence how friendly a person is thought to be. Personality is an obvious one and sense of humour is another. However, there are also less obvious things, such as just knowing a person's name. A psychologist investigated this using an independent measures design experiment. They compared the ratings of friendliness given to shop assistants who wore a name badge to those who did not. Some of the data collected is presented in the table below.

Ratings of friendliness (0 to 20) given to shop assistants who were wearing a name badge compared to shop assistants who were not (0 = 'not friendly at all' to 20 = 'extremely friendly')

Wearing name badge		Not wearing name badge	
Rating	**Rank**	**Rating**	**Rank**
18	17	13	11
14	12	2	1
10	8	6	5
17	15	4	3
16	14	8	7
18	17	18	17
5	4	7	6
11	9	15	13
20	20	12	10
19	19	3	2

Identify **and** explain **two** factors that could have affected the reliability of the data collected in this study.

1

[6]

40. Put the letter of the correct answer in the box provided.

If the null hypothesis is incorrectly rejected, what type of error is said to have been made?

A a type 1 error
B a type 2 error
C both a type 1 and a type 2 error
D neither a type 1 nor a type 2 error

Your answer ☐ **[1]**

41. Put the letter of the correct answer in the box provided.

Which of these best describes what the variance informs us of?

A dispersion around the mean
B dispersion around the median
C dispersion around the mode
D dispersion around the range

Your answer ☐ **[1]**

42. Put the letter of the correct answer in the box provided.

In Milgram's (1963) study of obedience, 14 of the 40 participants were disobedient. What is this expressed as a percentage?

A 5.6%
B 14%
C 35%
D 65%

Your answer ☐ **[1]**

43. Put the letter of the correct answer in the box provided.

What is 5% expressed as a fraction?

A $\frac{1}{5}$
B $\frac{1}{20}$
C $\frac{1}{5}$
D $\frac{1}{50}$

Your answer ☐ **[1]**

44. Put the letter of the correct answer in the box provided.

If this data (48, 24, 39, 50, 32, 39) was ranked, what rank would be assigned to 39?

A 2.5
B 3.5
C 4
D 4.5

Your answer ☐ **[1]**

45. Put the letter of the correct answer in the box provided.

In which of these distribution curves is the mean, median and mode all the same value?

A bimodal
B negatively skewed
C normal
D positively skewed

Your answer ☐ **[1]**

46. Put the letter of the correct answer in the box provided.

For which of these inferential statistical tests can the sample size in each condition be different?

A Binomial Sign test
B Mann-Whitney U test
C Wilcoxon Signed Ranks test
D none of them (sample sizes always have to be equal in each condition)

Your answer ☐ **[1]**

47. Put the letter of the correct answer in the box provided.

What is 25.8961 written to **two** decimal places?

A 25
B 25.88
C 25.89
D 25.90

Your answer ☐ **[1]**

48. Put the letter of the correct answer in the box provided.

Which type of validity refers to the agreement between two measures or assessments taken at the same time?

A concurrent
B construct
C criterion
D face

Your answer ☐ **[1]**

49. Put the letter of the correct answer in the box provided.

Which measure of central tendency looks for the most frequently occurring response in the data?

A mean
B median
C mode
D none of them

Your answer ☐ **[1]**

50.

A clean smell?

The aroma of freshly baked bread, the smell of newly ground coffee and the scent of a nice fragrance. We associate different smells with different people and situations, and our sense of smell can even influence how we behave. Sometimes this can be immediate and direct, such as making us feel hungry, but it can also be more indirect. For example, the smell of lemons is often associated with cleanliness. To study this further, psychologists want to use the experimental method to investigate if people leave less litter in a room filled with the smell of lemons compared to one that smells of nothing.

Explain how you would conduct a study using the laboratory experimental method to investigate if there is a difference in the amount of litter left in a room filled with the smell of lemons compared to a room that has no smell. Justify your decisions as part of your explanation. You must refer to:

- the sampling technique used to obtain participants for the study
- how you would operationalise the dependent variable to obtain quantitative data
- details of how one ethical consideration would be addressed
- the control of one extraneous variable.

You should use your own experience of practical activities to inform your response.

[15]

51(a).

A friendly name?

Research suggests many things can influence how friendly a person is thought to be. Personality is an obvious one and sense of humour is another. However, there are also less obvious things, such as just knowing a person's name. A psychologist investigated this using an independent measures design experiment. They compared the ratings of friendliness given to shop assistants who wore a name badge to those who did not. Some of the data collected is presented in the table below.

Ratings of friendliness (0 to 20) given to shop assistants who were wearing a name badge compared to shop assistants who were not (0 = 'not friendly at all' to 20 = 'extremely friendly')

Wearing name badge		Not wearing name badge	
Rating	**Rank**	**Rating**	**Rank**
18	17	13	11
14	12	2	1
10	8	6	5
17	15	4	3
16	14	8	7
18	17	18	17
5	4	7	6
11	9	15	13
20	20	12	10
19	19	3	2

Explain what ranking the data means.

[2]

(b). Explain why there are three ranks of 17.

[2]

52.

A friendly name?

Research suggests many things can influence how friendly a person is thought to be. Personality is an obvious one and sense of humour is another. However, there are also less obvious things, such as just knowing a person's name. A psychologist investigated this using an independent measures design experiment. They compared the ratings of friendliness given to shop assistants who wore a name badge to those who did not. Some of the data collected is presented in the table below.

Ratings of friendliness (0 to 20) given to shop assistants who were wearing a name badge compared to shop assistants who were not (0 = 'not friendly at all' to 20 = 'extremely friendly')

Wearing name badge		Not wearing name badge	
Rating	**Rank**	**Rating**	**Rank**
18	17	13	11
14	12	2	1
10	8	6	5
17	15	4	3
16	14	8	7
18	17	18	17
5	4	7	6
11	9	15	13
20	20	12	10
19	19	3	2

Calculate the mean rating of friendliness in each condition. Show your workings.

[3]

53. Put the letter of the correct answer in the box provided.

Which of these is an ethical consideration acknowledged by the BPS (British Psychological Society)?

A compassion
B competence
C complacence
D compliance

Your answer ☐ **[1]**

54. Put the letter of the correct answer in the box provided.

A group of 220 people answer a question about whether they prefer the colour blue or the colour red. 160 people state that they prefer the colour blue and 60 people state that they prefer the colour red. Which of the following shows these findings expressed as a ratio in its simplest form?

A 8 : 3
B 9 : 5
C 16 : 5
D 40 : 15

Your answer ☐ **[1]**

55. Put the letter of the correct answer in the box provided.

Which of these descriptive statistics involves a comparison to 100 as a reference?

A mean
B median
C mode
D percentage

Your answer ☐ **[1]**

56. Put the letter of the correct answer in the box provided.

Which inferential statistical test at some stage involves a calculation based on the number of rows of data and the number of columns of data?

A Chi-square
B Mann-Whitney U test
C Spearman's Rho
D Wilcoxon Signed Ranks test

Your answer ☐ **[1]**

57. Put the letter of the correct answer in the box provided.

Which of these would be the appropriate inferential test to use to analyse the data in a study investigating the difference in reaction times to two types of stimuli (a light versus a bell) that obtains data from the same people in each condition?

A Chi Square
B Mann-Whitney U test
C Spearman's Rho
D Wilcoxon Signed Ranks test

Your answer ☐ **[1]**

58. Put the letter of the correct answer in the box provided.

In research, which of the following best describes what 'primary data' is?

A data that has **not** been collected by the researcher(s)
B data that is obtained directly from the sample by the researcher(s)
C data that is obtained first in the research
D data that is the most important to use

Your answer ☐ **[1]**

59.

Driving me crazy!

Driving is something that many people do on a regular basis, and there are many things that motorists do that cause others to become frustrated and even angry. To investigate this, a psychologist used the self-report method to find out more about what one thing people find most annoying about the behaviour of other motorists. Below is a table displaying the data collected from the 300 people who took part in the study.

Table displaying the number of responses in each category when participants were asked to identify which one of the following behaviours they find most annoying in other motorists

	Driving too close to another car	**Driving too fast**	**Driving too slow**	**Overtaking**	**Inappropriate use of horn**	**Parking incorrectly**
Male participants' responses	23	20	47	18	19	23
Female participants' responses	33	51	12	24	18	12

Outline the meaning of $p < 0.05$ if it appeared as part of the significance statement when presenting the findings from an inferential statistical test for this study.

[2]

60. Put the letter of the correct answer in the box provided.

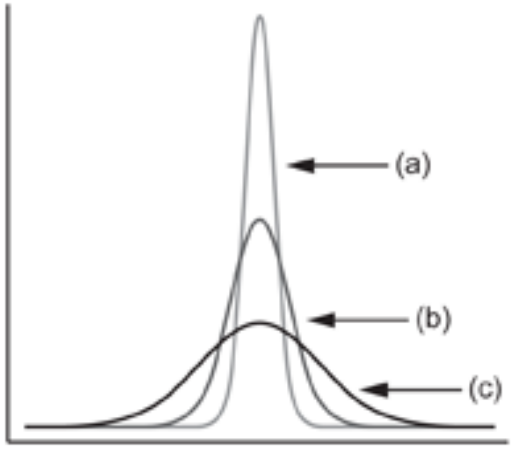
Which of the following means that 'a is proportional to b'?

A a < b
B a > b
C a ~ b
D a ∝ b

Your answer ☐ **[1]**

61. Put the letter of the correct answer in the box provided.

Which of these distributions shows the largest variance?



A distribution (a)
B distribution (b)
C distribution (c)
D they are all the same

Your answer ☐ **[1]**

62. Put the letter of the correct answer in the box provided.

In a study about phobias, which of the following would produce interval level data?

A asking people to put in order, from 'least frightening' to 'most frightening', pictures of five different spiders
B classifying people as 'not frightened', 'frightened' or 'very frightened'
C recording heart rate whilst being shown pictures of spiders
D responses to the open question asking 'What makes you feel frightened?'

Your answer ☐ **[1]**

63. Put the letter of the correct answer in the box provided.

In Milgram's study of obedience, what kind of data were the descriptions of the signs of anxiety shown by the participants?

A interval
B qualitative
C quantitative
D secondary

Your answer ☐ **[1]**

64. Put the letter of the correct answer in the box provided.

Which of these measures of dispersion only uses the highest and lowest values in a data set for its calculation?

A range
B standard deviation
C variance
D standard deviation and variance

Your answer ☐ **[1]**

65. Put the letter of the correct answer in the box provided.

Which measure of central tendency involves ranking the data first?

A mean
B median
C mode
D none of them

Your answer ☐ **[1]**

66.

Sounds friendly?

There are many different styles of music and some can be very emotional and even influence how we behave. A psychologist wants to use the laboratory experimental method to investigate if listening to music with prosocial lyrics (words related to love and kindness) makes people more helpful afterwards.

Outline **two** weaknesses of having quantitative data when trying to investigate whether prosocial lyrics influence how helpful people are.

1

2

[6]

67.

Driving me crazy!

Driving is something that many people do on a regular basis, and there are many things that motorists do that cause others to become frustrated and even angry. To investigate this, a psychologist used the self-report method to find out more about what one thing people find most annoying about the behaviour of other motorists. Below is a table displaying the data collected from the 300 people who took part in the study.

Table displaying the number of responses in each category when participants were asked to identify which one of the following behaviours they find most annoying in other motorists

	Driving too close to another car	Driving too fast	Driving too slow	Overtaking	Inappropriate use of horn	Parking incorrectly
Male participants' responses	23	20	47	18	19	23
Female participants' responses	33	51	12	24	18	12

What type of question has been used to produce the data displayed in the table above?
Give reasons for your answer.

[2]

68.

Driving me crazy!

Driving is something that many people do on a regular basis, and there are many things that motorists do that cause others to become frustrated and even angry. To investigate this, a psychologist used the self-report method to find out more about what one thing people find most annoying about the behaviour of other motorists. Below is a table displaying the data collected from the 300 people who took part in the study.

Table displaying the number of responses in each category when participants were asked to identify which one of the following behaviours they find most annoying in other motorists

	Driving too close to another car	Driving too fast	Driving too slow	Overtaking	Inappropriate use of horn	Parking incorrectly
Male participants' responses	23	20	47	18	19	23
Female participants' responses	33	51	12	24	18	12

Outline **two** conclusions that can be made from the data collected in this study.

1

2

[4]

69.

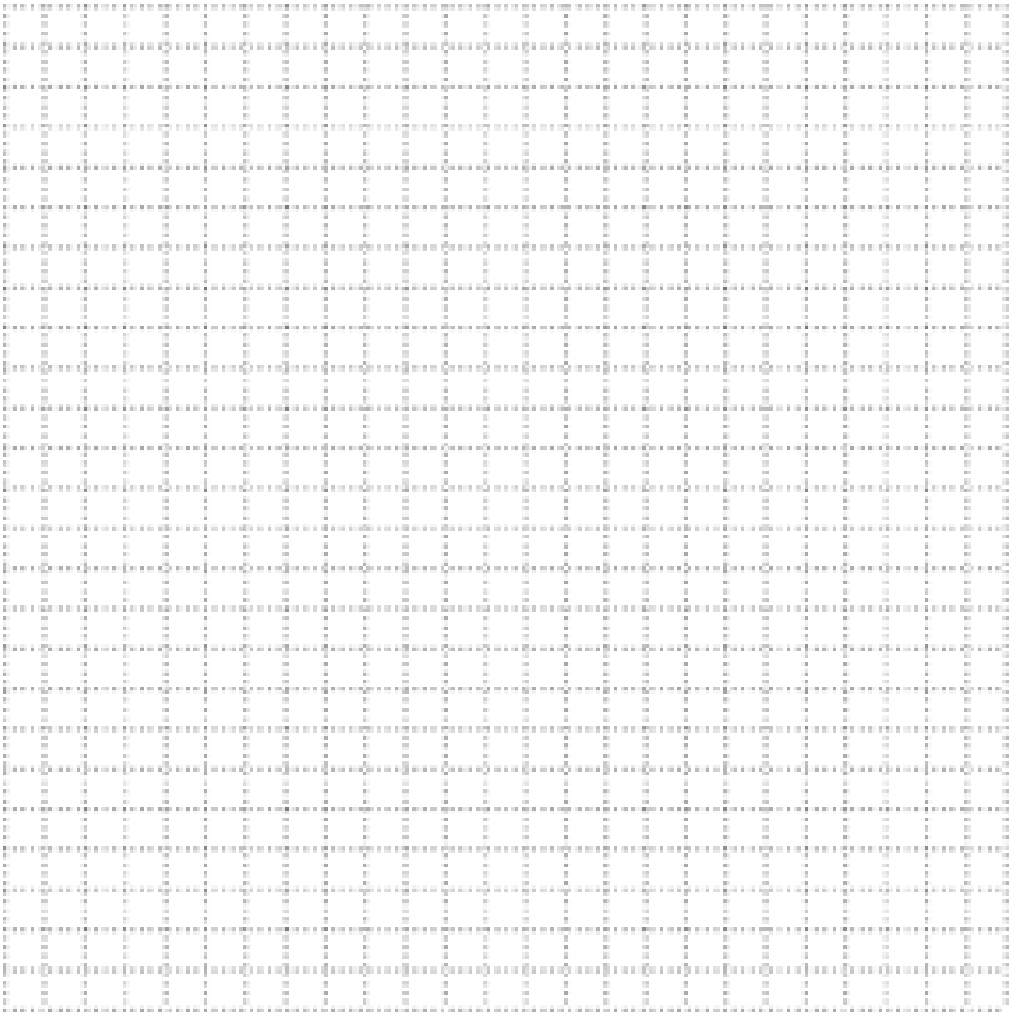
Driving me crazy!

Driving is something that many people do on a regular basis, and there are many things that motorists do that cause others to become frustrated and even angry. To investigate this, a psychologist used the self-report method to find out more about what one thing people find most annoying about the behaviour of other motorists. Below is a table displaying the data collected from the 300 people who took part in the study.

Table displaying the number of responses in each category when participants were asked to identify which one of the following behaviours they find most annoying in other motorists

	Driving too close to another car	**Driving too fast**	**Driving too slow**	**Overtaking**	**Inappropriate use of horn**	**Parking incorrectly**
Male participants' responses	23	20	47	18	19	23
Female participants' responses	33	51	12	24	18	12

Draw a fully labelled bar chart displaying the data from the **male participants** who took part in this study.



[4]

70.

Driving me crazy!

Driving is something that many people do on a regular basis, and there are many things that motorists do that cause others to become frustrated and even angry. To investigate this, a psychologist used the self-report method to find out more about what one thing people find most annoying about the behaviour of other motorists. Below is a table displaying the data collected from the 300 people who took part in the study.

Table displaying the number of responses in each category when participants were asked to identify which one of the following behaviours they find most annoying in other motorists

	Driving too close to another car	Driving too fast	Driving too slow	Overtaking	Inappropriate use of horn	Parking incorrectly
Male participants' responses	23	20	47	18	19	23
Female participants' responses	33	51	12	24	18	12

Calculate the overall percentage of people reporting 'overtaking' as being most annoying. Show your workings. The table displaying the data collected from the 300 people who took part in the study is repeated below.

[3]

71(a).

Driving me crazy!

Driving is something that many people do on a regular basis, and there are many things that motorists do that cause others to become frustrated and even angry. To investigate this, a psychologist used the self-report method to find out more about what one thing people find most annoying about the behaviour of other motorists. Below is a table displaying the data collected from the 300 people who took part in the study.

Table displaying the number of responses in each category when participants were asked to identify which one of the following behaviours they find most annoying in other motorists

	Driving too close to another car	Driving too fast	Driving too slow	Overtaking	Inappropriate use of horn	Parking incorrectly
Male participants' responses	23	20	47	18	19	23
Female participants' responses	33	51	12	24	18	12

Explain why the Chi square test would be the appropriate non-parametric inferential statistical test to use to analyse the data from this study.

[3]

(b). Calculate the ‘degrees of freedom’ required to check for significance when using the Chi square test to analyse the data collected in this study.

[1]

(c). After it is calculated, how is the ‘degrees of freedom’ used when checking for significance?

[2]

72. Put the letter of the correct answer in the box provided.

Which of these shows the ratio 42 : 6 in its simplest form?

A 5 : 1
B 6 : 1
C 7 : 1
D 21 : 3

Your answer ☐ **[1]**

73. Put the letter of the correct answer in the box provided.

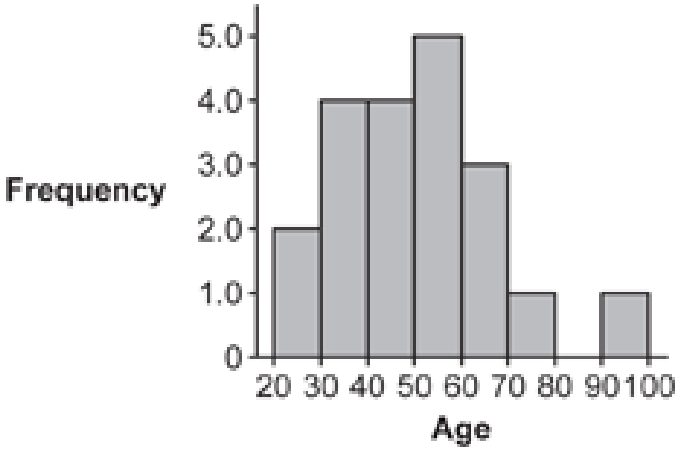
What is the range for this set of scores? 52, 63, 48, 37, 24

A 24
B 28
C 39
D 52

Your answer ☐ **[1]**

74. Put the letter of the correct answer in the box provided.

What type of graph is this?



A bar chart
B histogram
C line graph
D pie chart

Your answer ☐ **[1]**

75.

Knowing me, knowing you

How well do we know each other and how much does this reflect what we think of our self? To investigate this a psychologist asked ten people to rate their own level of extroversion on a scale of 0 ('not very extrovert at all') to 100 ('extremely extrovert') and then asked a close friend of each person to rate them using the same scale. The results are presented in the table below.

Participant	Self-rating of extroversion (0 to 100)	Friend's rating (0 to 100)
a	100	90
b	80	25
c	90	10
d	40	70
e	15	95
f	70	45
g	35	85
h	50	60
i	60	45
j	25	100

Given that the value of $d^2 = 270.5$, use the formula for Spearman's rho provided here to calculate the value of r_s. Show your workings.

$$r_s = 1 - \frac{6(\Sigma d^2)}{n(n^2 - 1)}$$

[3]

76.

Knowing me, knowing you

How well do we know each other and how much does this reflect what we think of our self? To investigate this a psychologist asked ten people to rate their own level of extroversion on a scale of 0 ('not very extrovert at all') to 100 ('extremely extrovert') and then asked a close friend of each person to rate them using the same scale. The results are presented in the table below.

Participant	Self-rating of extroversion (0 to 100)	Friend's rating (0 to 100)
a	100	90
b	80	25
c	90	10
d	40	70
e	15	95
f	70	45
g	35	85
h	50	60
i	60	45
j	25	100

Explain what $p>0.05$ would mean in the significance statement for the result of the Spearman's rho test from this study.

[3]

77. Put the letter of the correct answer in the box provided.

Which of these indicates the research is significant at the 3% level of probability?

A $p<3.0$
B $p<0.3$
C $p<0.03$
D $p<0.003$

Your answer ☐ **[1]**

78. Put the letter of the correct answer in the box provided.

Which of these is not a type of validity?

A construct
B criterion
C face
D visual

Your answer ☐ **[1]**

79. Put the letter of the correct answer in the box provided.

Which of these best describes the kind of data collected in Milgram's (1963) study of obedience?

A both quantitative and qualitative
B neither quantitative or qualitative
C quantitative
D qualitative

Your answer ☐ **[1]**

80. Put the letter of the correct answer in the box provided.

Which of these could be the result of Spearman's rho inferential statistical test?

A 0.05
B 1.05
C −1.05
D 11.05

Your answer ☐ **[1]**

81.

Mind changing?

Attitudes towards mental health and how it is treated have changed a lot over the years. However, despite this there are still differences in how people with physical, compared to mental illnesses are thought of and cared for. Psychologists want to study this further using the self-report method to investigate peoples' attitudes towards mental health.

Outline **two** ways that you would try to reduce the influence of demand characteristics in this self-report study.

[6]

82.

Knowing me, knowing you

How well do we know each other and how much does this reflect what we think of our self? To investigate this a psychologist asked ten people to rate their own level of extroversion on a scale of 0 ('not very extrovert at all') to 100 ('extremely extrovert') and then asked a close friend of each person to rate them using the same scale. The results are presented in the table below.

Participant	Self-rating of extroversion (0 to 100)	Friend's rating (0 to 100)
a	100	90
b	80	25
c	90	10
d	40	70
e	15	95
f	70	45
g	35	85
h	50	60
i	60	45
j	25	100

Outline **two** conclusions that can be made from the data collected in this study.

[6]

83.

Knowing me, knowing you

How well do we know each other and how much does this reflect what we think of our self? To investigate this a psychologist asked ten people to rate their own level of extroversion on a scale of 0 ('not very extrovert at all') to 100 ('extremely extrovert') and then asked a close friend of each person to rate them using the same scale. The results are presented in the table below.

Participant	Self-rating of extroversion (0 to 100)	Friend's rating (0 to 100)
a	100	90
b	80	25
c	90	10
d	40	70
e	15	95
f	70	45
g	35	85
h	50	60
i	60	45
j	25	100

Outline **one** strength and **one** weakness of the use of quantifiable data in this study.

[6]

84.

Knowing me, knowing you

How well do we know each other and how much does this reflect what we think of our self? To investigate this a psychologist asked ten people to rate their own level of extroversion on a scale of 0 ('not very extrovert at all') to 100 ('extremely extrovert') and then asked a close friend of each person to rate them using the same scale. The results are presented in the table below.

Participant	Self-rating of extroversion (0 to 100)	Friend's rating (0 to 100)
a	100	90
b	80	25
c	90	10
d	40	70
e	15	95
f	70	45
g	35	85
h	50	60
i	60	45
j	25	100

Explain why the Spearman's rho would be the appropriate non-parametric inferential statistical test to use to analyse the data from this study.

[3]

85. Put the letter of the correct answer in the box provided.

Which of these is not covered by the ethical issue of 'respect'?

A confidentiality
B consent
C payment
D withdrawal

Your answer ☐ **[1]**

86. Put the letter of the correct answer in the box provided.

Which of these is a type of reliability?

A criterion
B ecological
C face
D test-retest

Your answer ☐ **[1]**

87. Put the letter of the correct answer in the box provided.

What type of data is used to calculate the Chi-square test?

A Interval
B Nominal
C Ordinal
D ordinal and interval

Your answer ☐ **[1]**

88. Put the letter of the correct answer in the box provided.

Which is the simplest form of the ratio 12:8?

A 2:3
B 3:1
C 3:2
D 6:4

Your answer ☐ **[1]**

89. Put the letter of the correct answer in the box provided.

What decimal is represented by the fraction 1/25?

A 0.25
B 0.04
C 0.02
D 0.05

Your answer ☐ **[1]**

90. Put the letter of the correct answer in the box provided.

Which inferential statistical test simply involves counting the number of times the values in one condition are higher or lower than those in the other?

A Chi-square
B Binomial Sign
C Spearman's Rho
D Wilcoxon Signed Ranks

Your answer ☐ **[1]**

91. Put the letter of the correct answer in the box provided.

Which of these is **not** a criterion for the use of parametric inferential statistical tests?

A data must be interval level
B data must be normally distributed in the population
C sample size
D variance between conditions should be similar

Your answer ☐ **[1]**

92. Put the letter of the correct answer in the box provided.

In which type of distribution is the mean greater than both the median and the mode?

A bimodal
B negatively skewed
C normal
D positively skewed

Your answer ☐ **[1]**

93.

Get a grip

Hand shaking is a fairly common greeting between people in some cultures, especially when meeting someone for the first time. However, the way that we shake hands with someone can vary quite a lot, and this might influence what we think of the person whose hand we are shaking. Some research suggests that simply the length of time that the hand is shaken can influence how friendly or not we perceive the person to be. Psychologists want to study this using the experimental method to investigate the effect of a 'short' compared to a 'long' handshake on how friendly a person is thought to be.

Outline **one** way to help reduce demand characteristics in this study.

[2]

94.

Mobile mobile

Most people have a mobile phone so making or receiving a call while out and about is very easy. However, there may be differences in people's behaviour when talking on the phone. To study this a psychologist conducted an observation study to investigate if there are differences in how much men and women gesture (e.g. move hands or head) while talking on the phone. To do this they sat on a bench on a busy high street for two hours one Saturday afternoon and recorded the number of times people made gestures of any kind or not while talking on the phone. The data is presented in the table below.

Table showing the number of males and females who made gestures or not while making or receiving a mobile phone call

Use of gestures	Males	Females
Yes	(a) 11	(b) 5
No	(c) 3	(d) 14

Outline **one** conclusion that can be obtained from this study from the data presented in this table.

[3]

95.

Mobile mobile

Most people have a mobile phone so making or receiving a call while out and about is very easy. However, there may be differences in people's behaviour when talking on the phone. To study this a psychologist conducted an observation study to investigate if there are differences in how much men and women gesture (e.g. move hands or head) while talking on the phone. To do this they sat on a bench on a busy high street for two hours one Saturday afternoon and recorded the number of times people made gestures of any kind or not while talking on the phone. The data is presented in the table below.

Table showing the number of males and females who made gestures or not while making or receiving a mobile phone call

Use of gestures	Males	Females
Yes	(a) 11	(b) 5
No	(c) 3	(d) 14

Draw a fully labelled pie chart showing the percentage of males compared to females who made gestures whilst on the phone.

[4]

96.

Mobile mobile

Most people have a mobile phone so making or receiving a call while out and about is very easy. However, there may be differences in people's behaviour when talking on the phone. To study this a psychologist conducted an observation study to investigate if there are differences in how much men and women gesture (e.g. move hands or head) while talking on the phone. To do this they sat on a bench on a busy high street for two hours one Saturday afternoon and recorded the number of times people made gestures of any kind or not while talking on the phone. The data is presented in the table below.

Table showing the number of males and females who made gestures or not while making or receiving a mobile phone call

Use of gestures	Males	Females
Yes	(a) 11	(b) 5
No	(c) 3	(d) 14

The psychologist used the Chi-square test to analyse the data from this study. Give **one** reason why this would be the appropriate non-parametric inferential test to use in this study.

[2]

97(a).

Mobile mobile

Most people have a mobile phone so making or receiving a call while out and about is very easy. However, there may be differences in people's behaviour when talking on the phone. To study this a psychologist conducted an observation study to investigate if there are differences in how much men and women gesture (e.g. move hands or head) while talking on the phone. To do this they sat on a bench on a busy high street for two hours one Saturday afternoon and recorded the number of times people made gestures of any kind or not while talking on the phone. The data is presented in the table below.

Table showing the number of males and females who made gestures or not while making or receiving a mobile phone call

Use of gestures	Males	Females
Yes	(a) 11	(b) 5
No	(c) 3	(d) 14

The formula for the Chi-square test requires the use of expected frequencies (E). The expected frequencies for three of the cells are provided in the table below.

Cell	Observed frequency (O)	Expected frequency (E)	O – E	$(O – E)^2$	$(O – E)^2 / E$
A	11	6.79	4.21	17.72	2.61
B	5	9.21	-4.21	17.72	1.92
C	3	7.21	-4.21	17.72	2.46
D	14				

Calculate the expected frequency for cell D. Present your answer to two decimal places and show your workings.

[3]

Formula for X^2

$$X^2 = \sum \frac{(O-E)^2}{E}$$

(b). Calculate the overall value of Chi-square. Show your workings.

[3]

(c). Calculate the degrees of freedom for use with the Chi-square test in this study. Show your workings.

[2]

(d). Using the extract from the table of critical values presented below, what is the critical value for use with the Chi-square test in this study at the 1% level of probability?

[1]

	Probability level					
df	**0.5**	**0.10**	**0.05**	**0.02**	**0.01**	**0.001**
1	0.455	2.706	3.841	5.412	6.635	10.827
2	1.386	4.605	5.991	7.824	9.210	13.815
3	2.366	6.251	7.815	9.834	11.345	16.268
4	3.357	7.779	9.488	11.668	13.277	18.465
5	4.351	9.236	11.070	13.388	15.086	20.517

(e). What conclusions can be reached from the calculation of Chi-square in this study?

[3]

98(a).

Mobile mobile

Most people have a mobile phone so making or receiving a call while out and about is very easy. However, there may be differences in people's behaviour when talking on the phone. To study this a psychologist conducted an observation study to investigate if there are differences in how much men and women gesture (e.g. move hands or head) while talking on the phone. To do this they sat on a bench on a busy high street for two hours one Saturday afternoon and recorded the number of times people made gestures of any kind or not while talking on the phone. The data is presented in the table below.

Table showing the number of males and females who made gestures or not while making or receiving a mobile phone call

Use of gestures	Males	Females
Yes	(a) 11	(b) 5
No	(c) 3	(d) 14

Outline **one** strength of the use of the quantitative data collected in this study.

[3]

(b). Outline **one** weakness of the use of the quantitative data collected in this study.

[3]

99.

Mobile mobile

Most people have a mobile phone so making or receiving a call while out and about is very easy. However, there may be differences in people's behaviour when talking on the phone. To study this a psychologist conducted an observation study to investigate if there are differences in how much men and women gesture (e.g. move hands or head) while talking on the phone. To do this they sat on a bench on a busy high street for two hours one Saturday afternoon and recorded the number of times people made gestures of any kind or not while talking on the phone. The data is presented in the table below.

Table showing the number of males and females who made gestures or not while making or receiving a mobile phone call

Use of gestures	Males	Females
Yes	(a) 11	(b) 5
No	(c) 3	(d) 14

Outline **two** ways in which the naturalistic observation method used in this study could affect the validity of the data collected.

1

[6]

100. Put the letter of the correct answer in the box provided.

Which of these is an inferential rather than a descriptive statistical analysis?

A Mann-Whitney U
B mean
C standard deviation
D Variance

Your answer ☐ **[1]**

101. Put the letter of the correct answer in the box provided.

If the variance is 36, what is the standard deviation?

A 3
B 6
C 18
D 1296

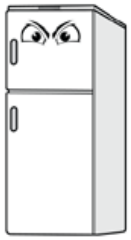
Your answer ☐ **[1]**

102(a).

Cold stare, you dare!

Controlling what we eat can be a challenge and any new idea may help. One psychologist has been investigating the effect of putting a picture of a pair of eyes on a fridge door to see if this can help reduce going to the fridge for snacks. She conducted an experiment using a small sample of 12 participants obtained using snowball sampling. For one week participants kept a record of the number of times they went into the fridge in-between meals for a snack (cheese, chocolate, fruit etc). The following week a pair of eyes was placed on the fridge door and the same participants were again asked to keep a record of the number of times they got a snack from the fridge. The data is presented below.

Participant	Number of times snacks taken from fridge during one week	
	Without eyes on fridge door	With eyes on fridge door
a	14	10
b	21	18
c	32	24
d	28	24
e	20	16
f	8	10
g	26	24
h	12	12
i	22	23
j	18	12
k	20	19
l	21	21



Using the Wilcoxon T Signed Ranks test, calculate the value of T.

Formula / process for calculating T: T = the sum of differences of the ranks with the least frequent sign

[5]

(b). Write the significance statement for the analysis carried out in above part with a probability of 5%, assuming a non-directional hypothesis.

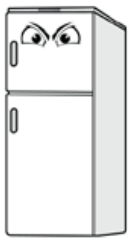
[2]

103(a).

Cold stare, you dare!

Controlling what we eat can be a challenge and any new idea may help. One psychologist has been investigating the effect of putting a picture of a pair of eyes on a fridge door to see if this can help reduce going to the fridge for snacks. She conducted an experiment using a small sample of 12 participants obtained using snowball sampling. For one week participants kept a record of the number of times they went into the fridge in-between meals for a snack (cheese, chocolate, fruit etc). The following week a pair of eyes was placed on the fridge door and the same participants were again asked to keep a record of the number of times they got a snack from the fridge. The data is presented below.

Participant	Number of times snacks taken from fridge during one week	
	Without eyes on fridge door	With eyes on fridge door
a	14	10
b	21	18
c	32	24
d	28	24
e	20	16
f	8	10
g	26	24
h	12	12
i	22	23
j	18	12
k	20	19
l	21	21



Using the extract of the table of critical values for the Wilcoxon T Signed Ranks test presented below, what is the critical value for a two-tailed test at the 5% probability level for data collected in this study?

Extract of table of critical values for Wilcoxon T Signed Ranks test
(*T* is significant if it is **less than or equal** to the table value)

N	level of significance for a one-tailed test					
	.10	.05	.025	.01	.005	.001
	level of significance for a two-tailed test					
	.20	.10	.05	.02	.01	.002
4	0					
5	2	0				
6	4	2	0			
7	6	3	2	0		
8	8	5	3	1	0	
9	11	8	5	3	1	
10	14	10	8	5	3	0
11	17	13	10	8	5	3
12	21	17	13	10	7	5

[2]

(b). The Wilcoxon T Signed Ranks test was used as the data was considered **not** to be parametric. Identify **two** of the criteria for using a parametric test to analyse data.

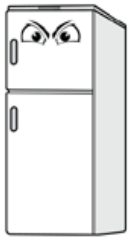
[2]

104(a).

Cold stare, you dare!

Controlling what we eat can be a challenge and any new idea may help. One psychologist has been investigating the effect of putting a picture of a pair of eyes on a fridge door to see if this can help reduce going to the fridge for snacks. She conducted an experiment using a small sample of 12 participants obtained using snowball sampling. For one week participants kept a record of the number of times they went into the fridge in-between meals for a snack (cheese, chocolate, fruit etc). The following week a pair of eyes was placed on the fridge door and the same participants were again asked to keep a record of the number of times they got a snack from the fridge. The data is presented below.

Participant	Number of times snacks taken from fridge during one week	
	Without eyes on fridge door	With eyes on fridge door
a	14	10
b	21	18
c	32	24
d	28	24
e	20	16
f	8	10
g	26	24
h	12	12
i	22	23
j	18	12
k	20	19
l	21	21



Outline **one** advantage of having quantitative data in this study.

[3]

(b). Outline **one** disadvantage of not having qualitative data in this study.

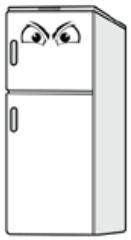
[3]

105(a).

Cold stare, you dare!

Controlling what we eat can be a challenge and any new idea may help. One psychologist has been investigating the effect of putting a picture of a pair of eyes on a fridge door to see if this can help reduce going to the fridge for snacks. She conducted an experiment using a small sample of 12 participants obtained using snowball sampling. For one week participants kept a record of the number of times they went into the fridge in-between meals for a snack (cheese, chocolate, fruit etc). The following week a pair of eyes was placed on the fridge door and the same participants were again asked to keep a record of the number of times they got a snack from the fridge. The data is presented below.

Participant	Number of times snacks taken from fridge during one week	
	Without eyes on fridge door	With eyes on fridge door
a	14	10
b	21	18
c	32	24
d	28	24
e	20	16
f	8	10
g	26	24
h	12	12
i	22	23
j	18	12
k	20	19
l	21	21



Calculate the mean for the number of snacks taken from the fridge in the week **without** eyes on the door. Show your workings and write your answer to three significant figures.

[3]

(b). Calculate the median and range for the number of snacks taken from the fridge in the week **with** eyes on the door. Show your workings.

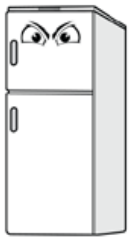
[3]

106.

Cold stare, you dare!

Controlling what we eat can be a challenge and any new idea may help. One psychologist has been investigating the effect of putting a picture of a pair of eyes on a fridge door to see if this can help reduce going to the fridge for snacks. She conducted an experiment using a small sample of 12 participants obtained using snowball sampling. For one week participants kept a record of the number of times they went into the fridge in-between meals for a snack (cheese, chocolate, fruit etc). The following week a pair of eyes was placed on the fridge door and the same participants were again asked to keep a record of the number of times they got a snack from the fridge. The data is presented below.

Participant	Number of times snacks taken from fridge during one week	
	Without eyes on fridge door	With eyes on fridge door
a	14	10
b	21	18
c	32	24
d	28	24
e	20	16
f	8	10
g	26	24
h	12	12
i	22	23
j	18	12
k	20	19
l	21	21



The snacking of participant (c) was less during the week with eyes on the fridge than in the week without eyes on the fridge. Calculate this reduction as a percentage. Show your workings.

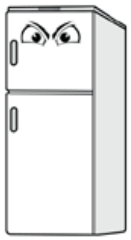
[3]

107(a).

Cold stare, you dare!

Controlling what we eat can be a challenge and any new idea may help. One psychologist has been investigating the effect of putting a picture of a pair of eyes on a fridge door to see if this can help reduce going to the fridge for snacks. She conducted an experiment using a small sample of 12 participants obtained using snowball sampling. For one week participants kept a record of the number of times they went into the fridge in-between meals for a snack (cheese, chocolate, fruit etc). The following week a pair of eyes was placed on the fridge door and the same participants were again asked to keep a record of the number of times they got a snack from the fridge. The data is presented below.

Participant	Number of times snacks taken from fridge during one week	
	Without eyes on fridge door	With eyes on fridge door
a	14	10
b	21	18
c	32	24
d	28	24
e	20	16
f	8	10
g	26	24
h	12	12
i	22	23
j	18	12
k	20	19
l	21	21



What does the information from the standard deviation for each condition from this study presented below inform us?

Standard deviation for each condition	
Without eyes on fridge door	With eyes on fridge door
6.726	5.613

[3]

(b). Outline **one** strength of using the standard deviation rather than the variance as the measure of dispersion for the data collected in this study.

[3]

108. Which of the following could help reduce demand characteristics?

A debriefing the participants
B deceiving the participants
C harming the participants
D paying the participants

Your answer ☐ **[1]**

109. Which of these is a type of inferential statistical test?

A parabolic
B paradoxical
C parametric
D parasympathetic

Your answer ☐ **[1]**

110. Which of these statements is true?

A interval data can be treated as ordinal data for analysis
B nominal data can be ranked for analysis
C nominal data cannot be statistically analysed
D ordinal data can be treated as interval data for analysis

Your answer ☐ **[1]**

111. Which of these is a British Psychological Society ethical consideration when conducting psychological research?

A compassion
B competence
C compliance
D compromise

Your answer ☐ **[1]**

112. In order to use the chi-square test, the expected values (E) of each category of data collected must be worked out first. What is the E value of the data in category cell (a) in the following table?

(a) 12	(b) 8
(c) 16	(d) 4

$$\text{Expected value} = \frac{\text{Row total} \times \text{Column total}}{N}$$

A 14
B 20
C 28
D 40

Your answer ☐ **[1]**

113. In a study investigating use of mobile phones, the following behavioural categories are used: texting; phoning; browsing; gaming and listening to music. What kind of data will be collected?

A interval
B nominal
C ordinal
D secondary

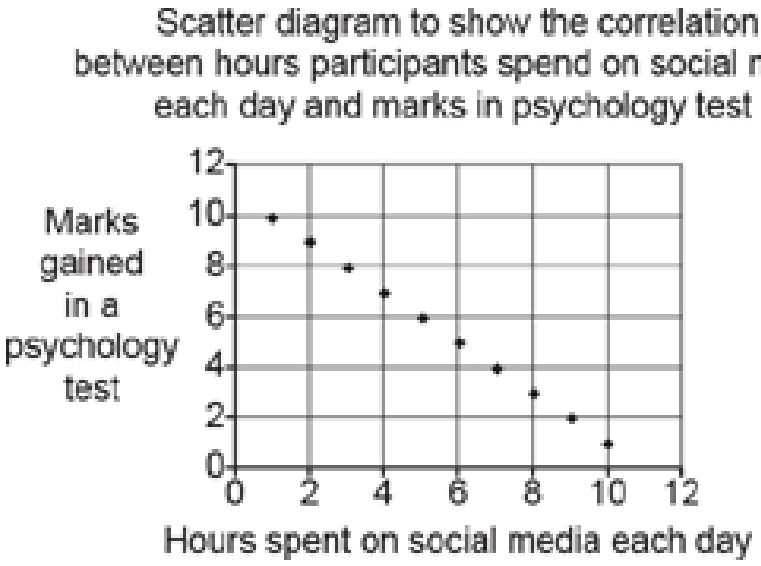
Your answer ☐ **[1]**

114. If the standard deviation for a set of scores is 6.3, what is the variance?

A 2.51
B 6.3
C 12.6
D 39.69

Your answer ☐ **[1]**

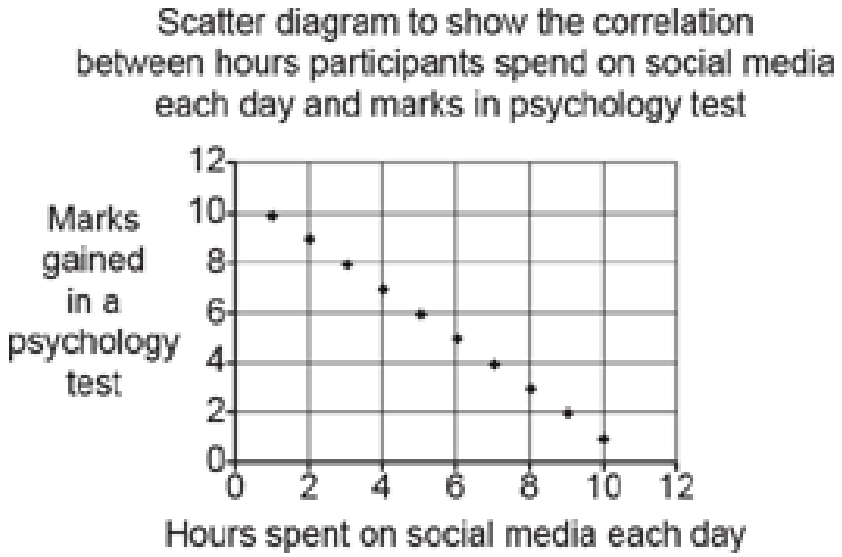
115(a). Which correlation coefficient is represented in the data displayed in this scatter diagram?



A −1.0
B −10.0
C +1.0
D +10.0

Your answer ☐ **[1]**

(b). How many participants' data are displayed in this scatter diagram?



A 5
B 10
C 12
D 20

Your answer ☐ **[1]**

116. What level of data and type of design are required to use the Mann-Whitney U test?

A nominal data with independent measures design
B nominal data with repeated measures design
C ordinal data with independent measures design
D ordinal data with repeated measures design

Your answer ☐ **[1]**

117.

Outline **one** strength of the type of data collected in Maguire et al.'s (2000) study into the brains of taxi drivers.

[2]

118.

Briefly discuss the extent to which Kohlberg's (1968) study into moral development may be considered ethnocentric.

[4]

119.

i. Explain **one** way in which the procedure of Grant et al.'s (1998) study into memory increased the reliability of the research.

[3]

ii. Outline **one** conclusion that was drawn from this study.

[2]

120.

i. Explain **one** weakness of the type of data collected in Loftus & Palmer's (1974) study into eyewitness testimony.

[3]

ii. Outline **one** conclusion that was drawn from this study.

[2]

121(a).

Extroversion = Extra money?

A psychologist investigated whether how much a person earns is related to the personality trait of extroversion. Participants were asked to provide details of their current annual salary. They then completed a general lifestyle questionnaire, which included a question asking them to rate how extrovert they were on a scale of 0 ('not at all extrovert') to 100 ('extremely extrovert'). The table below presents the data collected in the study.

Participant	Annual salary (£s)	Extroversion rating (0–100)	Ranks of the data collected	
			Annual salary (£s)	Extroversion rating (0–100)
a	50 000	85	9	9
b	55 000	65	10	6.5
c	10 000	95	1	11.5
d	40 000	60	7	5
e	30 000	45	5	4
f	15 000	30	2	1
g	25 000	40	4	3
h	60 000	90	11	10
i	35 000	70	6	8
j	20 000	35	3	2
k	70 000	95	12	11.5
l	45 000	65	8	6.5

Calculate the correlation coefficient for the data collected in this study using the formula for Spearman's ranked correlation coefficient presented below. Show your workings.

$$r_s = 1 - \frac{6(\Sigma d^2)}{n(n^2-1)}$$

[5]

(b). An extract from the table of critical values for Spearman's ranked correlation coefficient is shown below.

	Level of significance for a two-tailed test			
	0.10	0.05	0.02	0.01
$n = 4$	1.000			
5	0.900	1.000	1.000	
6	0.829	0.886	0.943	1.000
7	0.714	0.786	0.893	0.929
8	0.643	0.738	0.833	0.881
9	0.600	0.700	0.783	0.833
10	0.564	0.648	0.745	0.794
11	0.536	0.618	0.709	0.755
12	0.503	0.587	0.671	0.727
13	0.484	0.560	0.648	0.703
14	0.464	0.538	0.622	0.675
15	0.443	0.521	0.604	0.654
16	0.429	0.503	0.582	0.635
17	0.414	0.485	0.566	0.615
18	0.401	0.472	0.550	0.600
19	0.391	0.460	0.535	0.584
20	0.380	0.447	0.520	0.570

r_s must equal or exceed the table critical value to be significant at the stated level of probability.

Using this table, identify the critical value at the 5% probability level for data collected in this study.

[2]

(c). Write the significance statement for the analysis performed on this data using Spearman's ranked correlation coefficient.

[2]

122.

Extroversion = Extra money?

A psychologist investigated whether how much a person earns is related to the personality trait of extroversion. Participants were asked to provide details of their current annual salary. They then completed a general lifestyle questionnaire, which included a question asking them to rate how extrovert they were on a scale of 0 ('not at all extrovert') to 100 ('extremely extrovert'). The table below presents the data collected in the study.

Participant	Annual salary (£s)	Extroversion rating (0–100)	Ranks of the data collected	
			Annual salary (£s)	Extroversion rating (0–100)
a	50 000	85	9	9
b	55 000	65	10	6.5
c	10 000	95	1	11.5
d	40 000	60	7	5
e	30 000	45	5	4
f	15 000	30	2	1
g	25 000	40	4	3
h	60 000	90	11	10
i	35 000	70	6	8
j	20 000	35	3	2
k	70 000	95	12	11.5
l	45 000	65	8	6.5

Outline **one** weakness of not having qualitative data in this study.

[3]

123.

Extroversion = Extra money?

A psychologist investigated whether how much a person earns is related to the personality trait of extroversion. Participants were asked to provide details of their current annual salary. They then completed a general lifestyle questionnaire, which included a question asking them to rate how extrovert they were on a scale of 0 ('not at all extrovert') to 100 ('extremely extrovert'). The table below presents the data collected in the study.

Participant	Annual salary (£s)	Extroversion rating (0–100)	Ranks of the data collected	
			Annual salary (£s)	Extroversion rating (0–100)
a	50 000	85	9	9
b	55 000	65	10	6.5
c	10 000	95	1	11.5
d	40 000	60	7	5
e	30 000	45	5	4
f	15 000	30	2	1
g	25 000	40	4	3
h	60 000	90	11	10
i	35 000	70	6	8
j	20 000	35	3	2
k	70 000	95	12	11.5
l	45 000	65	8	6.5

Outline **two** things that could have affected the validity of the data collected in this study.

[6]

124(a).

Ready meal

Many things can influence people's eating habits and how much food we consume, including the appearance of food and how it is served. For example, some research suggests that people put less food on a red plate, and eat less from it, than a white plate. To investigate this further, psychologists want to use the experimental method to study the effect of plate colour on food consumption.

Outline **one** way you could obtain qualitative data in this study.

[4]

(b). Evaluate the use of qualitative data in this study.

[3]

125.

Ready meal

Many things can influence people's eating habits and how much food we consume, including the appearance of food and how it is served. For example, some research suggests that people put less food on a red plate, and eat less from it, than a white plate. To investigate this further, psychologists want to use the experimental method to study the effect of plate colour on food consumption.

Evaluate the external validity of this study.

[6]

126(a).

Extroversion = Extra money?

A psychologist investigated whether how much a person earns is related to the personality trait of extroversion. Participants were asked to provide details of their current annual salary. They then completed a general lifestyle questionnaire, which included a question asking them to rate how extrovert they were on a scale of 0 ('not at all extrovert') to 100 ('extremely extrovert'). The table below presents the data collected in the study.

Participant	Annual salary (£s)	Extroversion rating (0–100)	Ranks of the data collected	
			Annual salary (£s)	Extroversion rating (0–100)
a	50 000	85	9	9
b	55 000	65	10	6.5
c	10 000	95	1	11.5
d	40 000	60	7	5
e	30 000	45	5	4
f	15 000	30	2	1
g	25 000	40	4	3
h	60 000	90	11	10
i	35 000	70	6	8
j	20 000	35	3	2
k	70 000	95	12	11.5
l	45 000	65	8	6.5

Explain how the data has been ranked in this study.

[4]

(b). Explain why the data for the rating of extroversion for participants (b) and (l) have both been given rank 6.5.

[2]

127(a).

Extroversion = Extra money?

A psychologist investigated whether how much a person earns is related to the personality trait of extroversion. Participants were asked to provide details of their current annual salary. They then completed a general lifestyle questionnaire, which included a question asking them to rate how extrovert they were on a scale of 0 ('not at all extrovert') to 100 ('extremely extrovert'). The table below presents the data collected in the study.

Participant	Annual salary (£s)	Extroversion rating (0–100)	Ranks of the data collected	
			Annual salary (£s)	Extroversion rating (0–100)
a	50 000	85	9	9
b	55 000	65	10	6.5
c	10 000	95	1	11.5
d	40 000	60	7	5
e	30 000	45	5	4
f	15 000	30	2	1
g	25 000	40	4	3
h	60 000	90	11	10
i	35 000	70	6	8
j	20 000	35	3	2
k	70 000	95	12	11.5
l	45 000	65	8	6.5

Draw a fully labelled scatter diagram showing the data collected in this study.

[4]

(b). Outline **two** conclusions from this scatter diagram.

[4]

128.

Extroversion = Extra money?

A psychologist investigated whether how much a person earns is related to the personality trait of extroversion. Participants were asked to provide details of their current annual salary. They then completed a general lifestyle questionnaire, which included a question asking them to rate how extrovert they were on a scale of 0 ('not at all extrovert') to 100 ('extremely extrovert'). The table below presents the data collected in the study.

Participant	Annual salary (£s)	Extroversion rating (0–100)	Ranks of the data collected	
			Annual salary (£s)	Extroversion rating (0–100)
a	50 000	85	9	9
b	55 000	65	10	6.5
c	10 000	95	1	11.5
d	40 000	60	7	5
e	30 000	45	5	4
f	15 000	30	2	1
g	25 000	40	4	3
h	60 000	90	11	10
i	35 000	70	6	8
j	20 000	35	3	2
k	70 000	95	12	11.5
l	45 000	65	8	6.5

Calculate the mean rating of extroversion. Show your workings and present your finding to 2 significant figures.

[3]

129.

Extroversion = Extra money?

A psychologist investigated whether how much a person earns is related to the personality trait of extroversion. Participants were asked to provide details of their current annual salary. They then completed a general lifestyle questionnaire, which included a question asking them to rate how extrovert they were on a scale of 0 ('not at all extrovert') to 100 ('extremely extrovert'). The table below presents the data collected in the study.

Participant	Annual salary (£s)	Extroversion rating (0–100)	Ranks of the data collected	
			Annual salary (£s)	Extroversion rating (0–100)
a	50 000	85	9	9
b	55 000	65	10	6.5
c	10 000	95	1	11.5
d	40 000	60	7	5
e	30 000	45	5	4
f	15 000	30	2	1
g	25 000	40	4	3
h	60 000	90	11	10
i	35 000	70	6	8
j	20 000	35	3	2
k	70 000	95	12	11.5
l	45 000	65	8	6.5

Explain why it is more appropriate to use the mean, rather than the median, to calculate central tendency in this study.

[2]

130. Which of these best describes what the standard deviation informs us?

A dispersion around the mean
B dispersion around the median
C dispersion around the mode
D dispersion around the range

Your answer ☐ **[1]**

131. In the study by Levine *et al*. (2001) investigating cross cultural differences in helping behaviour, what level of data was recorded for the helping measure that involved responses to a dropped pen?

A both ordinal and interval
B interval
C nominal
D ordinal

Your answer ☐ **[1]**

132. Which decimal is represented by the fraction $\frac{1}{20}$?

A 0.2
B 0.02
C 0.5
D 0.05

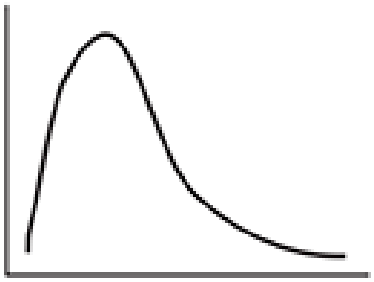
Your answer ☐ **[1]**

133. If the variance for a set of data is 16, what is the standard deviation?

A 3
B 4
C 8
D 256

Your answer ☐ **[1]**

134. What type of distribution curve is this?



A negatively skewed
B non-skewed
C normal
D positively skewed

Your answer ☐ **[1]**

135. Which of these is an assumption of parametric inferential statistical tests?

A data is asymmetrically distributed in the population
B data is negatively skewed in the population
C data is normally distributed in the population
D data is positively skewed in the population

Your answer ☐ **[1]**

136. What is 0.006089 written to two significant figures?

A 0.00
B 0.0060
C 0.0061
D 0.61

Your answer ☐ **[1]**

137. What name refers to data before any analysis is done?

A nominal
B parametric
C primary
D raw

Your answer ☐ **[1]**

138. Which of these is not a measure of dispersion?

A variance
B median
C range
D standard deviation

Your answer ☐ **[1]**

139. When data contains anomalies, which of these is it best to use?

A mean
B median
C mode
D percentage

Your answer ☐ **[1]**

140. Which inferential statistical test uses degrees of freedom (df) to find the critical value?

A Chi-square
B Mann-Whitney U Test
C Spearman's Rho
D Wilcoxon Signed Ranks Test

Your answer ☐ **[1]**

141. In a study investigating the difference in ratings (on a scale of 0 to 100) of self-esteem between males and females, which would be the appropriate inferential test to use to analyse the data?

A Binomial Sign Test
B Chi-square
C Mann Whitney U test
D Wilcoxon Signed Ranks test

Your answer ☐ **[1]**

142. Which of these indicates the probability is less than 5%?

A $p<0.5$
B $p>0.05$
C $p<0.05$
D $p>0.005$

Your answer ☐ **[1]**

143. Which of these features is needed to choose a test of statistical significance?

A central tendency
B level of data
C sample size
D standard deviation

Your answer ☐ **[1]**

144. Explain **one** way in which Freud's study of Little Hans could be criticised for its lack of reliability.

[3]

END OF QUESTION PAPER

Mark scheme

<table>
<tr><th colspan="3">Question</th><th>Answer/Indicative content</th><th>Marks</th><th>Guidance</th></tr>
<tr><td>1</td><td></td><td></td><td>A</td><td></td><td>range</td></tr>
<tr><td></td><td></td><td></td><td>Total</td><td>1</td><td></td></tr>
<tr><td>2</td><td>a</td><td></td><td>B</td><td></td><td>15</td></tr>
<tr><td></td><td>b</td><td></td><td>A</td><td></td><td>5</td></tr>
<tr><td></td><td></td><td></td><td>Total</td><td>2</td><td></td></tr>
<tr><td>3</td><td></td><td></td><td>Outline one way you could uphold the ethical consideration of respect in this study.

Under BPS ethical considerations ...

respect = informed consent, right to withdraw and privacy/confidentiality.

For example, one way to do this would be to inform primary school children at the beginning of the study on concentration that they can withdraw at any time.

One way to uphold ethical consideration of respect clearly presented in context.

Attempt to outline one way to uphold ethical consideration of respect (whether in context or not)

The candidate has not provided any creditworthy information</td><td>Max 2
2
1
0</td><td>Context = concentration, slippers, shoes, learning, primary school children etc.

Mere identification of the ethical guideline gains no credit.

Informed consent is only creditworthy if linked to parental/headteachers consent.

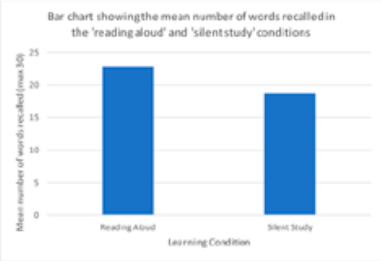
NB: Only first response is marked.

<u>Examiner’s Comments</u>

This question differentiated well as answers covered the full range of the mark scheme. The best responses identified relevant ethical guideline that comes under the principle of respect and explained how they would uphold it in the context of the scenario presented.

A significant proportion of candidates got zero marks because they chose an ethical consideration that is not under respect such as debrief or they did not consider the implications of the scenario, i.e. they referred to gaining consent from the children themselves. Candidates need to be aware that informed consent can only be gained from a consenting adult - either the participant themselves or a parent/guardian.</td></tr>
</table>

			Total	**2**	
4			**Explain two factors that could affect the external validity of this study.** Likely answers: size/diversity/representativeness of sample; ecological validity of task set to assess concentration, ecological validity of the setting, etc. **3 marks for each factor outlined ...** Clear explanation of how external validity could be affected in context. Clear explanation of how external validity could be affected, but not in context. **OR** attempted explanation of how external validity could be affected in context. Identification of relevant factor/type of external validity or weak attempt to explain how external validity could be affected (whether in context or not). The candidate has not provided any creditworthy information	Max 6 [3+3] 3 2 1 0	Context = concentration, slippers, shoes, learning, primary school children etc. NB. If candidates refer to types of external validity that are not listed on the specification these can be credited, e.g. temporal validity (teaching methods change over time) **Examiner's Comments** Good responses accurately identified factors which could impact external validity. The strongest referred to population and ecological validity as separate points with contextualisation. A notable number of candidates was unable to differentiate between internal and external validity, which meant low (or zero) marks.
			Total	**6**	
5			**Outline one conclusion that can be made from the raw data presented in this table.** Conclusions could include: • Reading aloud seems to facilitate memory, perhaps because the act of reading enables the words to be practiced more and processed at a deeper level • There are some individual differences, so reading aloud does not improve memory for everyone, indicating cognitive processes work differently for different people. Accept any other appropriate conclusions here. Clear, detailed conclusion in context (or supported by data). Clear, detailed conclusion but not in context. **OR** attempt in context. Brief and/or weak attempt (whether in context or not)	3 3 2 1 0	Context = reading aloud, silence, recall, memory, etc. A conclusion must be an interpretation/application of the findings / data (not simply a statement of the result(s) obtained). **Max 1 mark** for presentation of a finding (involving comparison of data) with no interpretation/explanation of it. Zero marks if just data is given. **NB: Only first response is marked.** **Examiner's Comments** Candidates really struggled to differentiate between findings and conclusions. Although almost all candidates included context in their

			The candidate has not provided any creditworthy information		response, the vast majority of those simply noted a finding and did not explain what the implications of the finding are for memory/recall. For those who attempted to explain the implications, often they made a general statement that memory is improved, rather than attempting to explain why reading aloud may improve memory. **Assessment for learning** Candidates need to practise forming a conclusion. Phrases such as: 'this is because...', 'this suggests that...' might be useful.
			Total	**3**	
6			**Draw a fully labelled bar chart showing the mean number of words recalled in each condition. Plot the means to two significant figures.**  **1 mark** is awarded for **correctly** calculating then **presenting** by value each bar representing the **mean** number of words recalled (to 2 significant figures) in the 'reading aloud' and 'silent' conditions **1 mark** is awarded for **clear labelling of the x** axis 1 mark is awarded for **clear labelling of the y** axis including measurement (which must start at 0). 1 mark is awarded for a fully operationalised title.	Max 4 [1+1+1+1] 1 1 1 1	Context = reading aloud, silence, recall, memory, etc. Mean values for each condition need to be calculated first. Mean for **reading aloud** condition = **23** to **2** sf. Mean for **silent** condition = **19** to **2** sf. Title must include both variables: (mean) number of words recalled and reading aloud/silent study conditions). Response must make it clear that this is the mean number of words (max 30) recalled in either title or y axis, if not max 3. Labels on axes must be clear. X axis - reading aloud/silent study Y axis - mean number of words recalled - and measurement must start at 0 (does not need to go up to 30 and can go beyond 30). If two bars are together (or if the bar touches y axis) do not award

<table>
<tr><td></td><td></td><td></td><td></td><td></td><td>mark for correct presentation of data.

<u>Examiner's Comments</u>

Almost all candidates knew how the bar chart is supposed to look. However, a significant proportion of candidates lost at least 1 mark. The most common mistakes included: not plotting the bar to two significant figures, not operationalising conditions in the title, incorrect labelling of the y axis by stating the number of words recalled rather than mean number. Compared to previous series, fewer candidates placed the bars together.</td></tr>
<tr><td></td><td></td><td></td><td>Total</td><td>4</td><td></td></tr>
<tr><td>7</td><td></td><td></td><td>Give one reason why the Mann Whitney U test is the appropriate inferential test to use to analyse the data from this study.

Any one reason (in bold) in context from:

• It is a test for independent measures design, and this study had different participants in the reading aloud condition compared to the silent condition.
• It is a test that uses <u>at least</u> ordinal data and this study had ordinal data as it was scores out of 30 in a memory test.
• It is a test that assesses differences between conditions, and this study assessed differences in memory between the reading aloud and silent conditions.

One appropriate reason in context

One appropriate reason but not in context. OR attempt to give one appropriate reason in context.

The candidate has not provided any creditworthy information.</td><td>Max 2
2
1
0</td><td>Context = reading aloud, silence, <u>recall of words</u>,memory, etc.

'6 participants' not enough for context.

NB: Only first response is marked

<u>Examiner's Comments</u>

Candidates were able to identify at least one reason for the use of Mann-Whitney U test, however the use of context was inconsistent. The most common mistake was to give two reasons and not to include context, which resulted in only 1 mark.</td></tr>
<tr><td></td><td></td><td></td><td>Total</td><td>2</td><td></td></tr>
<tr><td>8</td><td></td><td></td><td>Before using the formula for the Mann Whitney U test, the data obtained must be ranked. In the result two participants have the same</td><td>Max 3
3
2</td><td>Context = reading aloud, silence, recall, memory, etc.

NB. Specific scores could be</td></tr>
</table>

score of 24. Explain how this is dealt with when ranking the data.

As the two scores of 24 were the same, so they receive the same rank of 9.5 as the ranks 9 and 10 have been shared, i.e. 9 + 10 / 2 = 9.5	
Clear explanation in context of how having the same scores is dealt with including references to the same rank and a way of finding the same rank.	
Clear explanation of how having the same scores is dealt with including references to the same rank (or simply 9 and 10) and a way of finding the same rank (no context).	**OR** Explanation of why two ranks are the same in context.
Attempt to explain why two ranks are the same (whether in context or not).	
The candidate has not provided any creditworthy information	

1
0

credited as context.

Examiner's Comments

Most candidates were able to identify the issue of tied ranks, which was contextualised. Not all explained how to find the shared rank (by either calculating the mean of ranks or finding median/middle value).

			Total	**3**	

9 a

Calculate the U value for the Mann Whitney U test for the data collected in this study. Show your workings. You may use the formula presented below.

U = the smaller of U_1 and U_2

Where U_1 is ... and U_2 is ...

$$U_1 = R_1 - \frac{n_1(n_1+1)}{2} \qquad U_2 = R_2 - \frac{n_2(n_2+1)}{2}$$

1 mark for each of the following correct / evident in answer ..
1 mark for sum of ranks for reading aloud condition calculated correctly (46.5)
1 mark for sum of ranks for silent study group calculated correctly (31.5).
1 mark for correct calculation of U_1 (25.5) value OR correct calculation of U_2 value (10.5).
1 mark for <u>all</u> workings of U_2 value shown.
1 mark for choosing the final U value.
The candidate has not provided any creditworthy information.

Max 5
[1+1+1+1+1]
1
1
1
1
1
0

U_1:	**U_2:**
$= 46.5 - \frac{6\,(6+1)}{2}$	$= 31.5 - 6\,\frac{(6+1)}{2}$
$= 46.5 - \frac{6 \times 7}{2}$	$= 31.5 - \frac{6 \times 7}{2}$
$= 46.5 - \frac{42}{2}$	$= 31.5 - \frac{42}{2}$
$= 46.5 - 21$	$= 31.5 - 21$
= 25.5	**= 10.5**
U_2 is the smaller of the two scores so U = 10.5 (candidates may indicate this by circling the correct U value)	
NB. Accept alternative workings for U_1 and U_2 if the correct answer is calculated.	

Examiner's Comments

The majority of candidates responded really well to this question showing full workings and calculatingcorrect U value. Errors on this question typically stemmed from incorrectly adding the ranks for the conditions. A fair number of candidates lost a mark as they did not indicate the final U value.

	b		**How is the critical value used to determine if the findings are statistically significant?** For one mark this must refer to comparison with calculated value. It is **compared** to the appropriate/correct **calculated/observed** value.	1	Accept answers related directly to the Mann-Whitney test. **Examiner's Comments** Overall, the question was answered well. Some candidates lost marks as they referred to expected rather than observed/calculated value.
			Total	**6**	
10			**Explain what p > 0.05 would mean if it appeared as part of the significance statement when reporting the findings from this study.** Answer could refer to the following features: 1. Greater than 5% probability that the results are due to chance. 2. Null hypothesis is accepted AND/OR the alternative hypothesis is rejected. 3. The results are not significant/there is no significant difference. 4. Less than 95% confident that the results are significant. 5. Observed value was greater than the critical value, therefore results are not significant. 2 correct features in context. 2 correct features, but not in context. **OR** 1 correct feature in context. 1 correct feature, not in context. The candidate has not provided any creditworthy information	Max 3 3 2 1 0	Context = reading aloud, silence, recall, memory, etc. Example 3-mark answer: The results are **not significant**, so **the null hypothesis is accepted** (i.e. there is no difference in memory when words are read aloud compared to when studied in silence). Example 2-mark answer: There is **no significant difference** between how many words are recalled when words are read aloud or studied in silence. Example 1-mark answer: The **null hypothesis should be accepted**. **Examiner's Comments** The vast majority of candidates misread the greater than sign, which meant that their interpretation of significance was incorrect. It was quite common for candidates to incorrectly state that the research is significant. Some candidates did not understand what a significance statement is and instead provided suggestions as to how to look up a critical value. Exemplar 3 There is no significant difference as there is a more than 5% probability the difference between correct recall of words in a silent study or reading aloud condition are due to chance. Accept the null hypothesis and reject the alternate as the conditions of study do not have a significant effect on correct recall /3 on a memory test.

<table>
<tr><td></td><td></td><td></td><td></td><td></td><td>Exemplar 3 shows correct interpretation of the significance statement <u>in context</u>.</td></tr>
<tr><td></td><td></td><td></td><td>Total</td><td>3</td><td></td></tr>
<tr><td>11</td><td>a</td><td></td><td>Explain one strength of using quantitative data in this study.

Strengths could include:

• Easier data analysis and comparison across conditions (reading aloud vs silent)
• Easier to record data
• Easier to interpret.
• Objective
• Easier to check for consistency etc.

Clear explanation of strength in context.

Clear explanation of strength but not in context. OR attempted explanation of strength in context.

Brief and/or weak attempt to explain 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 = reading aloud, silence, recall, memory, etc.

NB: Only first response is marked.</td></tr>
<tr><td></td><td>b</td><td></td><td>Explain one weakness of using quantitative data in this study.

Weaknesses could include:

• Doesn’t inform us about reasons why there is a difference or not in memory between reading aloud and studying words in silence, because it lacks detail.
• Construct validity issues (simplifying complex behaviours to a score) etc.

Clear explanation of weakness in context.

Clear explanation of weakness but not in context. OR attempted explanation of weakness in context.

Brief and/or weak attempt to explain 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 = reading aloud, silence, recall, memory, etc.

NB: Only first response is marked.

<u>Examiner’s Comments</u>

As with previous strengths/weaknesses questions, the best responses followed a PEC structure, although sometimes the comment did not link back to the point identified and therefore did not get the final mark. Some simply stated ‘therefore more/less valid and more reliable’ rather than being specific.</td></tr>
<tr><td></td><td></td><td></td><td>Total</td><td>6</td><td></td></tr>
</table>

<table>
<tr>
<td>12</td>
<td></td>
<td></td>
<td>Outline two ways that an independent measures design could affect the validity of this study.

Answers could include:

• No order effects from having already studied words in previous condition, so increased validity.
• Individual differences in memory may have influenced the findings irrespective of learning condition (reading aloud vs silent), so decreased validity.
• Can keep the learning material (words) the same in each condition, so words in one condition are not easier or harder to remember than others, so increased internal validity.

Accept any other appropriate strengths or weaknesses.

Identified strength/weakness explained in relation to impact on validity and in context.

Some understanding of impact on validity, not in context.

The candidate has not provided any creditworthy information.</td>
<td>4
[2 + 2]
2
1
0</td>
<td>Context = reading aloud, silence, recall, memory, etc.

Evaluation points are likely to be based on those in part (a) but other points can be used. However, do not credit a point that is just a repetition of a strength/weakness given in Q32a, unless it is specifically related to the effect on validity.

Accept both positive and negative impacts on validity.

Do not credit ‘eradicates/eliminates/rules out demand characteristics’.

Do not accept population validity.

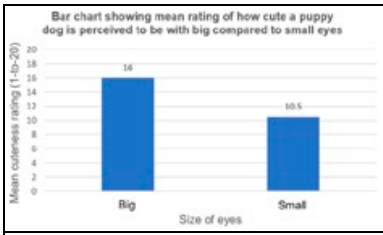
Do not credit issues that relate to reliability (e.g. therefore increases consistency).

<u>Examiner’s Comments</u>

Common responses to this question included reference to demand characteristics, order effects and participant variables. These were clearly explained within the context of the scenario. Weaker responses did not make a judgement as to how validity would be affected, i.e. increased or decreased. A number of candidates gave answers focusing on the use of a lab experiment rather than the independent measures design (specifically making references to ecological validity), which was not creditworthy.</td>
</tr>
<tr>
<td></td><td></td><td></td><td>Total</td><td>4</td><td></td>
</tr>
<tr>
<td>13</td><td></td><td></td><td>A</td><td></td><td>inter-rater</td>
</tr>
<tr>
<td></td><td></td><td></td><td>Total</td><td>1</td><td></td>
</tr>
<tr>
<td>14</td><td></td><td></td><td>C</td><td></td><td>$p < 0.025$</td>
</tr>
<tr>
<td></td><td></td><td></td><td>Total</td><td>1</td><td></td>
</tr>
</table>

15			C		Median **Examiner's Comments** A lot of candidates responded well to this question, however some candidates opted for D incorrectly.
			Total	**1**	
16			D		81 **Examiner's Comments** A lot of candidates responded well to this question, however some candidates opted for A incorrectly
			Total	**1**	
17			B		it is based on the data collected
			Total	**1**	
18			D		58
			Total	**1**	
19			D		an observation study which records whether customers say 'thank you' or not to the checkout assistant in a shop
			Total	**1**	
20			**B**	1	Peer-review **Examiner's Comments** Many answered correctly but there was still a large number of incorrect answers.
			Total	**1**	
21			**B**	1	Social desirability **Examiner's Comments** The majority answered correctly.
			Total	**1**	
22			**D**	1	Secondary **Examiner's Comments** Most answered correctly.

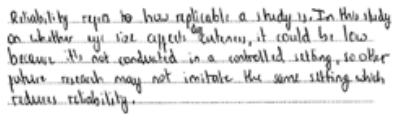
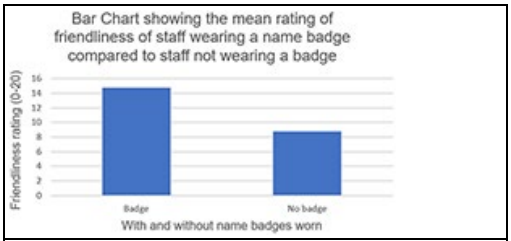
			Total	**1**	
23			**D**	1	29 **Examiner's Comments** Most answered correctly.
			Total	**1**	
24			**D**	1	Right to withdraw **Examiner's Comments** Nearly all answered correctly.
			Total	**1**	
25			**Identify one reason why the study might have low population validity.** **Answers could include:** • Low response rate. • People may be embarrassed and so not want to take part. • People may not complete the self-report. • Only people with certain characteristics might take part. • A small sample size might be collected. One correct reason identified The candidate has not provided any creditworthy information	Max 1 1 0	Context not needed here, but if used as part of an example provided by the candidate it is ok (context = mental health, grateful, feelings, etc.) **Examiner's Comments** Many candidates did not get this question correct; they did not identify an issue from the study in general and focused too heavily on the statistics given in the stem of '1 in 4 people experience mental health issues'. Candidates need to understand that this section on design and response is about a response to the stem, not a requirement for them to evaluate the stem.
			Total	**1**	
26			**Outline how you would address two ethical considerations in the design of this study.** **Answers could include:** • Respect (consent, withdrawal and confidentiality) • Competence (having a researcher trained in mental health conduct the interview) • Responsibility (protection from harm and debrief)	Max 6 [3+3] 3 2 1 0	Context = mental health, grateful, feelings, etc. Candidates do not need to use the headings Respect, Competence, Responsibility, Integrity. Candidates must outline two separate ethical considerations. **Examiner's Comments** Candidates were often able to give two separate ethical considerations in context however often they didn't

			• Integrity (maintain high standards and no use of deception). 3 marks for each way to address an ethical consideration: 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		explain 'how' these considerations would be employed.
			Total	**6**	
27			**Draw a fully labelled bar chart displaying the mean rating for how cute the puppy was perceived to be in each condition for this study.**  **1 mark** is awarded for correctly calculating then presenting by value each bar representing the mean rating of how cute the puppy dog is perceived to be in each condition **1 mark** is awarded for named categories/clear labelling of the x axis **1 mark** is awarded for clear labelling of the y axis (must be contextualised to 'cuteness') **1 mark** is awarded for a clearly contextualised title	Max 4 [1+1+1+1] 1 1 1 1	If a clearly contextualised title has been included, this could make up for any shortfall of labelling on an x and y axis. Candidates must make clear in the title that this bar chart is showing the 'mean'. **<u>Examiner's Comments</u>** Many candidates had a good understanding of how to draw a bar chart. Candidates often lost marks due to not fully operationalising their title or the label on their y axis. There was a significant number of candidates who presented all raw data on their graph rather than their mean.
			Total	**4**	
28			**Outline <u>two</u> conclusions that can be made from the data collected in this study.** **<u>Conclusions could include:</u>** • A puppy with big eyes is regarded as cuter than a puppy with smaller eyes perhaps because the larger eyes make us focus on the dog for longer which makes it more familiar. • A puppy with big eyes is regarded as cuter than a puppy with smaller eyes as the larger	Max 6 [3+3] 3 2 1 0	Context = puppy, dog, cute, etc. Clear (explicit) interpretation of findings (not simply stating a finding) is required to achieve beyond 1 mark. **<u>Examiner's Comments</u>** This question gained a variety of responses with few candidates accessing the full range of marks. Most candidates were restricted in their marks as they only identified

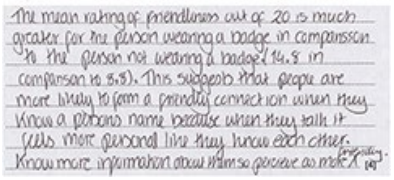
			eyes are more distinctive and therefore perceived as more human like. Accept any other appropriate conclusions here. **3 marks for each conclusion** Clear, detailed conclusion in context Clear, detailed conclusion but not in context **OR** Attempt at conclusion in context Brief and/or weak attempt (whether in context or not) **OR** Simply stating a finding The candidate has not provided any creditworthy information		findings rather than giving conclusions. **Assessment for learning** Candidates would benefit from being given opportunities to examine raw data tables and practise extrapolating conclusions that could be drawn from them by looking at the nature of the data and any patterns that can be inferred. These skills can also be illustrated through data collected through practical investigations conducted in preparation for this exam. The difference in findings and conclusions from the core studies component could be highlighted to assist in demonstrating to candidates how to differentiate between findings and conclusions.
			Total	**6**	
29			**Outline <u>one</u> strength and <u>one</u> weakness of the use of quantitative data in this study.** **<u>Strengths could include:</u>** • Allows participants to be more precise about level of cuteness as the scale (1 – 20) is quite broad and has verbal labels at each end (not very cute and extremely cute). • Easy to analyse and compare findings on ratings of cuteness of puppy. • Easy to present findings and plot visually (e.g., in a bar chart displaying mean cuteness rating across each condition) therefore easier to compare. **<u>Weaknesses could include:</u>** • Numbers in the ratings lack detail therefore don't indicate why participants thought the puppy was cute or not.	Max 6 [3+3] 3 2 1 0	Context = puppy, dog, cute, etc. **<u>Examiner's Comments</u>** Candidates were able to access a wide range of marks on this question, with most accessing the highest band. Reponses that couldn't access top band usually lacked context or elaboration on their evaluation point.

			• Some participants may interpret the scale differently to others (e.g. one person's idea of what 7 on the scale is, might be different to another person's idea of what 7 indicates in terms of cuteness). **For each strength/weakness:** Clear outline of strength/weakness in context Attempted outline of strength/weakness in context **OR** Clear outline of strength/ weakness, but not in context Brief and/or weak attempt to outline strength/weakness (whether in context or not) The candidate has not provided any creditworthy information		
			Total	**6**	
30	a		**The inferential statistical test used to analyse the data collected in this study is the Wilcoxon Signed Ranks test. Explain how data is ranked for this test in this study.** Ranking involves assigning data values that denote their position in an ordered sequence (i.e. putting the values in order from lowest to highest). For the data in this study, this involves ***ranking the difference*** between the ratings of cuteness for each condition (i.e. not ranking the scores/ratings themselves). Clear explanation in context of what ranking involves and how it is done using the Wilcoxon test for the data in context. Explanation of what ranking in general involves and explanation of how ranking is done using the Wilcoxon test, not in context. **OR** Explanation of what ranking in general involves in context. **OR** Explanation of how ranking is done using the Wilcoxon test in context Explanation of what ranking in general involves OR Explanation of how ranking is done using the Wilcoxon test The candidate has not provided any creditworthy information	Max 3 3 2 1 0	Context = puppy, dog, cute, etc. Convention is that the lowest score is assigned rank one and then proceed on this basis (but can be done vice versa (highest score is assigned rank 1). **<u>Examiner's Comments</u>** Many candidates misunderstood this question and outlined why the Wilcoxon test would be used. A significant number of candidates were however able to answer this question effectively but often missed out context for the final mark.
	b		**Using the Wilcoxon Signed Ranks test, calculate the value of T for the data collected in this study by completing the table below.**	Max 3 1+1+1 1 1	If candidate has the correct answer but not shown any workings, award 1 mark.

			Formula / process: T = the sum of the ranks of the differences with the least frequent sign Ratings of how cute a puppy dog is with big eyes compared to small eyes (1 = 'not very cute', 20 = 'extremely cute') Participant / Condition (a) with big eyes / Concition (b) with small eyes / Difference (a) – (b) / Ranks of differences / T value (sum of the ranks of the differences with the least frequent sign) a 18 7 11 5.5 b 14 3 11 5.5 — 1 + 2 = 3 c 12 14 -2 1 — T = 3 d 17 20 -3 2 e 19 9 10 4 f 16 10 6 3 1 mark for the correct answer (T=3) 1 mark for Ranks of differences 1 mark for working out the Difference (a) – (b) The candidate has not provided any creditworthy information	1 0	**Examiner's Comments** Some candidates were able to achieve full marks on this question. Many candidates were able to complete the difference column but nothing further. Those achieving 2 marks were normally able to complete the difference and get the final T value correct - however their ranking was often incorrect due to ranking minus 3 as 1 and minus 2 as 2. **Assessment for learning** Candidates would benefit from having opportunities to calculate practice inferential statistical tests. It would also be useful to remind them that in a Wilcoxon Signed Ranks test they are to ignore the minus signs when ranking the data.
			Total	**6**	
31			**Outline one reason why the reliability of the data collected in this study could be low.** **Answers could include:** • Some participants interpreting the rating scale for cuteness differently to others. • If a different type/breed of dog was used in the photo with big eyes compared to the photo with small eyes. • Type of people used in the sample (if dog owners or not) because they might interpret the photos differently. Clear outline in context Clear outline, but not in context / An attempt in context Brief and/or weak outline, whether in context or not The candidate has not provided any creditworthy information	Max 3 3 2 1 0	Context = puppy, dog, cute, etc. To score above 1 mark, the candidate must show an understanding of the concept of reliability (not just giving a generic example that could also affect the validity). No credit for reference to population validity or generalisability. **Examiner's Comments** Many candidates did not achieve marks on this question, often confusing reliability with validity. The more common incorrect responses focused on social desirability bias/population validity. Exemplar 3

					 Exemplar 3 demonstrates one possible 3 mark answer to this question. This candidate has identified the control of a setting and how if it is not kept the same then this would reduce the reliability of the data collection. Their response is also in context.
			Total	**3**	
32			**C**	1	Whether the measurement of a variable can predict outcomes for similar behaviour in the future **Examiner's Comments** Most answered this correctly.
			Total	**1**	
33			**D**	1	22% **Examiner's Comments** Nearly all answered this correctly
			Total	**1**	
34			**D**	1	Parametric statistical tests **Examiner's Comments** A number of candidates were unable to give a correct response.
			Total	**1**	
35	a		 1 mark is awarded for correct presentation of data 1 mark is awarded for clear labelling of the x axis 1 mark is awarded for clear labelling of the y axis including measurement must start at 0 (does not need to go up to 20) 1 mark is awarded for fully operationalised title	Max 4	Context = friendliness, friend, name, badge etc Title must include both variables ([mean] rating of friendliness and name badges/no name badges). Response must make it clear that this is the **mean** rating of friendliness in either title or y axis, if not max 3 Labels on axes must be clear. X axis – badge/not wearing badge (or similar wording) Y axis – rating of friendliness and

			<table><tr><td>All 4 features from above - 4</td></tr><tr><td>Any 3 features from above - 3</td></tr><tr><td>Any 2 features from above - 2</td></tr><tr><td>Any 1 feature from above - 1</td></tr><tr><td>The candidate has not provided any creditworthy information - 0</td></tr></table>		measurement must start at 0 (does not need to go up to 20 and can go beyond 20). If two bars are together do not award mark for correct presentation of data. **<u>Examiner's Comments</u>** The majority of candidates plotted the data accurately and labelled the X axis correctly. Some responses did not get the mark for the title as either it was missing or did not include both conditions of the independent variable in the title or not mentioning mean rating of friendliness. Some responses did not get the mark for plotting the data as it was either incorrect or the bars of the bar chart were placed together. The majority of candidates achieved the mark for the Y axis although a few did just state 'mean' or 'mean rating' which did not achieve the mark for the Y axis.
	b		<table><tr><td colspan="2">For example: Staff wearing a name badge were perceived as more friendly compared to those not wearing a name badge. This is perhaps because it was more personal and made customers feel that they knew the person better and could interact and relate to them more.</td></tr><tr><td colspan="2">Clear conclusion in context that refers to the findings. - 4</td></tr><tr><td>Attempted conclusion in context that refers to the findings. - 3</td><td>Clear conclusion in context and attempt to refer to findings. - 3</td></tr><tr><td>Attempted conclusion in context. - 2</td><td>Clear conclusion not in context - 2</td></tr><tr><td colspan="2">Unclear and/or brief conclusion (whether in context or not) - 1</td></tr><tr><td colspan="2">The candidate has not provided any creditworthy information - 0</td></tr></table>	Max 4	Context = friendliness, friend, name, badge etc Note: a conclusion must be an interpretation/application of the findings / data (not simply a statement of the result(s) obtained) No credit for stating just the mean scores (e.g. the mean friendliness score for wearing a badge is 14.8=0) Max 2 marks if only findings presented. For example - Staff wearing a name badge were perceived as more friendly (1) compared to staff not wearing a name badge(1). Needs to state the direction of the findings. If not, max 1 if just findings presented e.g. there is a difference in perceived friendliness rating if

					the staff wore a badge or not. **<u>Examiner's Comments</u>** The majority of candidates were able to make a brief statement about what the results show (i.e. that shop assistants wearing a name badge are perceived as more friendly than those who don't) but many struggled to elaborate beyond the findings. Better responses tended to say that knowing someone's name means you feel you know them more personally or you feel more comfortable to speak with the shop assistant. **Assessment for learning** It would be helpful to candidates to practise differentiating between a finding and a conclusion. This could be practised through the core studies where the results of the study (the findings) can be given and then the conclusions outlined (what can be inferred from the results). Exemplar 3  The mean rating of friendliness out of 20 is much greater for the person wearing a badge in comparison to the person not wearing a badge (14.8 in comparison to 8.8). This suggests that people are more likely to form a friendly connection when they know a persons name because when they talk it feels more personal like they know each other. Know more information about them so perceive as more X Exemplar 3 shows a response that received full marks. The response outlines the difference in the findings but then makes inferences about these differences and outlines why participants may have found a shop assistant wearing a name badge to be more friendly.
			Total	**8**	

<table>
<tr>
<td>36</td>
<td></td>
<td></td>
<td>
<table>
<tr><td colspan="2">Because the data contains outliers (e.g. in the wearing badge condition the rating of 5, which is much lower than any other rating) and the median is less sensitive measure of central tendency when there are outliers.</td></tr>
<tr><td colspan="2">Clear explanation in context - 2</td></tr>
<tr><td>Attempted explanation in context - 1</td><td>Clear explanation no context - 1</td></tr>
<tr><td colspan="2">The candidate has not provided any creditworthy information - 0</td></tr>
</table>
</td>
<td>Max 2</td>
<td>Context = friendliness, friend, name, badge etc OR 5 AND/OR 18

1 mark response – identifying the anomaly/extreme value
2 marks – example from the data (5 or 18) and reference to the median not being affected by anomalies (and the mean is).

Award credit to -
Median is the most appropriate with ordinal data. (1) and reason why-
This is because ordinal data is a less precise measurement than interval data.(1) OR
Median is less sensitive and is therefore more appropriate for ordinal data.(1)

<u>Examiner’s Comments</u>

The majority of candidates were able to achieve 1 mark by correctly identifying that outlier/extreme scores would affect the mean but not the median. Better responses then identified the extreme score in the data (5 or 18). However, many responses did not achieve this second mark for not giving examples of these extreme scores within the data set (5 or 18). It was common for responses to identify an incorrect extreme score such a one of the rankings of friendliness rather than the ratings of friendliness. A common error in responses that did not gain marks was to state that the median is better due to the large range of data.</td>
</tr>
<tr>
<td></td><td></td><td></td><td>Total</td><td>2</td><td></td>
</tr>
<tr>
<td>37</td>
<td>a</td>
<td></td>
<td>U1 = 135 – 110/2 = 80

U2 = 75 – 110/2 = 20 (U1 and U2 are inter-changeable in terms of order candidates choose to do them)

U = 20</td>
<td>Max 5</td>
<td>= 20 is creditworthy (rather than U=20) or circling the correct U value.

<u>Examiner’s Comments</u>

The majority of candidates responded to this question well and identified the correct U value and</td>
</tr>
</table>

			1 mark for each of the following correct / evident in answer . • 135 • 75 • correct calculation of U2 value obtained OR correct calculation of U1 value • All workings shown for U2 • Correct U value (20) All 5 features from above - 5 4 features from above - 4 3 features from above - 3 2 features from above - 2 1 feature from above - 1 The candidate has not provided any creditworthy information - 0		showed their workings fully. Errors on this question typically stemmed from incorrectly adding the ranks for the conditions and therefore coming out with the incorrect U value. Other errors in terms of the analysis came from using the ratings, not the rankings, to find the R1 and R2 values. A few responses carried out the calculation but did not identify U=20 (as the smallest of the two calculated values) so were unable to achieve full marks for this question. There were some who did not attempt this question.
	b		Found by using the sample size in each condition Table critical value = 23 Explanation provided and correct critical value stated - 2 Explanation only provided - 1 **OR** critical value only provided - 1 The candidate has not provided any creditworthy information - 0	Max 2	Credit N1/N2 or Na/Nb Credit Na = 10; Nb = 10 as the explanation 1 mark for indicating the critical value in the table and not in the answer space **Examiner's Comments** Candidates demonstrated good understanding of critical value tables and were able to identify 23 as the critical value and explain how this was found (by using the number of participants, 10, in each condition). A common error was reading from 5 (for the 5%level of probability) and then either 20 (total participants) or 10 (participants in each condition) instead of 10 and 10.
	c		U = 20, n1 = 10, n2 = 10, p<0.05 OR U=20, critical value = 23, 20<23, p<0.05 **OR** In words ... e.g. the alternative hypothesis (that there would be/is a significant difference in perceived friendliness between	Max 2	Credit N1/N2 or Na/Nb Cannot credit any p value other than .05 or 5% For worded statements 1 mark for each of the following:

			shop assistants wearing a badge and those who don't) is supported. This is because the calculated value (20) is less than the critical value (23). Correctly written significance statement - 2 **OR** written in words rather than a formal statement - 2 Just stating p<0.05 - 1 **OR** weak and/or brief written response - 1 The candidate has not provided any creditworthy information - 0		• Comparison between calculated value and critical value • Significance level (p<0.05)/'it is significant'/alternative hypothesis is supported/null hypothesis is rejected Max 1 mark if candidate contradicts themselves about whether it is significant but does state the correct comparison of calculated value and critical value. **Examiner's Comments** The vast majority of candidates presented a written statement rather than a mathematical one and correctly compared the critical and calculated value and stated there was a significant difference. Some candidates did not make a comparison between the calculated and critical value so could not achieve full marks for this question. The most common errors were from earlier mistakes being carried forward e.g. incorrect U value, and incorrect critical value that led to an incorrect significance statement in the response. A few candidates did have the correct response in Question 30 (a) and Question 30 (b) but then concluded that the results were not significant which was incorrect.
			Total	**9**	
38			Likely answers: lacks depth so we don't know reasons why wearing a name badge or not influences perceived friendliness; doesn't allow for consideration of other (extraneous) influences (e.g. physical features of person etc); May lack ecological validity as not how we would consider friendliness i.e. in number format Clear and precise explanation in context - 3 Clear explanation but not in context - 2 **OR** attempted explanation in context - 2	Max 3	Context = friendliness, friend, name, badge etc **Examiner's Comments** The majority of candidates focused on the difficulty of not being able to find out why shop assistants with badges were rated more friendly. Better responses explained why this was a problem such as lack of detail/depth when only quantitative data is collected and the effect this

			Brief and/or weak attempt (whether in context or not) - 1 The candidate has not provided any creditworthy information - 0		could have on the validity or usefulness of the results. Some candidates focused on the issue of not being able to rule out other extraneous variables, such as whether the shop assistant was smiling or their general behaviour towards participants and this was often done well. Weaker responses often did not explain why this extra insight is a benefit of qualitative data or a weakness of excluding it in this study.
			Total	**3**	
39			Credit answers which explain how reliability was maintained, increased or reduced. Likely answers: Consistency may be reduced by: Participants variables may produce an inconsistent/subjective view of the shop assistants who may be perceived as more friendly than others by different participants in different conditions. Different physical features of the shop assistants may be perceived as more friendly than others by different participants used in the badge wearing / non-badge wearing conditions. Different interpretations of the 0-20 rating scale for friendliness by different participants. Different behaviour of the shop assistants (this would probably be standardised as it was a research set up) when interacting with customers (and how long spent in their company, etc). Social desirability bias could have affected participants differently, (make sure it is linked clearly to lack of consistency, not accuracy). Consistency could have been increased	3 + 3	Context = friendliness, friend, name, badge, shop assistants etc **Annotation – up to 3 ticks per factor** No credit for reference to population validity or generalisability. Credit responses that refer to replicability. **<u>Examiner's Comments</u>** Most responses were able to identify at least one factor that could have affected the reliability of the data collected in this study. Common creditworthy responses focused on the rating scale being subjective (with different participants interpreting the scale differently), the shop assistants not acting in a consistent way, or following a standardised procedure, participant variables and the environment of the shop. However, the explanations were not always well developed with quite a bit of confusion between validity and reliability. Weaker responses were often brief with just a statement that the factor would affect either replicability or consistency of ratings without any explanations. Some responses would discuss issues that related to

<table>
<tr>
<td></td><td></td><td></td>
<td>by:

Standardised question increasing consistency as pre-set scale given.

Shop assistants are likely to behave in a similar way with each customer so this could be considered standardised behaviour of shop assistants allowing replicability.

Quantitative data is an objective measure so there is no inconsistency in the researcher interpreting the 0-20 ratings given by the participant.

Up to 3 marks for each factor

<table>
<tr><td colspan="2">Identification of a relevant factor with a clear and precise explanation in context - 3</td></tr>
<tr><td>Identification of a relevant factor with a clear explanation but not in context - 2</td><td>Identification of a relevant factor with an attempt at an explanation in context - 2</td></tr>
<tr><td colspan="2">Identification of a relevant (contextual) factor but no/very weak explanation - 1</td></tr>
</table>
</td>
<td></td>
<td>validity, and wouldn’t explain why it related to consistency in sufficient detail (e.g. demand characteristics, socially desirable responses). There were a significant number of responses that explained issues which were not creditworthy e.g. generalisability, or confuse the stem and suggest that the shop assistants were rating how friendly the customers were. Many of the responses were an attempt in context and not properly developed to get full marks.

Misconception

Many responses for this question were confused between reliability and validity.</td>
</tr>
<tr>
<td></td><td></td><td></td><td>Total</td><td>6</td><td></td>
</tr>
<tr>
<td>40</td><td></td><td></td><td>A</td><td></td><td>a type 1 error

<u>Examiner’s Comments</u>

The majority of candidates responded well to this question. Some candidates chose option B incorrectly.</td>
</tr>
<tr>
<td></td><td></td><td></td><td>Total</td><td>1</td><td></td>
</tr>
<tr>
<td>41</td><td></td><td></td><td>A</td><td></td><td>dispersion around the mean

<u>Examiner’s Comments</u>

The majority of candidates responded well to this question. Incorrect responses were varied.</td>
</tr>
<tr>
<td></td><td></td><td></td><td>Total</td><td>1</td><td></td>
</tr>
<tr>
<td>42</td><td></td><td></td><td>C</td><td></td><td>35%

<u>Examiner’s Comments</u>

The majority of candidates</td>
</tr>
</table>

					responded well to this question. Some candidates chose option D incorrectly.
			Total	**1**	
43			B		1/20 **Examiner's Comments** The majority of candidates responded well to this question. Some candidates chose option C incorrectly.
			Total	**1**	
44			B		3.5 **Examiner's Comments** The majority of candidates responded well to this question. Some candidates chose option C incorrectly.
			Total	**1**	
45			C		normal **Examiner's Comments** The majority of candidates responded well to this question. Incorrect responses were varied.
			Total	**1**	
46			B		Mann-Whitney U test **Examiner's Comments** The majority of candidates responded well to this question. Some candidates chose option C or D incorrectly.
			Total	**1**	
47			D		25.90 **Examiner's Comments** The majority of candidates responded well to this question. Some candidates chose option C incorrectly.

<table>
<tr><td></td><td></td><td></td><td>Total</td><td>1</td><td></td></tr>
<tr><td>48</td><td></td><td></td><td>A</td><td></td><td>concurrent

<u>Examiner's Comments</u>

The majority of candidates responded well to this question. Incorrect responses were varied.</td></tr>
<tr><td></td><td></td><td></td><td>Total</td><td>1</td><td></td></tr>
<tr><td>49</td><td></td><td></td><td>C</td><td></td><td>mode

<u>Examiner's Comments</u>

The majority of candidates responded well to this question. Some candidates chose option B incorrectly.</td></tr>
<tr><td></td><td></td><td></td><td>Total</td><td>1</td><td></td></tr>
<tr><td>50</td><td></td><td></td><td>What you are being driven by is the left-hand column of the grid ('details of the required features (RFs)'. That is always your starting point and 'locator' for the appropriate mark band before considering the other two columns ('justification of decisions made' and 'reference to own practical work').

<table>
<tr><th>Level of response</th><th>Details of required features (RFs) included</th><th>Justification of decisions made</th><th>Reference to own practical work</th></tr>
<tr><td>Good
12-15 marks</td><td>All 4 required features (RFs) addressed in context

Accurate and detailed knowledge and understanding of each feature in context

Good evidence of</td><td>Appropriate justification of all decisions and some is contextualised

Well-developed line of reasoning that is clear and logically structured</td><td>Explicit reference to own practical work and clear links between own work and the planned research for each required feature.
e.g. specific mention of aim or procedural features For top band</td></tr>
</table></td><td>Max = 15</td><td>Context = aroma, smell, scent, fragrance, litter, any example of litter (e.g. wrappers, rubbish) cleanliness etc
Annotations

RF (in the left column AND see next page for descriptors of the levels for description of the RF)
L4=Good; L3=Reasonable; L2= Limited;
L1= Basic

Annotate with CONT for context if RF in context. (under RF level annotation on left)

Tick for justification within the response Do not annotate the level, note the level of justification to decide on the mark given within the band

<u>Examiner's Comments</u>

Responses varied a lot to this extended question, with the majority of candidates finding it difficult to gain the higher band marks. The best responses were characterised by taking each of the four required features in turn and</td></tr>
</table>

						writing a separate paragraph relating to each one: first, demonstrating understanding of what was involved and how to address it for the research presented; next by justifying the decisions made regarding how to address it and finally, drawing on the candidates own experiences of conducting research themselves and how they learned from this how to conduct the research presented. All of this needed to be discussed in context to obtain marks in the highest band. It should also be noted that the candidates own experiences of conducting practical activities (especially the one using the same research method, which here was the laboratory experimental method) should be evident in their response to each required feature in terms of how this has helped inform their decision making for the planning of the current proposed research. There was also much variation in how candidates demonstrated knowledge and understanding of each of the individual required features (RFs). The best responses were characterised by first defining what the RF was/referred to (e.g. for RF1, defining the sampling method) before going on to describe exactly how the RF would be addressed in the proposed research. Often candidates did not provide enough detail. For example, in relation to RF1 just identifying the method without clearly describing how they would implement this sampling method in this research. For example, many responses did not clarify how they would end up with their final sample e.g. not stating that the first 20 volunteers were used. Some responses confused different sampling techniques, for example naming opportunity sample but then describing self-selected sampling. There were some very clear justifications in context, however,

	application of required features in context		(good) 12 marks if just one RF explicitly linked, 13 marks if two, 14 marks if three and 15 if all four are linked explicitly. If there is no explicit clear link between own practical work and *any* of the 4 required features caps the mark at 11 maximum. **Maximum 11 marks (reasonable) if *clearly* done as a field experiment.** **Overall Mark Decide on band and final mark** **Look at RF first** **L4 Good** – all 4 good (L4) in context **L3 Reasonable** – min 3 reasonable (L3) in context (could be 1 good and 2 reasonable OR 2 good and 1 reasonable) **L2 Limited** –
Reasonable 8-11 marks	At least **3 required features in context** Reasonably accurate and detailed knowledge and understanding of each feature	**Some** appropriate **justification** of decision related to required features (if no justification in context award 8 marks) There was a line of reasoning evident with some structure	
Limited 4-7 marks	At least **two** of the required features addressed in context **Limited application** of required features	**Attempt** to justify decision(s) but weak Evidence of some structure, but weak	
	OR three or all four required features referred to but in a limited way		
	If one required feature addressed in detail and justified in context and explicit links made to own practical work award 4 marks		
Basic 1-3 marks	At least **one** of the required features addressed **Weak application**	**None**, or if present very weak	

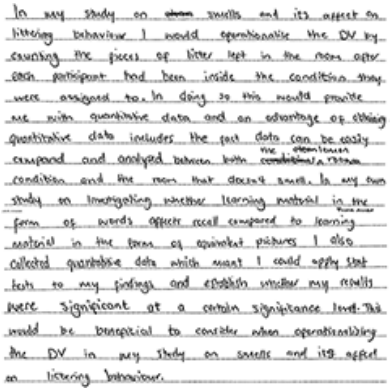
				of required features **OR** more than one of the required features referred to but in a very brief and/or basic way		min 2 limited (L2) in context or 3-4 limited (L2) with no context **L1 Basic** – 1 basic (L1) (no context needed). **THEN look at justifications** Make judgement of which mark to give the response within the band based on the justifications **L4 Good** – at least 2 reasonable (L3) AND at least 2 of the justifications are in context (does not have to be the reasonable ones) **L3 Reasonable** – at least 2 limited (L2) AND at least 1 of the justifications is in context (8 marks if none contextualised OR meets the minimum justification requirement). **L2 Limited** – at least 1 limited (L1) (none have to be in context) OR If one required feature		some did not use acceptable strengths of the sampling technique. For example, stating that opportunity sampling method would give a representative sample. There was great variation in the response to RF2. Better responses gave a clear outline of how the dependent variable of 'amount of litter' would be operationalised. Such as 'counting the pieces of litter left in the room on a tally chart' or 'weighing the amount of litter left in the room in grams'. Sometimes the operationalisation of dependent variable lacked when and where the collection of data would occur or where the litter came from which sometimes limited the response to the 'reasonable' rather than 'good' band for this RF. Common justifications for this RF included making comparisons, use of bar charts and doing statistical analysis. Some responses gave weaknesses of their dependent variable which did not gain marks. Some justifications were contextualised but this was inconsistent in the responses. Candidates used a wide variety of ethical considerations in their response to RF3. Popular responses including debriefing which told the aim of the study and revealed any deceptions used as to the nature of the study, confidentiality and allowing the participants to withdraw their data. Giving informed consent was another successful response. Clear justifications contextualised the response and often referred to respect for participants or the necessity to follow the BPS guidelines. Weaker responses to this RF did not address the assessment request to address one ethical issue and often attempted to address many ethical issues. This led to less detail being given to each ethical issue raised in the response as well as difficulty in justifying any one ethical issue in

					addressed in detail (good) and justified in context and explicit links made to own practical work award 4 marks **L1 Basic** – no justification or basic justification	

RF		Details of RF
1	Sampling technique	• Good – Identified the sampling method and clearly explained where and how this has been carried out in their study. Details of how the sampling method is enacted/procedural details e.g. how the P is contacted, is approached or gets in touch with experimenter or becomes part of the sampling pool (e.g using everyone in that location). ` • Reasonable – Identified the sampling method, possibly defined AND reasonable attempt to explain how this has been carried out in their study. • Limited – Sampling method identified and defined OR unclear attempt to explain how this has been carried out in their study. • Basic – Just identifying the sampling technique

detail. Some responses were in context, but this was less consistent compared to RF1 and RF2. The ethical consideration of confidentiality sometimes stated exactly how the researcher would keep the data from the participant confidential, such as using numbers in place of participants' names. However, many responses that had confidentiality as the ethical issue, often just stated they would keep it anonymous with no further explanation as to how they would do this. Some of the justifications were very brief, sometimes just stating it made the study 'more ethical'.

There was a wide variety of suggestions given in responses to RF4. The best responses often detailed control of situational variables with a recognition that keeping the lemon smell/no smell in the rooms consistent for each participant was important. Popular ideas including opening windows and not allowing participants to wear any form of perfume during the study. Other ideas that often achieved well standardised the environment by having the layout of the rooms, location of the bins and time spent in rooms standardised among participants. There were also some good responses about participant variables with a recognition that anyone who had problems with smell or was allergic to lemons should be excluded from the study. Weaker responses often just identified the extraneous variable to be controlled or gave a very brief outline of how to control. Some responses were contextualised while others were not. Justifications were often linked to increased validity and/or reliability and some were able to give a clear explanation of why the control would lead to this improvement in the study. Better responses had specific evaluation points included. Weaker candidates

					or confuses sampling methods.	justified by suggesting this would improve validity and/or reliability with no further elaboration. Links to own practical work were usually explicit with the responses outlining what they did in their study. Some candidates linked their own practical work to the choice of the RF while others just simply described what they did in their study. Many were able to further justify their choice of RF through their choices made in their own practical work. Exemplar 2  Exemplar 2 is an example of a candidate who has addressed RF2 reasonably and in context and provided some reasonable justification but in context and makes explicit reference to their practical work. This candidate achieved a mark in the reasonable band overall as the RFs were one good in context, two reasonable in context and one basic in context. The justifications for their response were mixed with some good and some reasonable and mainly in context.
			2	Operationalising DV	• Good – Clear details on how dependent variable will be operationalised. Outline how data is quantified and how/when the litter is counted (e.g. how weight is measured or unit of measure, comparison of amount before/after, number of pieces in the bin, tally of number of pieces of litter). • Reasonable – Reasonable details on how dependent variable will be operationalised that does lead to quantitative data e.g. count the number of pieces of litter/number of wrappers. May include a muddled/vague indication of where and how this has been carried out in their study. • Limited – Way DV is operationalised is quantitative and addressed in a limited/unclear way. E.g. does not indicate which litter is being counted, more than one measure indicated; Could indicate where the litter has come from and 'amount of litter left in the room'. • Basic – Vague indication of how DV would be measured (e.g. amount of litter left in the room).	
			3	One ethical consideration addressed	Integrity (deception) Respect (privacy/confidentiality/consent/right to withdraw), Responsibility (debrief/no psychological or physical harm	

					unlikely to be creditworthy unless clear how their study could be psychologically or physically harmful), Competence (refers to the competence of the researcher e.g. get an expert in to check for sensory impairments) • Good – Identifying the ethical consideration, explaining the ethical consideration and clarity on how it can be addressed. • Reasonable – Identifying the ethical consideration and reasonable explanation of how it can be addressed. Briefly addressed but lacks clarity. • Limited - Limited explanation with some understanding of the ethical consideration (e.g. outline of how to address ethical consideration possibly without identifying). • Basic – Just identifies the ethical consideration. If candidate clearly does more than one consideration, credit the first one. Allow ethical considerations which are clearly connected to each other/influence each other. There may be a mislabelling of the ethical principle. Therefore, if the RF meets the requirements of the description, it can be put at this level. This RF needs to focus on the way the ethical consideration is addressed. Any information given on the reason is justification.		

			4	Control of one Extraneous variable	• Good – Clear and somewhat detailed of how EV can be controlled. • Reasonable – Reasonable outline of how EV can be controlled. • Limited – Limited/brief outline of how EV can be controlled is unclear. • Basic – Identifies how EV can be controlled or is muddled. If more than one extraneous variable, credit the first one.	
			Total		**15**	
51	a		Ranking data refers to assigning numbers to denote position in an ordered sequence (lowest to highest or vice versa) Clear explanation - 2 Attempted explanation/e.g. how to do ranking - 1 The candidate has not provided any creditworthy information - 0		Max 2	For example - 1 mark – order the scores/numbers lowest to highest; highest to lowest; OR numerical order 2nd mark –explaining how the ranks are assigned e.g. 1 for the lowest, up to 20 for the highest OR explaining what ranking the data means **Examiner's Comments** The majority of candidates showed knowledge of ranking that involves putting the data in numerical order from lowest to highest. Better responses showed knowledge of how to rank (i.e. lowest number is rank of 1 etc.)
	b		Because there are three friendliness ratings of 18, covering ranks 16, 17 and 18 collectively. Therefore, the ranks must be shared (16 + 17 + 18 = 51/3= 17) Clear explanation - 2 Attempted explanation - 1 The candidate has not provided any creditworthy information - 0		Max 2	1 mark for identifying that identical numbers need to have the same rank OR identifying that there are three ratings of 18 2nd mark for how you calculate what that rank should be (mean or median or showing how this is done) **Examiner's Comments** The majority of candidates identified that there were three

<table>
<tr><td></td><td></td><td></td><td></td><td></td><td>ratings of 18 and therefore would need to have the same ranking. Many responses also explained that ranks 16, 17, 18 were assigned and the median/mean/middle value found. Some candidates did not gain the second mark due to not explaining how this was calculated (the mean/median of the rank positions).</td></tr>
<tr><td></td><td></td><td></td><td>Total</td><td>4</td><td></td></tr>
<tr><td>52</td><td></td><td></td><td>
<table>
<tr><td colspan="3">Wearing name badge condition = 14.8
Workings 18+14+10+17+16+18+5+11+20+19 = 148 (148/10 = 14.8)
No badge condition = 8.8
13+2+6+4+8+18+7+15+12+3 = 88 (88/10 = 8.8)</td></tr>
<tr><td colspan="3">Correct calculation of mean in both conditions with workings for both - 3</td></tr>
<tr><td>Correct calculation of mean in both conditions with workings for one - 2</td><td>OR correct calculation of mean in both conditions but no workings - 2</td><td>OR correct calculation of mean in one condition with workings - 2</td></tr>
<tr><td colspan="2">Correct calculation of mean for one condition with no workings - 1</td><td>Correct workings with incorrect mean - 1</td></tr>
<tr><td colspan="3">The candidate has not provided any creditworthy information - 0</td></tr>
</table>
</td><td>Max 3</td><td>Context not required
Credit workings for full marks as (both included)-
148/10 = 14.8
88/10 = 8.8
Credit all results for one condition on one line divided by 10 on the next e.g.
18+14+10+17+16+18+5+11+20+19 / 10
= 14.8
Credit rounding up
15 and 9

<u>Examiner’s Comments</u>

The majority of candidates were correct and most achieved full marks by including the correct workings and correct mean scores for both conditions. Some responses went beyond what was needed and also calculated the mean of both conditions together which was not a requirement of the question. A minority of responses incorrectly worked out the mean of the ranks rather than the ratings of friendliness which did not gain marks. There were a very small number of responses that made an error in their calculation and presented the incorrect mean score for one of the conditions.</td></tr>
<tr><td></td><td></td><td></td><td>Total</td><td>3</td><td></td></tr>
<tr><td>53</td><td></td><td></td><td>B</td><td></td><td>competence

<u>Examiner’s Comments</u>

The majority of candidates</td></tr>
</table>

					responded well to this question. Incorrect responses were varied.
			Total	**1**	
54			A		8:3 **Examiner's Comments** The majority of candidates responded well to this question. Incorrect responses were varied.
			Total	**1**	
55			D		percentage **Examiner's Comments** The majority of candidates responded well to this question. Incorrect responses were varied.
			Total	**1**	
56			A		Chi-square **Examiner's Comments** The majority of candidates responded well to this question. Incorrect responses were varied.
			Total	**1**	
57			D		Wilcoxon **Examiner's Comments** The majority of candidates responded well to this question. Some candidates chose option B incorrectly.
			Total	**1**	
58			B		data that is obtained directly from the sample by the researcher(s) **Examiner's Comments** The majority of candidates responded well to this question. Incorrect responses were varied.
			Total	**1**	

59			The probability of the null hypothesis being true (or results being due to chance) is less than 5% (so in this study, there is a significant difference in what male and female motorists find annoying in other motorists) Correct answer clearly explained - 2 Basic response such as 'There is a significant difference' - 1 The candidate has not provided any creditworthy information - 0	Max 2	Context is not required for full marks. For 1 mark accept 'The null hypothesis will be rejected' or 'The Alternate hypothesis will be accepted'. **Examiner's Comments** Many candidates didn't understand this question or didn't attempt it and gained zero marks. Some candidates gained 1 mark and recognised that data would be significant or that the null hypothesis would be rejected. The minority of candidates accessed full marks on this question.
			Total	**2**	
60			D	1	**Examiner's Comments** The majority of candidates responded correctly however a minority of students chose option C.
			Total	**1**	
61			C	1	**Examiner's Comments** The majority of candidates responded correctly with the minority of candidates incorrectly choosing option A.
			Total	**1**	
62			C	1	**Examiner's Comments** Some candidates were unable to give a correct response to this question. Probably because more focus was given to identifying ordinal and nominal data.
			Total	**1**	
63			B	1	**Examiner's Comments** The majority of candidates responded correctly.
			Total	**1**	

<table>
<tr><td>64</td><td></td><td></td><td>A</td><td>1</td><td><u>Examiner’s Comments</u>

The majority of candidates responded correctly.</td></tr>
<tr><td></td><td></td><td></td><td>Total</td><td>1</td><td></td></tr>
<tr><td>65</td><td></td><td></td><td>B</td><td>1</td><td><u>Examiner’s Comments</u>

The majority of candidates responded correctly.</td></tr>
<tr><td></td><td></td><td></td><td>Total</td><td>1</td><td></td></tr>
<tr><td>66</td><td></td><td></td><td>Weaknesses could include: problems understanding rating scale (if used) by participants; problems responding to forced choice, lack of insight and meaning relating to why participants behaved / responded as they did etc Up to 3 marks for each weakness

Clear outline of weakness in context - 3

Clear outline of weakness but not in context - 2 OR attempted outline of weakness in context - 2

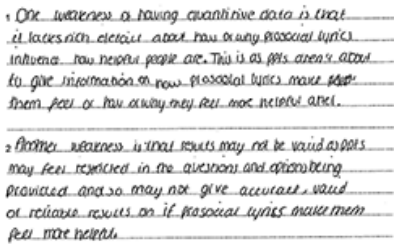
Brief and/or weak attempt to outline weakness (whether in context or not) - 1

The candidate has not provided any creditworthy information - 0</td><td>Max 3+3</td><td>-Context = music, prosocial, love, kindness, helpful etc

<u>Examiner’s Comments</u>

Most candidates were able to outline one weakness of quantitative data but were unable to outline a second weakness without becoming repetitive in their response.

Exemplar 2

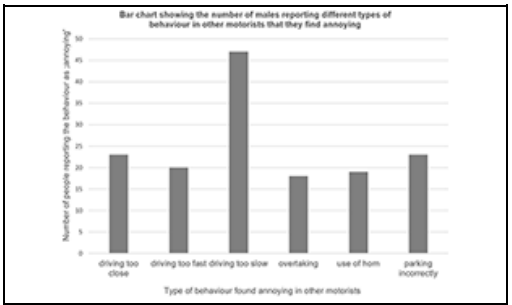
 
1. One weakness of having quantitive data is that it lacks rich detail about how or why prosocial lyrics influence how helpful people are. This is as this doesn't allow to give information on how prosocial lyrics make them feel or how likely they feel more helpful after.
2. Another weakness is that results may not be valid as ppts may feel rejected in the questions and options being provided and so may not give accurate, valid or reliable results on if prosocial lyrics make them feel more helpful.

This exemplar demonstrates how candidates can address the question without repetition of their first weakness. The responses identify separate weaknesses, explain why they are a weakness and are in context of the stem given.</td></tr>
<tr><td></td><td></td><td></td><td>Total</td><td>6</td><td></td></tr>
<tr><td>67</td><td></td><td></td><td>Closed question. Reason = the data is just frequency counts of the number of responses in each category of</td><td>Max 2</td><td>Context is not required for full marks

<u>Examiner’s Comments</u>

Most candidates got this answer</td></tr>
</table>

			things that motorists may find annoying. The data is participants responses to fixed choice response options with no opportunity to respond in any different way. Correct question type identified and clearly justified - 2 Correct question type identified - 1 **OR** correct question type identified and unclear attempt to justify why - 1 The candidate has not provided any creditworthy information - 0		correct and gained 2 marks. For those who only gained 1 mark, they were unable to provide an appropriate reason for it being a closed question.
			Total	**2**	
68			For example: Male and female motorists find different things annoying (males 'people driving too slow' was the most frequently reported category, and for females 'people driving too close'), which suggests the way men and women approach driving and their experiences of it are very different; Many more females (51) than males (20) reported being annoyed by other motorists driving too fast, which suggests females are more safety conscience than males and think more about the consequences of driving too fast; Etc Up to 2 marks for each conclusion ... Clear outline of conclusion in context - 2 Clear outline of conclusion but not in context - 1 **OR** attempted outline of conclusion in context - 1 The candidate has not provided any creditworthy information - 0	Max 2+2	-Context = driving, car(s), anger, frustration, annoying etc -Must be conclusions (interpretation of data), and not simply stating findings (if so, cap at 1 mark max, whether one or two findings stated) **<u>Examiner's Comments</u>** This question gained a variety of responses with few candidates accessing the full range of marks. Most candidates were restricted in their marks due to only identifying findings rather than a conclusion. **Assessment for learning** Candidates would benefit from being given opportunities to examine raw data tables and practice extrapolating conclusions that could be drawn from them by looking at the nature of the data and any patterns that can be inferred. These skills can also be illustrated through data collected through practical investigations conducted in preparation for this exam. The difference in findings and conclusions from the core studies component could be highlighted to assist in demonstrating to candidates how to differentiate between findings and conclusions.

			Total	**4**	
69			 1 mark is awarded for correctly presenting by value each bar representing all 6 categories of motorists' behaviour found annoying for males 1 mark is awarded for named categories on the x axis 1 mark is awarded for clear labelling of the y axis (must be contextualised to 'annoying behaviour') 1 mark is awarded for a clearly contextualised title All features included - 4 3 features included - 3 2 features included - 2 1 feature included - 1 The candidate has not provided any creditworthy information - 0	Max 4	-Context = driving, car(s), anger, frustration, annoying etc Full Contextualisation would include driving/motorists and annoying If presented as a histogram cap at 2m If both males and females are included on the graph cap at 2m If a clearly contextualised title has been included, this could make up for any shortfall of labelling on an x and y axis **Examiner's Comments** Many candidates could draw a bar chart but few accessed full marks due to not providing a title, labelled axes or incorrectly presenting data as a histogram rather than a bar chart.
			Total	**4**	
70			14% Workings ... 18 (males) + 24 (females) = 42 in total So ... 42/300 x 100 = 14 Correct answer with full workings shown - 3 Correct answer with some workings shown - 2 Correct answer shown with no workings OR some correct workings shown but incorrect final answer - 1 The candidate has not provided any creditworthy information - 0	Max 3	18 + 24 is not a necessary working to show for full marks. **Examiner's Comments** Most candidates achieved full marks on this question. A small number of candidates lost a mark for not showing their full workings.
			Total	**3**	
71	a		Reasons: -test of difference (the study investigated the difference	Max 3	-Context = driving, car(s), anger, annoying, frustration, males and females etc **Examiner's Comments**

			between what males and females found annoying in other motorists' behaviour) -nominal data collected (frequency count of different categories of annoying behaviour in motorists) -independent measures design (males v females) Two or three appropriate reasons provided, at least one in context - 3 Two appropriate reasons provided without context OR one appropriate reason in context - 2 One appropriate reason provided without context - 1 The candidate has not provided any creditworthy information - 0		Most students managed to identify two criteria for use of the Chi square test. Those who gained full marks were able to contextualise at least one of these criteria explicitly to the scenario.
	b		df = 5 Workings ... (R-1) x (C-1) (2-1)x(6-1) = 5 Correct answer **(no need for workings) - 1** The candidate has not provided any creditworthy information - 0	Max 1	**Examiner's Comments** Many candidates got this question incorrect or didn't write a response.
	c		A clear explanation The 'degrees of freedom' is mapped against levels of significance on a table of critical values to establish the critical value necessary to judge the significance of the study. Basic explanation: e.g. It is used to find the critical value. - 1 The candidate has not provided any creditworthy information - 0	Max 2	For two marks candidates must refer to level of significance or probability AND finding the critical value. Reference to number of participants is not creditworthy. **Examiner's Comments** Many candidates got this question incorrect or didn't write a response. Some candidates gained 1 mark for reference to finding a critical value. A minority of candidates gained full marks on this question.
			Total	**6**	
72			C	1	**Examiner's Comments** The majority of candidates responded correctly.
			Total	**1**	
73			C	1	**Examiner's Comments** The majority of candidates responded correctly.

			Total	**1**	
74			B	1	**Examiner's Comments** The minority of candidates found it difficult to identify this graph as a histogram. **Misconception** There was a common misconception here that this is a bar chart rather than a histogram.
			Total	**1**	
75			Accept −0.639 and −0.64 Workings … $r_s = 1 - \frac{6(\sum d^2)}{n(n^2-1)}$ 6 × 270.5 = 1623 1623 / 990 = 1.64 1 − 1.64 = −0.64 **3 mark answer**: Correct answer with all workings **2 mark answer**: Correct answer with some workings **1 mark answer**: Correct calculation withno workings **OR** Some correct workings,but incorrect answer **0 marks**: No credit worthy information	3	**Examiner's Comments** There was a spread of marks for this question. Candidates generally were able to calculate the value of r_s however were sometimes penalised for not including their workings.
			Total	**3**	
76			The probability of there being no relationship between self-ratings of extroversion and ratings by a friend is greater than 5%, meaning that this is not a statistically significant result. There is no relationship between self-ratings of extroversion and ratings made by a friend. **3 mark answer**:	3	Context = personality, extroversion etc **Examiner's Comments** There were quite a few non-responders to this question, in addition to many stating that this would mean there was a significant relationship. Some candidates

			Clear explanation in context **2 mark answer**: Clear explanation but not in context **1 mark answer**: Brief and/or weak explanation, whether in context or not **0 marks**: No credit worthy information		stated that there would not be a significant relationship but did not go beyond this. Very few included context and a minority of candidates were able to give a clear explanation.
			Total	**3**	
77			C	1	**Examiner's Comments** Most candidates answered this correctly.
			Total	**1**	
78			D	1	**Examiner's Comments** Many candidates answered this correctly with some giving an inaccurate response.
			Total	**1**	
79			A	1	**Examiner's Comments** Many candidates answered this correctly with some inaccurately answering C as their response.
			Total	**1**	
80			A	1	**Examiner's Comments** Some inaccurate responses were given to this question.  **Assessment for learning** Candidates should undertake calculations of inferential statistical tests to understand the outcomes.
			Total	**1**	
81			For example: inclusion of distractor questions; ensuring anonymity of respondents; disguising aim of research	6	Context = mental illness, psychiatric, attitudes etc

			etc **3 mark answer**: Clear outline in context **2 mark answer**: Clear outline but not in context **OR** Attempted outline in context **1 mark answer**: Brief and/or weak attempt to outline (whether in context or not) **0 marks**: No credit worthy information		3 marks for each way outlined to reduce demand characteristics. **Examiner's Comments** There was a mixed response to this question. Some candidates gained 3 marks for one suggestion but struggled to write an equally successful response for another suggestion. Some candidates were penalised due to the lack of detail or context in their response. Some candidates were unable to give a response or gave a response that wasn't relevant to the self-report study such as using independent measures design.
			Total	**6**	
82			Conclusions could include: there was a variety in how consistent some people rated their own personality compared to the rating from their best friend, suggesting that some people think about their personality in a similar way to how their friend perceives them, whereas other people think very differently. The differences may be due to how well or how long the person has known their best friend. Judging personality is a difficult thing to do, and the way we perceive our own personality may be different because it is based on different things to what anyone else used to judge personality, such as things that are private and only known to the individual etc **3 mark answer**: Clear detailed response, in context **2 mark answer**: Clear detailed response, but not in context OR Attempted response, in context **1 mark answer**: Attemped response, not in context **OR** simply stating a finding	6	• Context = personality, extroversion etc • Clear (explicit) interpretation of findings (not simply stating a finding) is required for top band Accept any other appropriate conclusions here. 3 marks for each conclusion **Examiner's Comments** This question gained a variety of responses with not many candidates accessing the full range of marks. Candidates were required to interpret a table of data and infer a finding. Candidates needed to go beyond merely restating an individual finding which many did not do. Credit was given for a finding but this restricted candidates to low band marks. **Assessment for learning** Candidates would benefit from

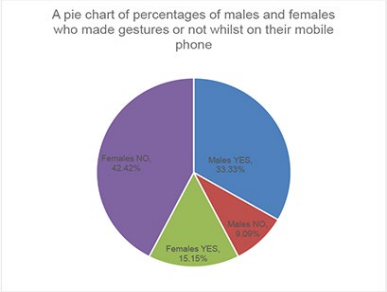
			0 marks: No credit worthy information		being given opportunities to examine raw data tables and practice extrapolating conclusions that could be drawn from them by looking at the nature of the data and any patterns that can be discerned. These skills can also be illustrated through practical work data gathered in preparation for this exam. The difference in findings and conclusions from the core studies component could be highlighted to assist in demonstrating to candidates how to differentiate between findings and conclusions.
			Total	**6**	
83			Strengths include easy to analyse, record and present findings on ratings of extroversion; easy to plot visually (scatter diagram) etc Weaknesses include lack of insight into why ratings on extroversion mean, or why given; possible differences in use / interpretation of extroversion scale between participants etc **3 mark answer**: Clear outline of strength/weakness, in context **2 mark answer**: Attempted outline of strength/weakness, in context **OR** Clear outline of strength/weakness, not in context **1 mark answer**: Brief and/or weak outline of strength/weakness, whether in context or not **0 marks**: No credit worthy information	6	Context - personality, extroversion etc 3 marks for strength 3 marks for weakness **<u>Examiner's Comments</u>** Most candidates were able to identify a strength and a weakness of quantitative data. For those not achieving full marks, often candidates were penalised for not contextualising their response or for not elaborating on the strength/weakness. For example, candidates need to go beyond 'it lacks detail' for a weakness.  **Assessment for learning** Centres should encourage candidates to make sure they always contextualise their responses when a question includes the phrase "in this study". Candidates could be encouraged to self-reflect on their own work and highlight their use of context. This will assist candidates in identifying where they are losing marks if they

					have not included context in their response.
			Total	**6**	
84			Reasons: • test of correlation/relationship (and the study investigated the relationship between self-rating and friends rating of extroversion) • ordinal data collected (ratings of extroversion on 0 to 100 scale) **3 mark answer**: Two appropriate reasons provided, both in context **2 mark answer**: Two appropriate reasons provided, but only one, or neither in context OR One appropriate reason provided in context **1 mark answer**: One appropriate reason provided but not in context **0 marks**: No credit worthy information	3	Context - personality, extroversion etc **<u>Examiner's Comments</u>** Candidates answered this question well in terms of knowledge. Very few candidates contextualised their response so were often penalised here.
			Total	**3**	
85			C	1	**<u>Examiner's Comments</u>** This question was answered correctly by most candidates.
			Total	**1**	
86			D test-retest		**<u>Examiner's Comments</u>** The majority of candidates answered correctly. Some candidates chose option A incorrectly.
			Total	**1**	

87			B nominal		**Examiner's Comments** This was answered correctly by the majority of candidates. Incorrect choices by candidates were varied.
			Total	**1**	
88			C 3:2	Enter text here.	**Examiner's Comments** Most candidates were successful in this question. A small number of candidates chose option D incorrectly.
			Total	**1**	
89			B 0.04		**Examiner's Comments** Answered correctly by most candidates. Some candidates chose option A incorrectly.
			Total	**1**	
90			B Binomial Sign		**Examiner's Comments** Answered correctly by less than half of all candidates. Many candidates chose option D incorrectly.
			Total	**1**	
91			C sample size		**Examiner's Comments** Answered correctly by over half of the candidates. Some candidates chose option A incorrectly.
			Total	**1**	
92			D positively skewed		**Examiner's Comments** Answered correctly by less than half of the candidates. Some candidates chose option B or C incorrectly.
			Total	**1**	
93			Possible examples Disguising aim of study Asking additional, unrelated questions etc. One way to reduce demand characteristics clearly presented in context	Max 2 2 1	Context = friendliness, hand-shaking etc. Experimenter isn't in room when study is taking place Independent measures (attempt unless explained) Experimenter keeps a neutral

			<table><tr><td>One way to reduce demand characteristics clearly presented but not in context</td><td>**OR** attempted way to reduce demand characteristics (whether in context or not)</td></tr></table> The candidate has not provided any creditworthy information	0	expression (attempt unless explained) Ensuring anonymity of responses. **Examiner's Comments** There were many good responses which used deception or distractor questions to disguise the aim or use independent measures design. Better responses used context examples as part of their explanation. Lower scoring responses gave a basic outline of how to reduce demand characteristics such as disguising the aim without outlining either how this could be done (e.g. being told a different aim) or why this would lead to a reduction in demand characteristics. A significant number of responses were confused between demand characteristics and social desirability or suggested deceiving the participants in the debrief as the candidate was clearly confused about when a debrief takes place.
			Total	**2**	
94			Conclusions could include: males gesture more, perhaps because they are less self-conscious than females, or that they are more visual in how they think and this is reflected in conversations they have whilst on the phone with other people making them more animated etc. Accept any other appropriate conclusions here. <table><tr><td colspan="2">**3 marks for each conclusion**</td></tr><tr><td colspan="2">Clear, detailed conclusion in context</td></tr><tr><td>Clear, detailed conclusion but not in context</td><td>**OR** attempt in context</td></tr><tr><td>Brief and/or weak attempt (whether in context or not)</td><td>**OR** simply stating a finding</td></tr></table> The candidate has not provided any creditworthy information	3 3 2 1 0	Context = mobile, phone, gestures etc. Must be an interpretation of findings (not simply presentation of finding). 1 mark for finding, 2 marks for explanation. Example of a finding - 'Males gesture more than females' (1) Zero marks if just data is given. **Examiner's Comments** Most responses were able to gain 1 mark for this question by concluding that men use more gestures than women while making or receiving a mobile phone call. However, many candidates did not understand the term 'conclusion' and often repeated the data from

					the table. Better responses gave an explanation as to why men might use more gestures such as men being more passionate or enthusiastic; females being more verbally articulate so did not need to gesture to emphasise their point.
			Total	**3**	
95			A pie chart showing the proportion of males compared to females who made gestures whilst on their mobile phone  Each feature can be awarded one mark as follows ... 1 mark is awarded for correctly calculating what % of the circle should represent the males and females who made gestures. 1 mark is awarded for drawing the sectors in correct proportion to data 1 mark is awarded for clear labelling of each sector of the pie chart 1 mark is awarded for a clear and appropriate title 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	1 mark for % 1 mark for proportions 1 mark – male/female labelled (or key) 1 mark title Males 11/16 × 100 = 68.75% accept 69 Females 5 / 16 × 100 = 31.25% accept 31 Title should refer to males and females and gestures. If no reference is made to percentage/proportion in title there must be % **within** the pie chart to be creditworthy. Reference to **number** of males/females is not creditworthy. If all data included – can be awarded full marks (see pie chart below) the candidate may alter the percentage slightly for one of the categories to show the pie chart adds up to 100% (43% below illustrates this) which is still creditworthy. Credit the whole numbers (15%, 42/43%, 9% and 33%) regardless of what is written after the decimal point. Yes Female – 15% (15 2) No Female – 42% (42.4)/43% No Male – 9% (9.1) Yes Male – 33% (33.3) **OR** drawing 2 pie charts – one for males (21.4%, 78.6% or 21%, 79%) and one for females (26.3%, 73.7%, 26%, 74%)also can be awarded 4 marks Drawing a pie chart where the total for males (79%) and total for

					females (26%) is not creditworthy. Be generous with drawing of the circle and indication of the proportions e.g., for the 31% need to be over a quarter of the circle. **Examiner's Comments** Many candidates achieved 3 or 4 marks for this question by working out how many males and females gestured overall and then the percentage of males and females within this group. There were many responses that gave a clear title and labelled all sections of their pie chart as well as putting in the percentages for each section. Some responses provided a key rather than a label within the pie chart which was creditworthy. Less successful responses often had no title or an inappropriate title such as 'number of males and females..' where the pie chart showed the percentage. Those responses that provided a pie chart for all the data from the table were also able to achieve full credit for this question.
			Total	**4**	
96			 Any one reason in context from: nominal/categorical data obtained; looking for a difference; independent/unrelated groups One appropriate reason in appropriate context One appropriate reason but not in context The candidate has not provided any creditworthy information	Max 2 2 1 0	Nominal data/categories – whether the participant makes a gesture or not OR yes/no Independent measures – male/female Looking for a difference – between males and females OR difference between the number of gestures made between males and females. **Reason and context must be linked**. **Examiner's Comments** Many candidates achieved full marks for this question by either identifying that the data was nominal with the example of the categories from the study (use of

					gestures or not) or identified that the study had an independent measures design with the two conditions of male and female. A small number of responses identified that the study was investigating a difference but often this lacked context. Less successful responses often gave two reasons such as nominal and independent measures without contextualising either of these.
			Total	**2**	
97	a		9.79 Row total (17) × column total (19) / overall total (33) = 9.78787878 9.79 with workings shown 9.78787878 with workings (not to 2 dps) OR 9.79 without workings 9.78787878 without workings (not to 2 dps) OR correct workings only presented The candidate has not provided any creditworthy information	Max 3 3 2 1 0	$\frac{17 \times 19}{33} = 9.79$ Award 3 marks 14−4.21 = 9.79 Award 3 marks 33−23.21 = 9.79 Award 3 marks Many candidates are putting workings out for 30(b) in the answer to this question. If possible, ignore these workings out for this question. 9.79 in correct column in table (regardless of what else is written) = 1 **Examiner's Comments** Candidates were well prepared for this question. A common good response was achieved via 17*19/33 method. Common error was not rounding up to 2 decimal places or just knocking off the remaining numbers and using 9.78 rather than rounding up to 9.79.
	b		$X^2 = 8.8$ Workings ... Cell (d) O − E − 14 − 9.79 = 4.21 $(O - E)^2 = 4.21 \times 4.21 = 17.72$ $(O - E)^2 / E = 17.72 / 9.79 = 1.81$ So, the sum of $(O-E)^2/ E = 2.61 + 1.92 + 2.46 + 1.81 = 8.8$ Correct calculation of overall chi square value with workings shown	Max 3 3	Can credit cell calculations in the table 17.72 / 9.79 = 1.81 (1) – on its own 2.61 + 1.92 + 2.46 + **1.81** = 8.8 award 3 marks If incorrect frequency worked out in 30(a) will mean that the calculation will be incorrect = 0 **Examiner's Comments**

			<table><tr><td colspan="3">Correct calculation with some workings shown</td></tr><tr><td>Correct calculation.</td><td>All workings shown correctly but no calculation given (or incorrect calculation)</td><td>Cell D calculated correctly.</td></tr></table> The candidate has not provided any creditworthy information	2 1 0	Candidates were well prepared for this question. The most common good response: 2.61 + 1.92 + 2.46 + 1.81 = 8.8. A common error was to work out **1.81 to 1.82** instead.
	c		df = 1 Workings ... (2−1) × (2−1) = 1 <table><tr><td colspan="2">Correct answer with workings</td></tr><tr><td>Correct answer but not workings (or workings incomplete or unclear)</td><td>Workings but no answer</td></tr><tr><td colspan="2">The candidate has not provided any creditworthy information</td></tr></table>	Max 2 2 1 0	No marks for just formula. (R−1) × (C−1) **<u>Examiner's Comments</u>** Well answered, most candidates understood how to calculate this and got the correct answer. Some counted the headings of the rows and columns in the table OR counted the rows and columns in the table used to work out cell d, making the response incorrect.
	d		6.635 Correct answer provided The candidate has not provided any creditworthy information	Max 1 1 0	Award 1 mark if candidate has circled (or otherwise correctly indicated) correct answer in table. **<u>Examiner's Comments</u>** Candidates achieved well on this question with many identifying the correct critical value of 6.635. Small number of candidates had the common error of using the 0.05 critical value so clearly misinterpreting its meaning. In addition, some chose the incorrect critical value due to having calculated the incorrect degree of freedom in Question 30 (d).
	e		Any **three** from the following statements. - There is a statistically significant difference between the number of males and females who made gestures whilst on the phone - The probability of there being no difference between the number of	Max 3	Context = mobile, phone, gestures etc **<u>Examiner's Comments</u>** There were many good responses to this question where candidates correctly interpreted that the

<table>
<tr><td></td><td></td><td></td><td>males and females who made gestures whilst on the phone is less than 1%
- Males are more likely to make gestures whilst on the phone compared to females
- Because calculated is higher than the critical (a significant difference has been found).
- Null hypothesis can be rejected/alternative hypothesis can be accepted.
- any probability level other than 0.001 (e.g., p ≤ 0.01)
3 correct statement – with at least 1 in context
2 correct statements
1 correct statements
The candidate has not provided any creditworthy information</td><td>3
2
1
0</td><td>calculated value was higher than the critical value and so the result of the test was significant. Better responses contextualised this in light of the study's overall conclusion. Many then correctly identified that the null hypothesis was rejected. Less successful responses often lacked context, accepting the null hypothesis, and not knowing how to refer to the meaning of the significance level accurately i.e. probability of confidence in the IV impacting the DV/ probability results were a fluke/chance.</td></tr>
<tr><td></td><td></td><td></td><td>Total</td><td>12</td><td></td></tr>
<tr><td>98</td><td>a</td><td></td><td>Answers could include: easier data analysis and comparison across groups (males and females); easier to record data; easier to interpret, no researcher bias to interpret, etc.

<table><tr><td colspan="2">Clear outline of strength in context</td></tr><tr><td>Clear outline of strength but not in context</td><td>Attempted outline of strength in context</td></tr><tr><td colspan="2">Brief and/or weak attempt at an outline of strength (whether in context or not)</td></tr></table>
The candidate has not provided any creditworthy information</td><td>Max 3

3
2

1

0</td><td>Context = mobile, phone, gestures, male/female etc.

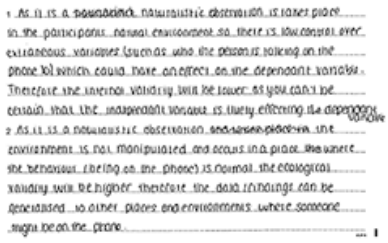
No credit for reference to being able to calculate the mean as this is not relevant for this data.

<u>Examiner's Comments</u>

There were many accurate strengths identified by candidates in this question such as easy to analyse or to compare. Some responses were clearly outlined as to why this was a strength in the context of the study. Lower achieving responses were where candidates did not fully elaborate on a point and just mentioned that this made the study more reliable or more objective, for example, without an explanation as to why. In addition, some candidates did not recognise that the data to which they were referring was nominal and mentioned factors such as calculating mean.</td></tr>
<tr><td></td><td>b</td><td></td><td>Answers could include: doesn't inform us about reasons why there is a difference in amount of gesturing made whilst on phone between males and</td><td>Max 3</td><td>Context = mobile, phone, gestures, male/female etc.

<u>Examiner's Comments</u></td></tr>
</table>

			females, mean cannot be calculated, etc. Clear outline of weakness in context Clear outline of weakness but not in context / Attempted outline of weakness in context Brief and/or weak attempt at an outline of weakness (whether in context or not) The candidate has not provided any creditworthy information	3 2 1 0	There were many good responses to this question which often referred to lack of insight or detail. The best responses included some relevant detail such as quantitative data being limited because we are unaware of why gestures were being made such as how annoyed the participant on the phone was or the content of their call. Responses that did not score as well did not outline why a lack of detail is a problem or did not include context.
			Total	**6**	
99			Evaluation points could include: higher validity as participants unlikely to know they are being observed; higher ecological validity as in natural setting of high street; low validity due to possibly missing/misinterpreting some use of gesturing if high street busy, lower validity due to more extraneous variables in the natural environment, etc. Accept any other appropriate conclusions here. **3 marks for each evaluation point** Clear, detailed evaluation in context Clear, detailed evaluation but not in context / **OR** attempted evaluation in context Brief and/or weak attempt at evaluation (whether in context or not) The candidate has not provided any creditworthy information	6 3 2 1 0	Context = mobile, phone, gestures, male/female, high street, bench etc. Full credit can be given for two positive effects on validity, two negative effects or one of each. Lacks external validity - from a busy high street to different locations is creditworthy Do not credit population validity. Do not credit references to reliability. **<u>Examiner's Comments</u>** The marks achieved by candidates for this question covered the full range of the mark scheme. Common points included ecological validity, lack of demand characteristics which improved validity and lack of control of extraneous variables which reduced validity. Examples of good responses fully explained how the natural environment enabled the observers to record true to life gesture behaviour or how behaviour was unaffected by the demand characteristics that occur within the laboratory. Candidates provided good examples of possible extraneous variables that occur within a busy street that could affect the gestures that a person

					makes while on the phone (e.g bumping into people, little space, used phone call content well). Less successful responses often identified the issue but did not explain how validity would actually be affected, lacked detail or did not include context; all of which restricted the marks available to them. A significant number of responses identified population validity as an issue in this study which is not specific to naturalistic observations and therefore was not creditworthy. Exemplar 3  Exemplar 3 is a full mark response where the two points raised are clear, detailed and in context.
			Total	**6**	
100			A Mann-Whitney U		**Examiner's Comments** Most candidates responded correctly to this question. Some candidates chose option C incorrectly.
			Total	**1**	
101			B 6		**Examiner's Comments** Most candidates answered this question correctly. Incorrect choices by candidates were varied.
			Total	**1**	
102	a		Answer T = 5 Workings p: Number of times snacks taken from fridge during one week (a) Without eyes on; (b) With eyes on; Diff (a – b); Ranks of diffs; Sum of ranks of neg' diffs; Sum of ranks of pos' diffs	5	Allow Diff to be b-a

				fridge door	fridge door			
			a	14	10	4	7	7
			b	21	18	3	5	5
			c	32	24	8	10	10
			d	28	24	4	7	7
			e	20	16	4	7	7
			f	8	10	-2	3.5	3.5
			g	26	24	2	3.5	3.5
			h	12	12	0	-	
			i	22	23	-1	1.5	1.5
			j	18	12	6	9	9
			k	20	19	1	1.5	1.5
			l	21	21	0	–	
							Σr= 5	Σr= 50
			1 mark for finding difference a-b 1 mark for ranking differences 1 mark for sum of negative ranks 1 mark for sum of positive ranks 1 mark for correct identification of T					
	b		T = 5 , n = 10, p<0.05 **OR** In words ... e.g. the null hypothesis (i.e. that there is no difference in amount of food consumed with eyes on fridge compared to without) is rejected as the calculated value is less than the critical value. There is a significant finding / the hypothesis is accepted (eyes on fridge *does* make a difference to amount of food consumed). **2 mark answer:** Correctly written significance statement (calculated value, sample size 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 mark answer:** No creditworthy response		Max 2	Context = food, snacks, eat(ing), fridge etc OR the correct numerical data (e.g. T=5). 1 mark for correctly saying null rejected, hypothesis accepted.

			Total	**7**	
103	a		Table critical value = 8 **2 mark answer:** Critical value correctly stated **1 mark answer:** Correct sample size used (10) but wrong critical value quoted **0 mark answer:** No creditworthy response	2	
	b		**Answers could include:** • Population / scores normally distributed (68% of our data should be less than ±1 standard deviation around the mean) • At least interval / ratio (continuous) data • No extreme scores **1 mark per criteria:** Appropriate criteria identified **0 marks:** No creditworthy response	2	
			Total	**4**	
104	a		Answer could include: ability to analyse data about snacking more easily; can compare across conditions and individual participants easily etc **3 mark answer:** Clear outline of advantage in context **2 mark answer:** Clear outline of advantage, but not in context OR Attempted outline of advantage in context **1 mark answer:** Brief and / or weak to outline advantage (whether in context or not) **0 mark answer:** No credit worthy response	3	Context = food, snacks, eat(ing), fridge etc

	b		Answer could include: no reasons for why snacks consumed obtained **3 mark answer:** Clear outline of disadvantage in context **2 mark answer:** Clear outline of disadvantage, but not in context OR Attempted outline of disadvantage in context **1 mark answer:** Brief and / or weak to outline disadvantage (whether in context or not) **0 mark answer:** No credit worthy response	3	Context = food, snacks, eat(ing), fridge etc
			Total	**6**	
105	a		Total 14,21,32,28,20,8,26,12,22,18,20,21 =242 /12 Mean = 20.16666 So to 3 SFs = 20.2 **1 mark** for correct working **1 mark** for correct mean **1 mark** for correctly stating to 3 significant figures **0 marks:** No creditworthy response	3	
	b		Order: 10, 10, 12, 12, 16, <u>18, 19,</u> 21, 23, 24, 24, 24 Median = 18.5 Range = 24-10 = 14 OR 24-10 = 14+1 = 15 **1 mark** for ordering data correctly **1 mark** for identifying median correctly **1 mark** for calculating range correctly **0 marks:** No creditworthy response	3	
			Total	**6**	
106			Answer = 25% less food consumed when eyes on fridge compared to when no eyes on fridge	3	

			Workings 8 items less when eyes on fridge (32-24) 8 as a percentage of 32 = 8/32 x 100 = 25% **3 mark answer:** Percentage correctly stated with full workings shown **2 mark answer:** Percentage correctly stated with some workings shown OR Percentage only shown (no workings) **1 mark answer:** Partial workings shown e.g. 8/32 or 32-24 **0 mark answer:** No credit worthy response		
			Total	**3**	
107	a		There was more variation in the consumption of snacks when eyes were not on the fridge compared to when they were. This informs us that having eyes on the fridge has a fairly consistent effect on the amount of food consumed across the different participants. Without eyes on the fridge there is more variation, which informs us that some people may be better than others at restricting taking food from the fridge. **3 mark answer:** Clear explanation in context **2 mark answer:** Clear explanation, but not in context OR Attempted explanation in context **1 mark answer:** Brief and / or weak explanation (whether in context or not) **0 mark answer:** No credit worthy response	3	Context = food, snacks, eat(ing), fridge etc
	b		The standard deviation is the square root of the variance. The advantage of this is that the answer is a figure that is much more typical of the actual	3	Context = food, snacks, eat(ing), fridge etc Responses that refer to the

			difference (actual data collected) in the amount of snacking taking place with or without eyes on the fridge. The variance is a much bigger (untypical) figure as it is the sum of the differences in each individual's snacking compared to the mean which is then squared (squaring makes the answer a much bigger figure). **3 mark answer:** Clear outline in context **2 mark answer:** Clear outline, but not in context OR Attempted outline in context **1 mark answer:** Brief and / or weak outline (whether in context or not) **0 mark answer:** No credit worthy response		standard deviation as more valid / accurate than the variance are not creditworthy.
			Total	**6**	
108			B	1	
			Total	**1**	
109			C	1	
			Total	**1**	
110			A	1	
			Total	**1**	
111			B	1	
			Total	**1**	
112			A	1	
			Total	**1**	
113			B	1	
			Total	**1**	
114			D	1	
			Total	**1**	
115	a		A	1	
	b		B	1	

			Total	**2**	
116			C	1	
			Total	**1**	
117			**Outline one strength of the type of data collected in this study.** Possible strengths: • Easy to make comparisons / identify patterns • Objective / less open to interpretation • Easy to replicate <u>Example of 1 mark answer</u> Easy to compare data from each condition. <u>Example of 2 mark answer</u> It was easy to make comparisons (1) to see if one group's hippocampi were structured differently from the other's (1).	2	**2 marks** for a response which identifies a relevant strength which is then developed in the context of the study. **1 mark** for identifying a relevant strength either explicitly, or implicitly through reference to the study. **0 marks** – no creditworthy response.
			Total	**2**	
118			**Briefly discuss the extent to which Kohlberg's** (1968) **study into moral development may be considered ethnocentric.** Possible content: • More focus on American – through sampling and longitudinal research. • However, five other cultures studied to support universality of theory. • Still Western bias in overall sample. • Stories may be more relevant to some cultures than others. • Story was changed for Taiwanese participants showing some awareness of cultural differences. • Implication that some cultures develop slower than others, implying a deficiency in cultural experience or possibly intellect.	4	**3-4 marks** for a clear, detailed and well informed discussion of the extent of ethnocentrism in the context of the study. **1-2 marks** for a brief or vague discussion of the extent of ethnocentrism in the study. There may be some muddling of ideas or applications as part of the discussion. **0 marks** – no creditworthy response. NB The question requires a *brief* discussion and marks should be awarded according to this. A full mark response does not necessarily have to consider both sides of the argument – a good answer could argue that the study is or is not ethnocentric alone. The discussion can focus on one

					aspect of the study (e.g. sample) and still earn full marks.
			Total	**4**	
119		i	Explain **one** way in which the procedure of Grant *et al"*s (1998) study into memory increased the reliability of the research**.** Possible answers: • same background noise within condition • headphones worn for both conditions • all participants read the same article • tests for recall were the same • order of testing was the same • break between study and test phase was about 2 minutes each time <u>Example of 1 mark answer</u> The break between study and testing was about the same time each trial. <u>Example of 2 mark answer</u> All participants read with headphones on (1) to ensure consistency in experience (1). <u>Examples of 3 mark answer</u> Standardisation was used (1) where experimenters used the same article on psychoimmunology (1) to ensure that what participants had to learn was not an extraneous variable affecting recall (1). The experimenters used an article on psychoimmunology (1) which was published and accessible to anyone (1) so that the study could be easily replicated to establish reliability (1).	3	**3 marks** for a clear response which outlines what is meant by reliability (e.g. consistency or replicability), identifies how this is achieved in study, and then applies specifically to this study through example. **2 marks** for a vague response with all three of the above features or for a clear response with two of the features. **1 mark** for identifying a relevant way the procedure addressed the issue of reliability.
		ii	**Outline one conclusion that was drawn from this study.** Possible answers: • There are context-dependency effects for newly learned meaningful material regardless	2	**2 marks** for a clearly identified and relevant conclusion arising from the study. **1 mark** for a vague, brief or muddled conclusion arising from the study.

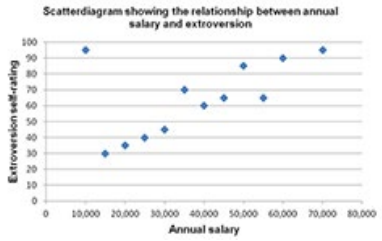
			of whether a short-answer test or a multiple-choice test is used to assess learning. • Studying and testing in the same environment leads to enhanced performance in recall through access to relevant cues. • Students are likely to perform better in exams if they study for them with a minimum of background noise because, although there was no overall effect of noise on performance, the fact that there was evidence for context-dependency suggests they are better off studying without background noise as it will not be present during actual testing.		**0 marks** – no creditworthy response.
			Total	**5**	
120		i	**Explain one weakness of the type of data collected in Loftus & Palmer's (1974) study into eye witness testimony.** Possible answer: Because the data was quantitative (1) there is a risk that it suggests similarity in response when there may not be (1) so for example a mean estimate of speed does not tell us whether certain individuals' estimates were similar across conditions (1). NB Other appropriate responses should be credited.	3	**3 marks** for a clear answer which; • identifies the data is quantitative or similar • identifies a weakness of quantitative data • identifies this weakness in the context of the study **2 marks** for an answer which addresses at least two of the above points. **1 mark** for a partial or vague answer which addresses at least one of the above points. **0 marks** – no creditworthy response. No credit for use of qualitative data.
		ii	**Outline one conclusion that was drawn from this study.** Possible answers: • The verb used in a question influences a participant's response i.e. the way a question is phrased influences the answer given.	2	**2 marks** for a clearly identified and relevant conclusion arising from the study. **1 mark** for a vague, brief or muddled conclusion arising from the study. **0 marks** – no creditworthy response.

			• People are not very good at judging vehicular speed regardless of the situation they find themselves in. • Misleading post event information can have a distorting effect on an individual's memory. Information gleaned during the perception of the original event and the post-event information integrate over time to affect memory. <u>Example of a 1 mark answer:</u> 'leading questions affected memory', 'eye witness testimonies are unreliable'.		No credit for just findings.
			Total	**5**	
121	a		$r_s = 1 - \frac{6(\Sigma d^2)}{n(n^2-1)}$ R_s = +0.517 **1 mark** for calculation of differences between ranks in each condition **1 mark** for calculation of sum of differences squared (Σd^2) = 138 **1 mark** for correct application / use of formula provided/substituting at least one number correctly **1 mark** for all workings at each stage shown **1 mark** for correct final overall answer (r_s = +0.517)	5	Won't necessarily show all these steps
	b		Table critical value = 0.587 **2 mark answer:** Critical value correctly stated **1 mark answer:** Correct sample size (12) used **or** 0.05 column, but wrong critical value quoted **0 marks:** No credit worthy information	2	
	c		As the Rs value (+0.517) is smaller than the critical value (0.587) with 12 participants and p<0.05, the null hypothesis is supported that there is no relationship between extroversion self-rating and salary earned).	2	Allow this written in numerical form As 0.517<0.587 for n=12, p>0.05

			2 mark answer: Correctly written significance statement with reference to calculated value, critical value, sample size and probability level **1 mark answer:** Weak and/or brief written response e.g. the results are not significant or there is no relationship between extroversion and salary **0 marks:** No credit worthy information		
			Total	**9**	
122			Likely answers: don't know reasons why personality may / may not be related to salary earned; doesn't differentiate between different types of extroversion; **3 mark answer:** Clear outline of weakness in context **2 mark answer:** Clear outline of weakness but not in context OR attempted outline in context **1 mark answer:** Brief and/or weak attempt (whether in context or not) **0 marks:** No credit worthy information	3	Context = personality, introvert, extrovert, money, salary etc Do not credit weaknesses of correlations
			Total	**3**	
123			Likely answers: demand characteristics affecting accuracy of extroversion self-rating; honesty regarding salary details; problems interpreting extroversion rating scale etc 3 marks for each thing **3 mark answer:** Clear outline in context **2 mark answer:** Clear outline but not in context OR attempted outline in context	6	Context = personality, introvert, extrovert, money, salary etc Can be strengths or weakness. Could outline two things that have a positive effect on the validity. Two things that have a negative affect on validity. One thing that is positive and one that is negative Could also outline a thing that might have affected the validity and the response explains that it is unclear

			1 mark answer: Brief and/or weak attempt (whether in context or not) **0 marks:** No credit worthy information		whether this affect would have been positive or negative. All types of validity creditworthy including, for example, population validity and temporal validity
			Total	**6**	
124	a		**Example response:** Why did you eat the food during the study? Describe to me why you think the colour of the plate might have influenced how much food you ate. **2 mark answer:** Clear outline in context **1 mark answer:** Clear outline but not in context OR Attempt in context **0 marks:** No credit worthy information	2	-Context = food, meal(s), eating, plate(s), crockery etc No marks awarded to closed questions. Can be awarded mark for both a question or a statement Do not necessarily have to ask a specific question for full marks
	b		Likely answers: provides rich and detailed responses, allows participants to explain their behaviour in the experiment, doesn't allows differences between participants to be ranked, is subjective, does not enables use of descriptive statistics (e.g. mean), some motivated participants may write more detailed responses compared to other participants, less reliable etc **3 mark answer:** Clear evaluation in context **2 mark answer:** Clear evaluation but not in context OR Attempt in context **1 mark answer:** Brief and/or weak attempt (whether in context or not) **0 marks:** No credit worthy information	3	-Context = food, meal(s), eating, plate(s), crockery etc Accept positive and/or negative evaluation points as creditworthy Can get full credit for one point although this will need to be a detailed point in context.
			Total	**5**	

125			External validity can refer to aspects of ecological validity and or population validity. **5-6 mark answer:** Clear evaluation with two or more points with some context (for 6 marks two of the points must be in context) **3-4 mark answer:** Clear evaluation with two or more points but not in context OR One clear evaluation point in context OR Two points, with one in context **1-2 mark answer:** Attempted evaluation (whether in context or not)	6	Context = food, meal(s), eating, plate(s), crockery etc Accept positive and/or negative evaluation points as creditworthy Do not accept evaluation points related to the methodology used in general Ignore reference to usefulness
			Total	**6**	
126	a		Ranking data refers to assigning numbers to denote position in an ordered sequence (lowest to highest or vice versa) **2 mark answer:** An explanation in context **1 mark answer:** An attempted explanation whether in context or not **0 marks:** No credit worthy information	2	Context = personality, introvert, extrovert, money, salary etc
	b		They both have the same extroversion rating (65), which in an ordered sequence of all the extroversion ratings occupy rank positions 6 and 7, so ranks 6 and 7 have been shared (added together and divided by two), resulting in a rank of 6.5 for both. **2 mark answer:** Clear explanation **1 mark answer:** Attempted explanation **0 marks:** No credit worthy information	2	Could refer to having a tied rank.

			Total	**4**	
127	a		 **1 mark for accurate plotting of data** **1 mark for title with reference to both co-variables and relationship/correlation** **1 mark for clear labelling of both axes** **1 mark for scales on both axes**	4	- If points joined up no mark for 'accurate plotting of data'
	b		Any 2 of the following: - There is a positive correlation between annual salary and extroversion/extroversion self-rating. (2 marks) There is a correlation between annual salary and extroversion. (1 mark) - Extroversion does not necessarily cause a high salary as correlations do not show cause and effect (2 marks) - There is an anomaly with one participant has a high extroversion score but a low annual salary.(2 marks) **2 mark answer:** Clear conclusion in context **1 mark answer:** Clear conclusion but not in context OR Attempted conclusion in context **0 marks:** No credit worthy information	4	-Context = personality, introvert, extrovert, money, salary etc
			Total	**8**	
128			775/12= Mean 64.583333 Two SFs = 65 **3 mark answer:** Mean correctly stated to two significant figures with all workings shown	3	

			2 mark answer: Mean correctly stated to two significant figures but with no workings shown **OR** Workings shown but mean not written to two significant figures **1 mark answer:** Mean only stated but not to two significant figures **0 marks:** No credit worthy information		
			Total	**3**	
129			Mean is a more sensitive measure of central tendency to use as there are no outliers in the results for extroversion self-ratings. The mean takes into account all of the data and is more mathematically accurate than the median which does not take into account all of the scores. **2 mark answer:** Clear explanation in context **1 mark answer:** Attempted explanation whether in context or not **0 marks:** No credit worthy information	2	Context = personality, introvert, extrovert, money, salary etc
			Total	**2**	
130			A dispersion around the mean	1	
			Total	**1**	
131			C nominal	1	
			Total	**1**	
132			D 0.05	1	
			Total	**1**	
133			B 4	1	

			Total	**1**	
134			D positively skewed	1	
			Total	**1**	
135			C data is normally distributed in the population	1	
			Total	**1**	
136			C 0.0061	1	
			Total	**1**	
137			D raw	1	
			Total	**1**	
138			B median	1	
			Total	**1**	
139			B median	1	
			Total	**1**	
140			A Chi-square	1	
			Total	**1**	
141			C Mann-Whitney U Test	1	
			Total	**1**	
142			C $p<0.05$	1	
			Total	**1**	
143			B level of data	1	
			Total	**1**	
144			**Explain one way in which Freud's study of Little Hans could be criticised for its lack of reliability.** <u>Possible content:</u>	3	**3 marks** for a clearly described issue associated with reliability which is well developed in the context of the study.

			• use of case study method • use of secondary data • many ideas open to interpretation • unrepresentative sample <u>Example of a 1-mark answer:</u> The use of the case study method produces unreliable findings as it cannot be replicated (1). (No context) <u>Example of a 2-mark answer:</u> Freud only studied Little Hans so his sample was unrepresentative (1) making the study unreliable (1). <u>Example of a 3-mark answer:</u> The case study method is unreliable as it not replicable (1) as we only have Freud's account and interpretation of Little Hans' experiences at the time (1) and it would be difficult for someone else to revisit this after the event to assess for reliability (1). Freud investigated Little Hans' fears and fantasies using the case study method which lacks reliability as it does not have a standardised procedure (1). The way in which Hans was observed and questioned by his father was not standardised as the situation was unique to Hans (1). The procedure cannot therefore be replicated to assess whether other boys experience the Oedipus complex/go through the five psychosexual stages of development (1). Other appropriate answer.		**2 marks** for a reasonably described and relevant issue associated with reliability loosely or vaguely linked to Freud's study **1 mark** for the mere identification of a relevant issue which is not in the context of the study. **0 marks** – no creditworthy response.
			Total	**3**	

1. Outline **one** weakness of the method used in Milgram's study into obedience.

[2]

2. Outline **one** weakness of the way the sample was organised in Loftus and Palmer's first experiment.

[2]

3.

Phombies (talk, don't walk!)

People seem to use their mobile phones more and more these days and it seems we prefer to text than actually talk to other people at times. When out and about some people even use their phone whilst still walking, looking like some kind of phone zombie (phombie), often so fixated on their display screen that they are completely oblivious of other people and their surroundings.

One psychologist conducted an observation study in a busy town centre to investigate this further and see if there was a difference in males' and females' use of the phone whilst walking. The data is presented below.



Number of males and females walking or not walking whilst using their mobile phone in a town centre		
	Walking whilst using phone	Not walking whilst using phone
Males	84	32
Females	58	26

Name the appropriate inferential statistical test to analyse the data from any one of your own practical activities. Give reasons for your answer.

[4]

4.

Doodling

Doodling can be described as ‘drawing whilst our attention is otherwise occupied’, such as when trying to listen to what someone is saying, or whilst trying to read something etc. Although it seems that this may make it less likely that we are able to focus on other things happening at the same time, some research suggests that it may actually improve our ability to concentrate. A psychologist wants to study this further by using the experimental method to investigate if people are able to concentrate better when doodling compared to when not doodling.



Outline **one** strength and **one** weakness of not having any qualitative data in this study.

[6]

5.

Doodling

Doodling can be described as ‘drawing whilst our attention is otherwise occupied’, such as when trying to listen to what someone is saying, or whilst trying to read something etc. Although it seems that this may make it less likely that we are able to focus on other things happening at the same time, some research suggests that it may actually improve our ability to concentrate. A psychologist wants to study this further by using the experimental method to investigate if people are able to concentrate better when doodling compared to when not doodling.



Outline one strength of having nominal data in this study.

[3]

6.

Phombies (talk, don’t walk!)

People seem to use their mobile phones more and more these days and it seems we prefer to text than actually talk to other people at times. When out and about some people even use their phone whilst still walking, looking like some kind of phone zombie (phombie), often so fixated on their display screen that they are completely oblivious of other people and their surroundings.

One psychologist conducted an observation study in a busy town centre to investigate this further and see if there was a difference in males’ and females’ use of the phone whilst walking. The data is presented below.



Number of males and females walking or not walking whilst using their mobile phone in a town centre		
	Walking whilst using phone	Not walking whilst using phone
Males	84	32
Females	58	26

What level of data has been collected in this study? Give reasons for your answer.

[2]

7.

Phombies (talk, don't walk!)

People seem to use their mobile phones more and more these days and it seems we prefer to text than actually talk to other people at times. When out and about some people even use their phone whilst still walking, looking like some kind of phone zombie (phombie), often so fixated on their display screen that they are completely oblivious of other people and their surroundings.

One psychologist conducted an observation study in a busy town centre to investigate this further and see if there was a difference in males' and females' use of the phone whilst walking. The data is presented below.



Number of males and females walking or not walking whilst using their mobile phone in a town centre		
	Walking whilst using phone	Not walking whilst using phone
Males	84	32
Females	58	26

Draw a fully labelled pie chart displaying the data from this study.

[4]

8.

Phombies (talk, don't walk!)

People seem to use their mobile phones more and more these days and it seems we prefer to text than actually talk to other people at times. When out and about some people even use their phone whilst still walking, looking like some kind of phone zombie (phombie), often so fixated on their display screen that they are completely oblivious of other people and their surroundings.

One psychologist conducted an observation study in a busy town centre to investigate this further and see if there was a difference in males' and females' use of the phone whilst walking. The data is presented below.



Number of males and females walking or not walking whilst using their mobile phone in a town centre		
	Walking whilst using phone	Not walking whilst using phone
Males	84	32
Females	58	26

Calculate the overall percentage of people using their phones whilst walking. Show your workings.

[3]

9(a).

Phombies (talk, don't walk!)

People seem to use their mobile phones more and more these days and it seems we prefer to text than actually talk to other people at times. When out and about some people even use their phone whilst still walking, looking like some kind of phone zombie (phombie), often so fixated on their display screen that they are completely oblivious of other people and their surroundings.

One psychologist conducted an observation study in a busy town centre to investigate this further and see if there was a difference in males' and females' use of the phone whilst walking. The data is presented below.



Number of males and females walking or not walking whilst using their mobile phone in a town centre		
	Walking whilst using phone	Not walking whilst using phone
Males	84	32
Females	58	26

Explain why the chi-square would be the appropriate non-parametric inferential statistical test to use to analyse the data from this study.

[3]

(b). Using the extract of the tables of critical values for the chi-square test presented below, what is the critical value at the 5% probability level for data collected in this study?

	Probability level		
Degrees of freedom (df)	0.5	0.05	0.01
1	0.455	3.841	6.635
2	1.386	5.991	9.210
3	2.366	7.815	11.345
4	3.357	9.488	13.277
5	4.351	11.070	15.086

[2]

(c). The calculated value of chi-square is 0.2681. Write the significance statement at the 5% level of probability for this study.

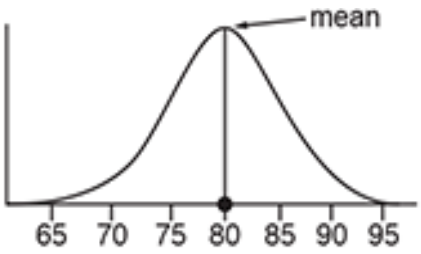
[3]

10. Which of these indicates the research is significant at the 2% level of probability?

A $p<2.0$
B $p<0.2$
C $p<0.02$
D $p>0.002$

Your answer ☐ **[1]**

11. What is the mode in the data displayed in this normal distribution curve where the mean is 80?



A 75
B 80
C 85
D 95

Your answer ☐ **[1]**

12. In a study of stress, which of the following would produce interval level data?

A asking participants to keep a diary of how they feel
B classifying people as ‘calm’ or ‘anxious’
C ratings on a zero to ten scale (0 = ‘relaxed’, 10 = ‘nervous’)
D temperature readings in degrees celsius

Your answer ☐ **[1]**

13. In the study by Freud, what kind of data was the information received about Little Hans in letters from his father?

A interval
B ordinal
C primary
D secondary

Your answer ☐ **[1]**

14. In the study of hemispheric deconnection by Sperry, what is the decimal for the duration in seconds that images were presented to the visual fields?

A 0.001
B 0.01
C 0.1
D 1.0

Your answer ☐ **[1]**

15. Which of these is an ethical consideration under the British Psychological Society (BPS) guidelines?

A rejection
B respect
C respiration
D respite

Your answer ☐ **[1]**

16. Which of these is the name given to the type of validity that refers to when research at least appears to be measuring what it claims to?

A face
B internal
C population
D test-retest

Your answer ☐ **[1]**

17. What is the median for this set of scores? 12, 8, 18, 16, 5, 9, 11, 18

A 11
B 11.5
C 12
D 18

Your answer ☐ **[1]**

18. What is the range for this set of scores? 32, 28, 15, 24, 9

A 9
B 19
C 23
D 32

Your answer ☐ **[1]**

19. Which correlation coefficient best relates to the data displayed in this scatter diagram?



A −0.7
B −0.07
C +0.7
D +0.07

Your answer ☐ **[1]**

20(a).

Weigh more outgoing?

Some research suggests that personality is linked to a person's body type (their structure and weight), with larger people having a more extrovert (lively and outgoing) personality than those who are smaller. To investigate this further, a psychologist wants to use the correlation technique to examine the relationship between weight and extroversion.

Outline how you could obtain secondary data to use as the measurement of the variable 'weight'.

[3]

(b). Outline **one** strength of the use of secondary data in this study.

[3]

21.

Weigh more outgoing?

Some research suggests that personality is linked to a person's body type (their structure and weight), with larger people having a more extrovert (lively and outgoing) personality than those who are smaller. To investigate this further, a psychologist wants to use the correlation technique to examine the relationship between weight and extroversion.

Outline **one** weakness of having quantitative data in this study.

[3]

22(a).

Weigh more outgoing?

Some research suggests that personality is linked to a person's body type (their structure and weight), with larger people having a more extrovert (lively and outgoing) personality than those who are smaller. To investigate this further, a psychologist wants to use the correlation technique to examine the relationship between weight and extroversion.

Suggest **one** open question that could provide information in the form of qualitative data for use in this study.

[2]

(b). Outline **one** strength of having some qualitative data in this study.

[3]

23.

Weigh more outgoing?

Some research suggests that personality is linked to a person's body type (their structure and weight), with larger people having a more extrovert (lively and outgoing) personality than those who are smaller. To investigate this further, a psychologist wants to use the correlation technique to examine the relationship between weight and extroversion.

Explain **one** way that the design of this study could increase the generalisability of the findings from this study.

[3]

24.

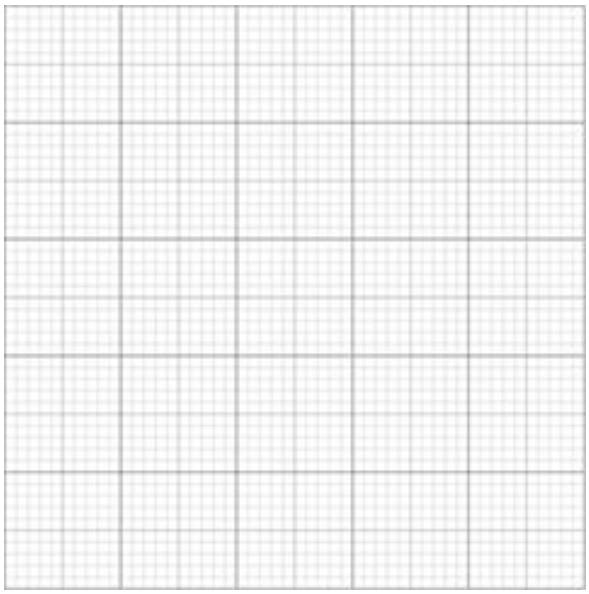
Steps to increase bin use

Research suggests that, when trying to encourage people to change their behaviour in some way, 'telling people what to do' is not always effective and a more subtle approach may be better ('nudge theory'). To investigate this, a psychologist conducted an observation study monitoring people's use of two different types of litter bin situated close to each other in the pedestrian area of one large town centre – one that had steps printed on the pavement leading up to it and another without the steps. Recordings were made each time any item was deposited in the bin throughout a continuous period from 10am to 2pm. The data on how many people used the bins is presented in the table below.

Table showing the number of times each bin was used by males and females

	Males	**Females**
Bin with steps	9	14
Bin without steps	5	12

Draw a fully labelled bar chart showing the overall use of the two different types of bin.



[4]

25.

Steps to increase bin use

Research suggests that, when trying to encourage people to change their behaviour in some way, 'telling people what to do' is not always effective and a more subtle approach may be better ('nudge theory'). To investigate this, a psychologist conducted an observation study monitoring people's use of two different types of litter bin situated close to each other in the pedestrian area of one large town centre – one that had steps printed on the pavement leading up to it and another without the steps. Recordings were made each time any item was deposited in the bin throughout a continuous period from 10am to 2pm. The data on how many people used the bins is presented in the table below.

Table showing the number of times each bin was used by males and females		
	Males	**Females**
Bin with steps	9	14
Bin without steps	5	12

Calculate the percentage of people who used the bin with steps leading up to it. Show your workings and present your finding to **two** significant figures.

[3]

26.

Steps to increase bin use

Research suggests that, when trying to encourage people to change their behaviour in some way, 'telling people what to do' is not always effective and a more subtle approach may be better ('nudge theory'). To investigate this, a psychologist conducted an observation study monitoring people's use of two different types of litter bin situated close to each other in the pedestrian area of one large town centre – one that had steps printed on the pavement leading up to it and another without the steps. Recordings were made each time any item was deposited in the bin throughout a continuous period from 10am to 2pm. The data on how many people used the bins is presented in the table below.

Table showing the number of times each bin was used by males and females

	Males	**Females**
Bin with steps	9	14
Bin without steps	5	12

Outline **two** conclusions that can be obtained from the data collected in this study.

[6]

27(a).

Steps to increase bin use

Research suggests that, when trying to encourage people to change their behaviour in some way, 'telling people what to do' is not always effective and a more subtle approach may be better ('nudge theory'). To investigate this, a psychologist conducted an observation study monitoring people's use of two different types of litter bin situated close to each other in the pedestrian area of one large town centre – one that had steps printed on the pavement leading up to it and another without the steps. Recordings were made each time any item was deposited in the bin throughout a continuous period from 10am to 2pm. The data on how many people used the bins is presented in the table below.

Table showing the number of times each bin was used by males and females

	Males	**Females**
Bin with steps	9	14
Bin without steps	5	12

The psychologist used the Chi-square test to analyse the data from this study. Give **one** reason why this would be the appropriate non-parametric inferential test to use.

[2]

(b). Calculate the degrees of freedom for use with the Chi-square test in this study. Show your workings.

[2]

(c). Using the extract from the table of critical values presented below, what is the critical value for use with the Chi-square test in this study at the 5% level of probability?

	Probability level					
df	**0.5**	**0.10**	**0.05**	**0.02**	**0.01**	**0.001**
1	0.455	2.706	3.841	5.412	6.635	10.827
2	1.386	4.605	5.991	7.824	9.210	13.815
3	2.366	6.251	7.815	9.837	11.345	16.268
4	3.357	7.779	9.488	11.668	13.277	18.465
5	4.351	9.236	11.070	13.388	15.086	20.517

[1]

(d). The psychologist obtained a calculated value of 0.4058 after analysing the data with the Chi-square test. Write a significance statement presenting this finding showing if the results are significant at the 5% level of probability or not.

[3]

(e). What does the analysis from the Chi-square test inform us regarding the use of the two different types of bin from this study?

[3]

28(a).

Steps to increase bin use

Research suggests that, when trying to encourage people to change their behaviour in some way, 'telling people what to do' is not always effective and a more subtle approach may be better ('nudge theory'). To investigate this, a psychologist conducted an observation study monitoring people's use of two different types of litter bin situated close to each other in the pedestrian area of one large town centre – one that had steps printed on the pavement leading up to it and another without the steps. Recordings were made each time any item was deposited in the bin throughout a continuous period from 10am to 2pm. The data on how many people used the bins is presented in the table below.

Table showing the number of times each bin was used by males and females

	Males	**Females**
Bin with steps	9	14
Bin without steps	5	12

Outline **one** strength of the use of the nominal data collected in this study.

[3]

(b). Outline **one** weakness of the use of the nominal data collected in this study.

[3]

29.

Steps to increase bin use

Research suggests that, when trying to encourage people to change their behaviour in some way, 'telling people what to do' is not always effective and a more subtle approach may be better ('nudge theory'). To investigate this, a psychologist conducted an observation study monitoring people's use of two different types of litter bin situated close to each other in the pedestrian area of one large town centre – one that had steps printed on the pavement leading up to it and another without the steps. Recordings were made each time any item was deposited in the bin throughout a continuous period from 10am to 2pm. The data on how many people used the bins is presented in the table below.

Table showing the number of times each bin was used by males and females

	Males	**Females**
Bin with steps	9	14
Bin without steps	5	12

Explain what it would mean if there was a 'Type 1 error' in this study.

[2]

30. Which of the calculated values from a Spearman's Rho test best reflect the correlation displayed in the scatter diagram below?



A $r_s = +0.2$
B $r_s = +0.8$
C $r_s = -0.2$
D $r_s = -0.8$

Your answer ☐ **[1]**

31. What name is given to the type of data collected directly by the researchers?

A preliminary
B precursor
C principal
D primary

Your answer ☐ **[1]**

32. What type of data is displayed in a histogram?

A anomalous
B continuous
C nominal
D ranked

Your answer ☐ **[1]**

33. What does the term 'representativeness' refer to?

A the extent to which the findings can be applied to everyday life
B the extent to which the findings can be applied to other related research
C the extent to which the findings can be applied to the population
D the extent to which the findings can be applied to the sample

Your answer ☐ **[1]**

34. What type of distribution refers to a situation in which the majority of participants score highly on the measure on which they are assessed?

A bimodal
B negatively skewed
C normal
D positively skewed

Your answer ☐ **[1]**

35. Which of these best describes what a 'Type 2' error refers to?

A incorrectly accepting the null hypothesis
B incorrectly rejecting the null hypothesis
C use of the incorrect inferential statistical test
D use of the incorrect tables of critical values

Your answer ☐ **[1]**

36. When would it be best to use the median rather than the mean?

A when there are a few scores much lower than the rest
B when there are a few scores very similar to each other
C when there is a large number of scores overall
D when there is a small number of scores overall

Your answer ☐ **[1]**

37. Which of these inferential statistical tests is the only one that does not involve ranking of the data at some stage?

A Chi-square
B Mann Whitney U test
C Spearman's Rho
D Wilcoxon Signed Ranks test

Your answer ☐ **[1]**

38. A study investigating the difference in aggression between boys and girls recorded the number of times a child shouted at someone during playtime. Which of the following would be the appropriate inferential test to use to analyse the data from this study?

A Binomial sign test
B Chi-square
C Mann Whitney U test
D Wilcoxon Signed Ranks test

Your answer ☐ **[1]**

39. Which symbol in statistics means ‘approximately’?

A α
B <
C ~
D Σ

Your answer ☐ **[1]**

40. What type of validity refers to the extent to which the scores on one measure are able to predict the outcome on another related measure?

A criterion
B ecological
C face
D population

Your answer ☐ **[1]**

41. For which of the following would it be appropriate to use a line graph to display the data?

A differences in the ability of males and females to identify six emotions displayed on the face
B number of acts of aggression displayed by boys and girls
C preference for four different learning styles displayed as percentages
D self-ratings of aggression (1 to 10) at different times of day (10am to 10pm)

Your answer ☐ **[1]**

42. What fraction is represented by the decimal 0.05?

A $\frac{1}{2}$
B $\frac{1}{5}$
C $\frac{1}{20}$
D $\frac{1}{50}$

Your answer ☐ **[1]**

43.

A question of morality
How do young children know right from wrong, what is good or bad or kind or unkind? Are there some things that are common to all children that underpin our sense of morality and shape the way we think about such things? The only way psychologists can investigate people's thoughts about such things is to ask them questions using the self-report method.

Outline **two** ways that you would address the ethical consideration of 'respect' in the design of this self-report study.

[6]

44.

A question of morality
How do young children know right from wrong, what is good or bad or kind or unkind? Are there some things that are common to all children that underpin our sense of morality and shape the way we think about such things? The only way psychologists can investigate people's thoughts about such things is to ask them questions using the self-report method.

Identify **one** thing that may lower the generalisability of the findings from this study.

[2]

45.

Exercise your stress

Stress affects many people, probably everyone at some stage of their life, and can be damaging to our health if it reaches a high level over a continuous period. Psychologists interested in finding ways to control and manage stress have investigated the correlation between exercise and stress by getting a group of ten volunteers to record the average amount of exercise taken each month and comparing this to the amount of stress experienced during the same period. The data obtained is presented in the table below.

Participant	Average number of hours exercise taken each month	Stress rating (0 = 'none' to 100 'high')
a	55	10
b	25	50
c	30	25
d	0	90
e	40	30
f	50	30
g	20	80
h	45	20
i	15	60
j	60	100

Outline **two** conclusions that can be made from the data collected in this study.

[6]

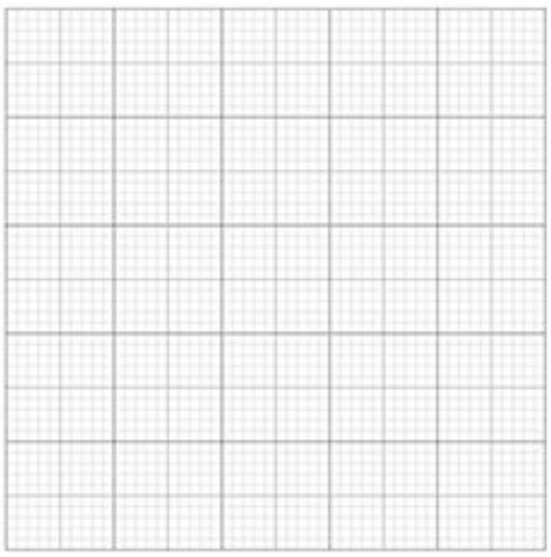
46.

Exercise your stress

Stress affects many people, probably everyone at some stage of their life, and can be damaging to our health if it reaches a high level over a continuous period. Psychologists interested in finding ways to control and manage stress have investigated the correlation between exercise and stress by getting a group of ten volunteers to record the average amount of exercise taken each month and comparing this to the amount of stress experienced during the same period. The data obtained is presented in the table below.

Participant	Average number of hours exercise taken each month	Stress rating (0 = 'none' to 100 'high')
a	55	10
b	25	50
c	30	25
d	0	90
e	40	30
f	50	30
g	20	80
h	45	20
i	15	60
j	60	100

Draw a fully labelled scatter diagram displaying the data from this study



[4]

47.

Exercise your stress

Stress affects many people, probably everyone at some stage of their life, and can be damaging to our health if it reaches a high level over a continuous period. Psychologists interested in finding ways to control and manage stress have investigated the correlation between exercise and stress by getting a group of ten volunteers to record the average amount of exercise taken each month and comparing this to the amount of stress experienced during the same period. The data obtained is presented in the table below.

Participant	Average number of hours exercise taken each month	Stress rating (0 = 'none' to 100 'high')
a	55	10
b	25	50
c	30	25
d	0	90
e	40	30
f	50	30
g	20	80
h	45	20
i	15	60
j	60	100

Outline **two** things that could have affected the validity of the data collected in this study.

[6]

48.

Exercise your stress

Stress affects many people, probably everyone at some stage of their life, and can be damaging to our health if it reaches a high level over a continuous period. Psychologists interested in finding ways to control and manage stress have investigated the correlation between exercise and stress by getting a group of ten volunteers to record the average amount of exercise taken each month and comparing this to the amount of stress experienced during the same period. The data obtained is presented in the table below.

Participant	Average number of hours exercise taken each month	Stress rating (0 = 'none' to 100 'high')
a	55	10
b	25	50
c	30	25
d	0	90
e	40	30
f	50	30
g	20	80
h	45	20
i	15	60
j	60	100

Explain why the Spearman's Rho would be the appropriate non-parametric inferential statistical test to use to analyse the data from this study.

[3]

49.

Exercise your stress

Stress affects many people, probably everyone at some stage of their life, and can be damaging to our health if it reaches a high level over a continuous period. Psychologists interested in finding ways to control and manage stress have investigated the correlation between exercise and stress by getting a group of ten volunteers to record the average amount of exercise taken each month and comparing this to the amount of stress experienced during the same period. The data obtained is presented in the table below.

Participant	Average number of hours exercise taken each month	Stress rating (0 = 'none' to 100 'high')
a	55	10
b	25	50
c	30	25
d	0	90
e	40	30
f	50	30
g	20	80
h	45	20
i	15	60
j	60	100

After carrying out a Spearman's Rho inferential statistical test the calculated value obtained was $r_s = -0.37$. Explain what this means.

[3]

50(a).

Exercise your stress

Stress affects many people, probably everyone at some stage of their life, and can be damaging to our health if it reaches a high level over a continuous period. Psychologists interested in finding ways to control and manage stress have investigated the correlation between exercise and stress by getting a group of ten volunteers to record the average amount of exercise taken each month and comparing this to the amount of stress experienced during the same period. The data obtained is presented in the table below.

Participant	Average number of hours exercise taken each month	Stress rating (0 = 'none' to 100 'high')
a	55	10
b	25	50
c	30	25
d	0	90
e	40	30
f	50	30
g	20	80
h	45	20
i	15	60
j	60	100

The standard deviation was calculated for the average number of hours exercised each month and found to be 18.275666882497

What is this written to two significant figures?

---**[1]**

(b). Explain what this informs us about the average number of hours exercised each month by participants in this study.

---**[2]**

51. Put the letter of the correct answer in the box provided.

Which of these indicates the research is significant at the 1% level of probability?

A $p<1.0$
B $p<0.1$
C $p<0.01$
D $p>0.01$

Your answer ☐ **[1]**

52. Put the letter of the correct answer in the box provided.

What type of data is obtained from other (often already existing) sources, rather than directly by the researchers?

A nominal
B primary
C secondary
D qualitative

Your answer ☐ **[1]**

53. Put the letter of the correct answer in the box provided.

In which type of distribution is the mean, median and mode all the same score?

A a negatively skewed distribution
B a normal distribution
C a positively skewed distribution
D an asymmetrical distribution

Your answer ☐ **[1]**

54. Put the letter of the correct answer in the box provided.

What is Na in the formula for the Mann Whitney U test below?

$$Ua = NaNb + \left[\frac{Na(Na + 1)}{2}\right] - Ra \qquad Ub = (NaNb) - Ua$$

A mean of condition a
B number of participants in condition a
C the number of participants in condition a multiplied by the number of participants in condition b
D total of the ranks for condition a

Your answer ☐ **[1]**

55. Put the letter of the correct answer in the box provided.

What is the median for this set of scores? 21, 14, 20, 21, 12, 17

A 17
B 17.5
C 18.5
D 20.5

Your answer ☐ **[1]**

56(a).

Stand to attention. An educational psychologist conducted a study to investigate if getting pupils to stand up when completing some of their work in class increased their performance by making them concentrate more. An independent measures design experiment was used testing a small group of 12 pupils. Half of the class stood up at their desks whilst completing a maths test whilst the others remained seated. The data obtained from the study is presented in the table below.

Score on maths test (max 20) when stood up or sat down when taking the test			
Stood up		Sat down	
participant	score	participant	score
1	18	1	14
2	20	2	8
3	17	3	20
4	15	4	4
5	18	5	15
6	19	6	12

Calculate the mean for the ‘stood up’ condition and present your findings to 2 decimal places. Show your workings.

---------- **[2]**

(b). Calculate the mean for the ‘sat down’ condition and present your findings to 2 significant figures. Show your workings.

[2]

(c). Calculate the mean percentage score in each condition. Show your workings.

[2]

57(a).

Stand to attention. An educational psychologist conducted a study to investigate if getting pupils to stand up when completing some of their work in class increased their performance by making them concentrate more. An independent measures design experiment was used testing a small group of 12 pupils. Half of the class stood up at their desks whilst completing a maths test whilst the others remained seated. The data obtained from the study is presented in the table below.

Score on maths test (max 20) when stood up or sat down when taking the test			
Stood up		Sat down	
participant	score	participant	score
1	18	1	14
2	20	2	8
3	17	3	20
4	15	4	4
5	18	5	15
6	19	6	12

Explain how you would calculate the standard deviation for each condition of this study.

[5]

(b). What information would the standard deviation provide if it was calculated for the data in this study?

[2]

(c). The standard deviation for each condition of this study is presented in the table below. What do these findings inform us about the effect of standing up or sitting down when performing a test?

Stood up	Sat down
1.72	5.60

[4]

58(a).

Stand to attention. An educational psychologist conducted a study to investigate if getting pupils to stand up when completing some of their work in class increased their performance by making them concentrate more. An independent measures design experiment was used testing a small group of 12 pupils. Half of the class stood up at their desks whilst completing a maths test whilst the others remained seated. The data obtained from the study is presented in the table below.

Score on maths test (max 20) when stood up or sat down when taking the test			
Stood up		Sat down	
participant	score	participant	score
1	18	1	14
2	20	2	8
3	17	3	20
4	15	4	4
5	18	5	15
6	19	6	12

What would be the appropriate non-parametric inferential statistical test to use to analyse the data from this study? Give reasons for your answer.

[2]

(b). Outline how the data in this study would be ranked before using the inferential statistical test.

[2]

59(a).

Curious findings. Curiosity can be regarded as simply ‘a desire to know things’ – having an inquisitive mind. However, some people may be more curious than others. To investigate this a psychologist secretly observed students who were left alone in a room waiting to be interviewed about their experiences at university shortly before they graduated. In the room was a box with a sign in front of it that clearly said ‘do not open the box’.

Type of student	DID open the box	Did NOT open the box
Science	39	21
Arts	15	45
Humanities	8	22

What would be the appropriate non-parametric inferential statistical test to use to analyse the data from this study? Give reasons for your answer.

[3]

(b). Explain how the expected frequency (E) values for each cell would be calculated before the formula for this inferential test could be used.

[2]

(c). Explain what the symbol (>) would mean in the significance statement when reporting the findings from the inferential statistical test for this study.

[2]

60(a). A psychologist conducted a survey assessing beliefs in the paranormal with a sample of 144 people.

25% of people said they believe in ghosts. How many people is this?

A 25
B 36
C 40
D 72

Your answer ☐ **[1]**

(b). 24 people said they believe aliens exist. What fraction of the sample is this?

A 1/3
B 1/4
C 1/6
D 2/3

Your answer ☐ **[1]**

(c). What is the ratio of people who believe in ghosts to aliens?

A 1:4
B 2:3
C 3:2
D 4:1

Your answer ☐ **[1]**

61(a).

Love is in the (question) **air.** We can fall 'into it' and 'out of it', it can make us feel both happy and sad or even angry and mad, but what is love? Psychologists decided to investigate this further by using the self-report method in which they posted out a questionnaire to members of the public to complete. Some of the data from the first twenty males and first twenty females to respond are presented in the tables below.

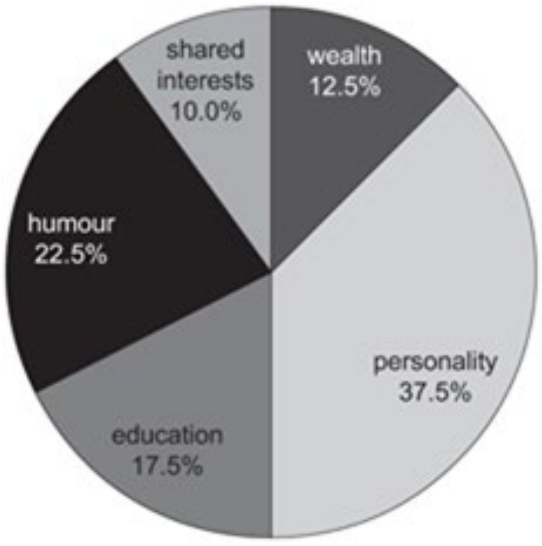
Table 1: Ratings of the importance of physical appearance for being in love given by male and female participants (1= 'not very important' to 10 = 'extremely important')

males	females
7	3
5	7
9	1
10	6
8	4
7	9
7	6
9	3
8	5
10	3
9	5
10	3
9	5
9	5
7	3
6	7
5	5
6	1
7	3
9	2

Table 2: Responses to the question … *'do you believe in love at first sight?'*

	yes	no
males	5	15
females	12	8

Figure 1
Pie chart showing the percentage of people who thought which attributes were the most important for being in love with someone



A partly completed table of measures of central tendency for the data collected in Table 1 is presented below.

Measures of central tendency for the ratings of the importance of physical appearance for being in love given by male and female participants		
	males	females
mean		4.3
median	8.0	
mode	9.0	

Answer the following questions so that the rest of the table can be completed. Show all your workings.

Calculate the mean rating of the importance of physical appearance for being in love given by males. Write your answer to two significant figures.

[3]

(b). Calculate the median rating of the importance of physical appearance for being in love given by females.

[2]

(c). What is the mode for the importance of physical appearance for being in love given by females?

[1]

62(a).

Love is in the (question) **air.** We can fall 'into it' and 'out of it', it can make us feel both happy and sad or even angry and mad, but what is love? Psychologists decided to investigate this further by using the self-report method in which they posted out a questionnaire to members of the public to complete. Some of the data from the first twenty males and first twenty females to respond are presented in the tables below.

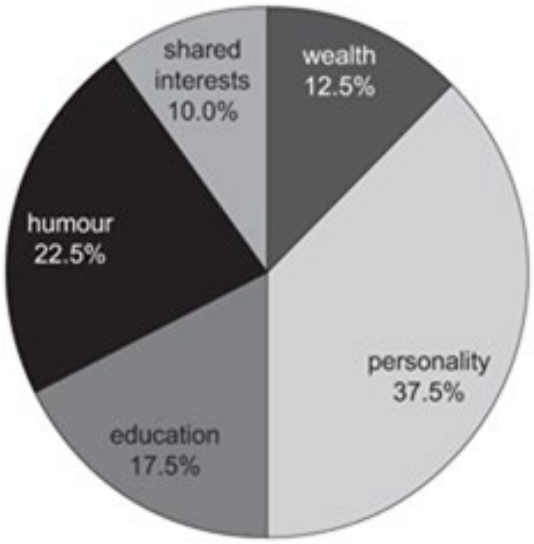
Table 1: Ratings of the importance of physical appearance for being in love given by male and female participants (1= 'not very important' to 10 = 'extremely important')

males	females
7	3
5	7
9	1
10	6
8	4
7	9
7	6
9	3
8	5
10	3
9	5
10	3
9	5
9	5
7	3
6	7
5	5
6	1
7	3
9	2

Table 2: Responses to the question … *'do you believe in love at first sight?'*

	yes	no
males	5	15
females	12	8

Figure 1
Pie chart showing the percentage of people who thought which attributes were the most important for being in love with someone



What is the range for the rating of the importance of physical appearance for being in love given by males **and** females? Show your workings.

[4]

(b). Outline **one** conclusion from the calculation of the range for the rating of the importance of physical appearance for being in love given by males and females.

[3]

63(a).

Love is in the (question) **air.** We can fall 'in to it' and 'out of it', it can make us feel both happy and sad or even angry and mad, but what is love? Psychologists decided to investigate this further by using the self-report method in which they posted out a questionnaire to members of the public to complete. Some of the data from the first twenty males and first twenty females to respond are presented in the tables below.

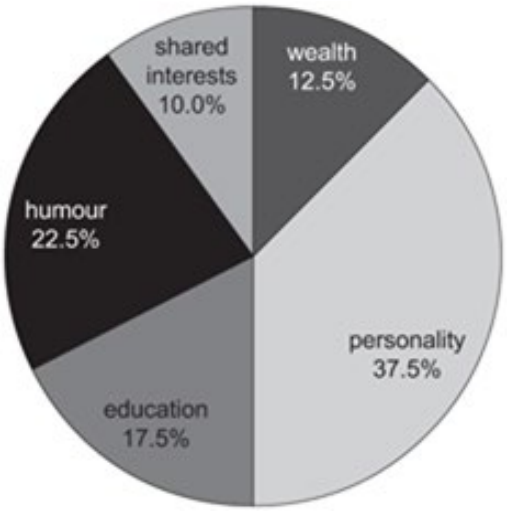
Table 1: Ratings of the importance of physical appearance for being in love given by male and female participants (1= 'not very important' to 10 = 'extremely important')

males	females
7	3
5	7
9	1
10	6
8	4
7	9
7	6
9	3
8	5
10	3
9	5
10	3
9	5
9	5
7	3
6	7
5	5
6	1

Table 2: Responses to the question … *'do you believe in love at first sight?'*

	yes	no
males	5	15
females	12	8

Figure 1
Pie chart showing the percentage of people who thought which attributes were the most important for being in love with someone



7	3
9	2

Using the formula provided calculate the value of Chi-square for the data in the table below. The E values (expected frequencies) have already been provided (in the table in italics). Show your workings.

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

Responses to the question … *'do you believe in love at first sight?'*		
	yes	no
males	5 *(8.50)*	15 *(11.50)*
females	12 *(8.50)*	8 *(11.50)*

[5]

(b). Using the extract of the tables of critical values for the Chi-square test presented below, what is the critical value at the 5% probability level for data collected in this study?

	Probability level		
Degrees of freedom (df)	0.5	0.05	0.01
1	0.455	3.841	6.635
2	1.386	5.991	9.210
3	2.366	7.815	11.345
4	3.357	9.488	13.277
5	4.351	11.070	15.086

[2]

(c). Write the significance statement for the analysis performed on this data using the Chi-square test.

[2]

64(a).

A psychologist was interested in the relationship between individuals' language ability and their physical co-ordination. She used a correlational analysis to investigate this in a self-selected sample of 34 adult participants.

Each participant completed a language ability test. This test involved both having to spell a list of 20 words and answering 20 closed questions on the meaning of words. The participants' physical co-ordination was then calculated by rating them on six different tasks including balancing on a beam, catching a ball and copying a dance routine.

Outline how each of the following ethical considerations could have been dealt with in this study.

Confidentiality

[2]

(b). Informed consent

[2]

(c). Protection of participant

[2]

65. What is 7.864 written to two significant figures?

A 7.8
B 7.86
C 7.9
D 8.0

Your answer ☐ **[1]**

66. What is the median for this set of data 10, 12, 6, 5, 14, 3?

A 6.5
B 7.5
C 8.0
D 16.0

Your answer ☐ **[1]**

67. What is the collective term for the mean, median and mode?

A measures of central dispersion
B measures of central range
C measures of central significance
D measures of central tendency

Your answer ☐ **[1]**

68. Which of these is a non-parametric inferential statistical test?

A Mann Whitney U
B Mean
C Range
D standard deviation

Your answer ☐ **[1]**

69. A newspaper includes 12 pages of sport and eight pages of TV. What is the ratio of sport to TV? Give your answer in its simplest form.

A 2:3
B 3:2
C 8:12
D 12:8

Your answer ☐ **[1]**

70. What is 2,830 written in standard form?

A 2.83×10^2
B 2.83×10^3
C 2.83×10^4
D 2.83×10^5

Your answer ☐ **[1]**

71. Which of these is nominal data?

A classifying participants as 'good', 'average' or 'poor'
B ratings by an independent observer on a scale of 1 (very good) to 10 (very poor)
C self-ratings by participants on a scale of 1 (very good) to 10 (very poor)
D use of a stopwatch to record reaction time

Your answer ☐ **[1]**

72. Which of these best describes the type of data collected in Loftus and Palmer's study investigating the effects of language on memory?

A both quantitative and qualitative
B neither qualitative or quantitative
C qualitative
D quantitative

Your answer ☐ **[1]**

73.

An educational psychologist conducted research investigating if there was any correlation between performance in two different but related subjects studied by GCSE students. If a student is good at one, are they also good at the other, and if they perform poorly in one does that mean that they will also struggle to do well in the other one? Data from a small sample is presented in the table below.

Table showing test results from two different subjects studied by a group of participants

Participant	Maths test score	Physics test score
a	20	23
b	15	17
c	6	24
d	28	30
e	8	12
f	11	14
g	22	24
h	4	7

Identify **two** findings from the data presented in this table.

[4]

74.

An educational psychologist conducted research investigating if there was any correlation between performance in two different, but related subjects studied by GCSE students. If a student is good at one, are they also good at the other, and if they perform poorly in one does that mean that they will also struggle to do well in the other one? Data from a small sample is presented in the table below.

Table showing test results from two different subjects studied by a group of participants

Participant	Maths test score	Physics test score
a	20	23
b	15	17
c	6	24
d	28	30
e	8	12
f	11	14
g	22	24
h	4	7

Draw a scatter diagram displaying the results of this study.

[4]

75(a).

An educational psychologist conducted research investigating if there was any correlation between performance in two different, but related subjects studied by GCSE students. If a student is good at one, are they also good at the other, and if they perform poorly in one does that mean that they will also struggle to do well in the other one? Data from a small sample is presented in the table below.

Table showing test results from two different subjects studied by a group of participants

Participant	Maths test score	Physics test score
a	20	23
b	15	17
c	6	24
d	28	30
e	8	12
f	11	14
g	22	24
h	4	7

Calculate the range for each test taken.

[2]

(b). What conclusion can be reached by interpreting the range for each test?

[4]

(c). Suggest **one** advantage of using standard deviation instead of the range to analyse the data from each test.

[3]

76(a).

An educational psychologist conducted research investigating if there was any correlation between performance in two different, but related subjects studied by GCSE students. If a student is good at one, are they also good at the other, and if they perform poorly in one does that mean that they will also struggle to do well in the other one? Data from a small sample is presented in the table below.

Table showing test results from two different subjects studied by a group of participants

Participant	Maths test score	Physics test score
a	20	23
b	15	17
c	6	24
d	28	30
e	8	12
f	11	14
g	22	24
h	4	7

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.

[2]

(b). Explain how the data would be ranked for use in this test.

[2]

(c). The inferential test result produced a calculated value of +0.7083. Explain what this means.

[3]

77(a).

An educational psychologist conducted research investigating if there was any correlation between performance in two different, but related subjects studied by GCSE students. If a student is good at one, are they also good at the other, and if they perform poorly in one does that mean that they will also struggle to do well in the other one? Data from a small sample is presented in the table below.

Table showing test results from two different subjects studied by a group of participants

Participant	Maths test score	Physics test score
a	20	23
b	15	17
c	6	24
d	28	30
e	8	12
f	11	14
g	22	24
h	4	7

What does the term ‘critical value’ refer to?

[1]

(b). How would the critical value be obtained in this study?

[2]

78. Which of these refers to making a Type 1 error?

A abandoning the null hypothesis
B ignoring the null hypothesis
C incorrectly accepting the null hypothesis
D incorrectly rejecting the null hypothesis

Your answer ☐ **[1]**

79. Which of these best describes what the variance is?

A measure of central tendency
B measure of dispersion around the mean
C measure of proportionality
D measure of quantifiable data

Your answer ☐ **[1]**

80. Which of these indicates the research is significant at the 5% level of probability?

A $p<0.005$
B $p<0.05$
C $p>0.005$
D $p>0.05$

Your answer ☐ **[1]**

81.

Love is in the (question) **air.** We can fall 'in to it' and 'out of it', it can make us feel both happy and sad or even angry and mad, but what is love? Psychologists decided to investigate this further by using the self-report method in which they posted out a questionnaire to members of the public to complete. Some of the data from the first twenty males and first twenty females to respond are presented in the tables below.

Table 1: Ratings of the importance of physical appearance for being in love given by male and female participants (1= 'not very important' to 10 = 'extremely important')

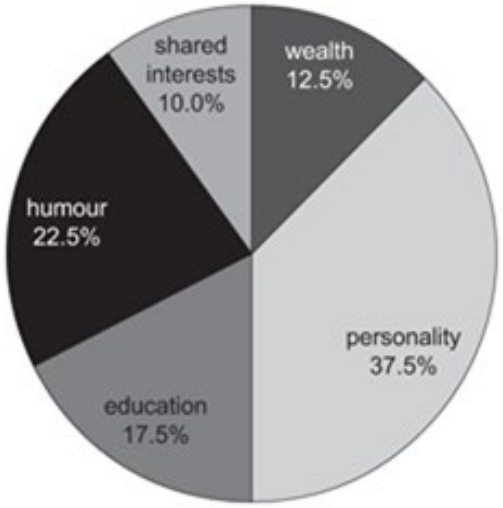
males	females
7	3
5	7
9	1
10	6
8	4
7	9
7	6
9	3
8	5
10	3
9	5
10	3
9	5

Table 2: Responses to the question … *'do you believe in love at first sight?'*

	yes	no
males	5	15
females	12	8

Figure 1
Pie chart showing the percentage of people who thought which attributes were the most important for being in love with someone

9	5
7	3
6	7
5	5
6	1
7	3
9	2



Evaluate the population validity of the data collected in this study.

[6]

82. Which of these inferential statistical tests does not require the data to be ranked as part of the calculation?

A Chi-square
B Mann-Whitney U test
C Spearman's Rho
D Wilcoxon Signed Ranks test

Your answer ☐ **[1]**

83. Which of these is an advantage of secondary data?

A already exists
B easy to analyse
C easy to interpret
D ecologically valid

Your answer ☐ **[1]**

84. What is meant by the term 'significant result'?

A it is a figure that you compare the answer from an inferential statistical test with
B it is an answer that exceeds a certain probability level
C it is an answer that tells us something important
D it is an approximate answer

Your answer ☐ **[1]**

85. In research terms, what is meant by 'social desirability'?

A responding in a way that is approved of by society
B responding in a way to be perceived as more friendly
C responding in a way to please the researcher
D responding in a way to provide the researcher with what is expected

Your answer ☐ **[1]**

86. Which of these different types of data would not include any information in the form of words?

A interval
B nominal
C qualitative
D secondary

Your answer ☐ **[1]**

87. Which of these could not be the answer from a Spearman's Rho inferential statistical test?

A –0.728
B 0.3
C 0.892
D 1.52

Your answer ☐ **[1]**

88. What is the mode in this set of data?

17, 19, 12, 23, 17, 25, 19, 17

A 17
B 17.5
C 18
D 19

Your answer ☐ **[1]**

89. The variance of a set of scores is 14.44. What is the standard deviation?

A 1.4
B 3.8
C 14.0
D 208.51

Your answer ☐ **[1]**

90. What type of data was collected in the Piliavin et al. study for the dependent variable of length of time that it took for help to be offered?

A interval
B non-parametric
C qualitative
D secondary

Your answer ☐ **[1]**

91.

Travel log. How do people pass the time on a long journey? Some people seem to cope better than others and manage to keep themselves occupied or simply don't mind (perhaps even enjoy) the rest. Others find it more difficult and can become bored and restless. Psychologists want to use the naturalistic observation method to investigate this and find out if some types of behaviour and actions are more common than others.

Describe **two** things that may influence the inter-rater reliability of this study.

[6]

92.

Travel log. How do people pass the time on a long journey? Some people seem to cope better than others and manage to keep themselves occupied or simply don't mind (perhaps even enjoy) the rest. Others find it more difficult and can become bored and restless. Psychologists want to use the naturalistic observation method to investigate this and find out if some types of behaviour and actions are more common than others.

Describe **two** ways you would address the ethical consideration of 'responsibility' in relation to this study.

[6]

93.

Love is in the (question) **air.** We can fall 'into it' and 'out of it', it can make us feel both happy and sad or even angry and mad, but what is love? Psychologists decided to investigate this further by using the self-report method in which they posted out a questionnaire to members of the public to complete. Some of the data from the first twenty males and first twenty females to respond are presented in the tables below.

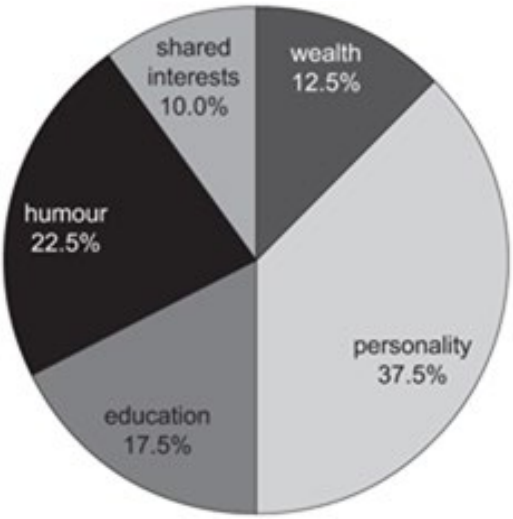
Table 1: Ratings of the importance of physical appearance for being in love given by male and female participants (1= 'not very important' to 10 = 'extremely important')

males	females
7	3
5	7
9	1
10	6
8	4
7	9
7	6
9	3
8	5
10	3
9	5
10	3
9	5
9	5
7	3
6	7
5	5
6	1
7	3
9	2

Table 2: Responses to the question … *'do you believe in love at first sight?'*

	yes	no
males	5	15
females	12	8

Figure 1
Pie chart showing the percentage of people who thought which attributes were the most important for being in love with someone



Using the data presented in the pie chart in Figure 1, calculate the ratio of how many people said that personality was the most important thing for love compared to those that said wealth was. Show your workings.

[4]

94. What is the name given to data before any analysis is performed?

A interval
B ordinal
C quantitative
D raw

Your answer [1]

95. What is the name for the type of reliability that involves dividing a test into two parts and comparing scores on both parts of the test?

A semi-structured
B split-half
C test-retest
D two-tailed

Your answer [1]

96.

Curious findings. Curiosity can be regarded as simply 'a desire to know things' – having an inquisitive mind. However, some people may be more curious than others. To investigate this a psychologist secretly observed students who were left alone in a room waiting to be interviewed about their experiences at university shortly before they graduated. In the room was a box with a sign in front of it that clearly said, 'do not open the box'.

Type of student	DID open the box	Did NOT open the box
Science	39	21
Arts	15	45
Humanities	8	22

Outline **two** things that may affect the reliability of the data collected in this study (and influence whether the same / similar findings would be obtained if repeated).

[6]

97(a). Under what circumstances would you use the Mann-Whitney U test?

A with independent measures design and nominal data
B with independent measures design and ordinal data
C with repeated measures design and nominal data
D with repeated measures design and ordinal data

Your answer ☐ **[1]**

(b). What is 'Ra' in the formula for the calculation of the Mann-Whitney U test?

$$Ua = NaNb + \left[\frac{Na(Na+1)}{2}\right] - Ra \qquad Ub = (NaNb) - Ua$$

A mean of condition 'a'
B number of participants in condition 'a'
C square root of condition 'a'
D sum of the ranks of condition 'a'

Your answer ☐ **[1]**

98. The following scores were obtained in a memory test from one group of participants:
10, 18, 24, 9, 11, 14, 23, 8, 19, 17, 14, 20

Using the formula below, what would the value of 'n' be when calculating the standard deviation for this set of scores?

$$SD = \sqrt{\frac{\sum(x - \bar{x})^2}{n}}$$

A 10
B 11
C 12
D 13

Your answer ☐ **[1]**

99. When would it be best to use the median rather than the mean?

A when there are a few scores much higher than the rest
B when there are few scores similar to the rest
C when there are lots of high scores
D when there are lots of low scores

Your answer ☐ **[1]**

100. Two different versions of a personality test are given to participants and then the results compared. What type of validity is being checked here?

A concurrent
B criterion
C ecological
D face

Your answer ☐ **[1]**

101. Which best describes what a 'type 1 error' refers to?

A null hypothesis incorrectly accepted
B null hypothesis incorrectly rejected
C use of the wrong inferential statistical test
D use of wrong table of critical values

Your answer ☐ **[1]**

102. Which of these is not a measure of dispersion?

A mean
B range
C standard deviation
D variance

Your answer ☐ **[1]**

103. What is the median of this set of scores?
15, 14, 8, 9, 12, 6, 10, 3, 2, 9

A 9
B 9.5
C 10.5
D 18

Your answer ☐ **[1]**

104. Which of the following is one of the criteria for use of a parametric inferential statistical test?

A a repeated measures design must have been used
B the level of data collected must be nominal
C the means of each condition must be similar
D the variance in the data in each condition must be similar

Your answer ☐ **[1]**

105(a). A researcher asked 50 people what their favourite type of food was and obtained the following results:

Type of food	No. of people reporting food as their 'favourite'
meat	10
fish	18
fruit	8
cereals	5
vegetables	9

What percentage of people reported fruit as their favourite?

A 4%
B 8%
C 16%
D 20%

Your answer ☐ **[1]**

(b). What kind of data is this?

A interval
B nominal
C ordinal
D qualitative

Your answer ☐ **[1]**

106. Which of the following best describes what the term 'representativeness' refers to?

A the extent to which the findings are able to be applied to the population
B the extent to which the findings are able to be published
C the extent to which the findings are reliable
D the extent to which the findings predict future outcomes

Your answer ☐ **[1]**

107(a).

Sleepy head? It's something we all do at least once a day yet varies a lot between people and can influence many other aspects of our life. The amount of sleep we get and the quality of the sleep can have both physical and cognitive consequences. However, too much sleep may be as bad as not getting enough. There are clearly lots of opportunities here to explore the relationship between sleep and behaviour. One such area of research could be to investigate if there is a correlation between the amount of sleep a person has and their ability to concentrate the next day.

Describe **one** strength of having quantitative data in this study.

---

[3]

(b). Describe **one** weakness of having quantitative data in this study.

---

[3]

108.

Sleepy head? It's something we all do at least once a day yet varies a lot between people and can influence many other aspects of our life. The amount of sleep we get and the quality of the sleep can have both physical and cognitive consequences. However, too much sleep may be as bad as not getting enough. There are clearly lots of opportunities here to explore the relationship between sleep and behaviour. One such area of research could be to investigate if there is a correlation between the amount of sleep a person has and their ability to concentrate the next day.

Outline **one** way that some qualitative data could be obtained in this study.

[3]

109(a).

Sleepy head? It's something we all do at least once a day yet varies a lot between people and can influence many other aspects of our life. The amount of sleep we get and the quality of the sleep can have both physical and cognitive consequences. However, too much sleep may be as bad as not getting enough. There are clearly lots of opportunities here to explore the relationship between sleep and behaviour. One such area of research could be to investigate if there is a correlation between the amount of sleep a person has and their ability to concentrate the next day.

Outline **one** way that the external validity of this study could be considered to be high.

[3]

(b). Outline **one** way that the external validity of this study could be considered to be low.

[3]

110.

Sleepy head? It's something we all do at least once a day yet varies a lot between people and can influence many other aspects of our life. The amount of sleep we get and the quality of the sleep can have both physical and cognitive consequences. However, too much sleep may be as bad as not getting enough. There are clearly lots of opportunities here to explore the relationship between sleep and behaviour. One such area of research could be to investigate if there is a correlation between the amount of sleep a person has and their ability to concentrate the next day.

Explain what the term 'demand characteristic' refers to in relation to your study.

[2]

111.

Curious findings. Curiosity can be regarded as simply 'a desire to know things' – having an inquisitive mind. However, some people may be more curious than others. To investigate this a psychologist secretly observed students who were left alone in a room waiting to be interviewed about their experiences at university shortly before they graduated. In the room was a box with a sign in front of it that clearly said, 'do not open the box'.

Type of student	DID open the box	Did NOT open the box
Science	39	21
Arts	15	45
Humanities	8	22

Calculate the percentage of science students who 'did' and 'did not' open the box. Show your workings.

[4]

112.

Curious findings. Curiosity can be regarded as simply ‘a desire to know things’ – having an inquisitive mind. However, some people may be more curious than others. To investigate this a psychologist secretly observed students who were left alone in a room waiting to be interviewed about their experiences at university shortly before they graduated. In the room was a box with a sign in front of it that clearly said, ‘do not open the box’.

Type of student	DID open the box	Did NOT open the box
Science	39	21
Arts	15	45
Humanities	8	22

What was the ratio of arts students who did not open the box to those who did?

[2]

113.

Curious findings. Curiosity can be regarded as simply ‘a desire to know things’ – having an inquisitive mind. However, some people may be more curious than others. To investigate this a psychologist secretly observed students who were left alone in a room waiting to be interviewed about their experiences at university shortly before they graduated. In the room was a box with a sign in front of it that clearly said, ‘do not open the box’.

Type of student	DID open the box	Did NOT open the box
Science	39	21
Arts	15	45
Humanities	8	22

Draw a fully labelled bar chart to display the data collected in this study.

[4]

114.

Curious findings. Curiosity can be regarded as simply ‘a desire to know things’ – having an inquisitive mind. However, some people may be more curious than others. To investigate this a psychologist secretly observed students who were left alone in a room waiting to be interviewed about their experiences at university shortly before they graduated. In the room was a box with a sign in front of it that clearly said, ‘do not open the box’.

Type of student	DID open the box	Did NOT open the box
Science	39	21
Arts	15	45
Humanities	8	22

Outline **two** conclusions that can be reached from the raw data collected in this study.

---

[6]

115.

Curious findings. Curiosity can be regarded as simply ‘a desire to know things’ – having an inquisitive mind. However, some people may be more curious than others. To investigate this a psychologist secretly observed students who were left alone in a room waiting to be interviewed about their experiences at university shortly before they graduated. In the room was a box with a sign in front of it that clearly said ‘do not open the box’.

Type of student	DID open the box	Did NOT open the box
Science	39	21
Arts	15	45
Humanities	8	22

Outline **one** ethical issue arising from how data was obtained in this study.

[3]

116. Which is an example of interval level data?

A the mass, in grams, of the brain of an individual with schizophrenia
B the number of nightmares experienced by an individual with schizophrenia
C the number of times an individual with schizophrenia has been admitted to hospital
D the rating of the severity of the delusions experienced by an individual with schizophrenia

Your answer ☐ **[1]**

117.

A psychologist used an observation to investigate the effect of environment on individuals' need for personal space. They decided to carry out a covert observation in three settings: a nightclub, a college library and the changing room in a leisure centre. Members of the public using the facilities made up the sample. The psychologist observed key behaviours, such as reduced eye contact, defensive body posture and movement away from people.

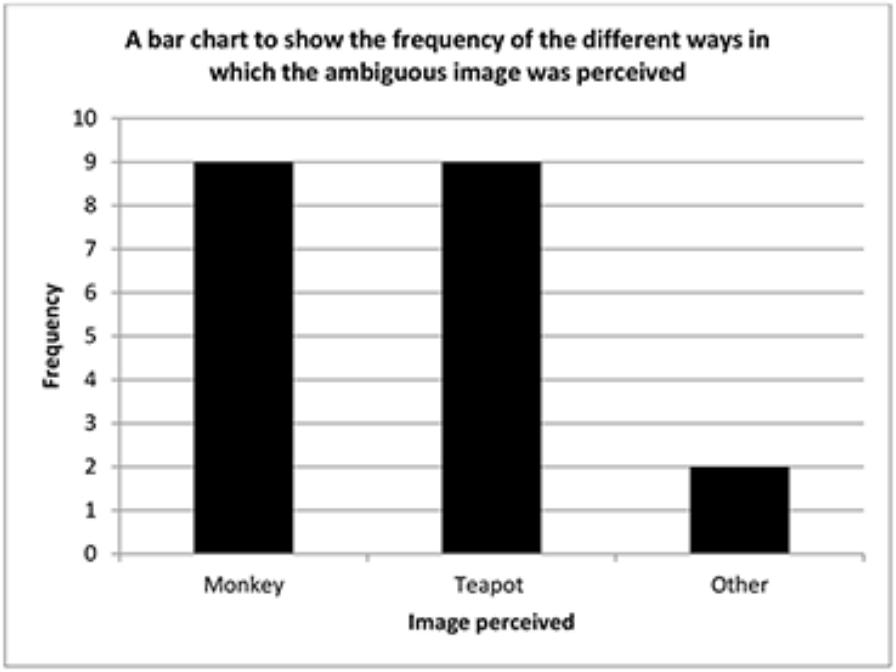
Identify **three** ethical issues that would need to be considered when carrying out this observation.

[3]

118(a).

A psychologist tested the effects of expectations on people's perceptions by carrying out the following experiment. The test item was an ambiguous image – an image that had been purposefully drawn to be perceived in one of two ways – either as a monkey or as a teapot. Participants had to say what they saw after they had viewed the image for one second.

Before carrying out the experiment, the psychologist had checked that the ambiguous image could be perceived in one of two ways. His findings, from this check, are presented in the bar chart below:



Identify **two** findings from the bar chart.

[2]

(b).

Explain why a bar chart is appropriate for presenting this data.

[2]

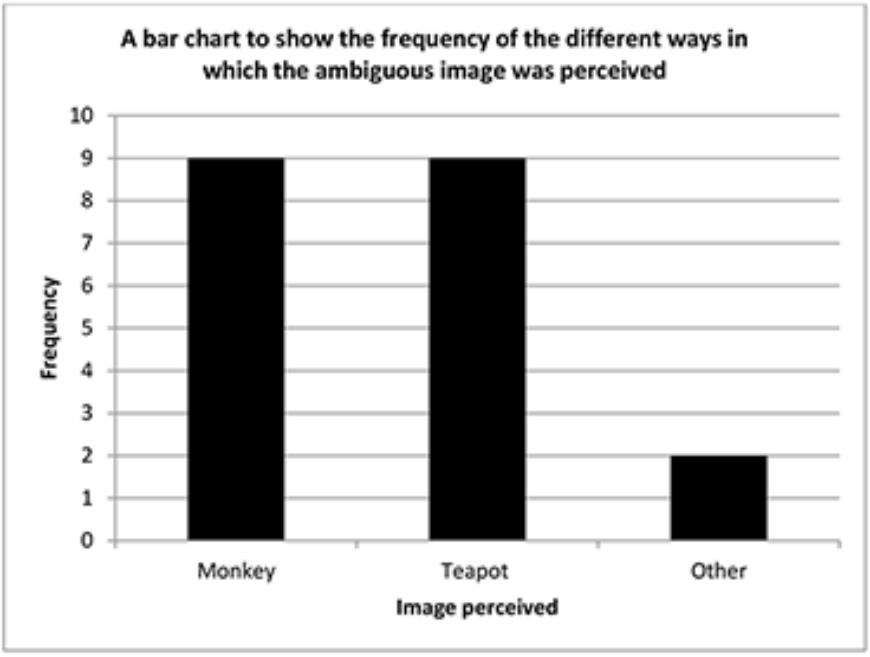
(c). Calculate the percentage number of times that the image was identified as neither a monkey nor a teapot. Show your workings.

---------- **[2]**

119(a).

A psychologist tested the effects of expectations on people's perceptions by carrying out the following experiment. The test item was an ambiguous image – an image that had been purposefully drawn to be perceived in one of two ways – either as a monkey or as a teapot. Participants had to say what they saw after they had viewed the image for one second.

Before carrying out the experiment, the psychologist had checked that the ambiguous image could be perceived in one of two ways. His findings, from this check, are presented in the bar chart below:



Fifty participants were recruited and then randomly allocated into two groups. In one condition, five drawings of other animals were presented, one after the other, before the ambiguous image. Participants had to name each one of these. In the second condition, the set up was the same but five images of kitchen items were used.

The findings from the study are presented below:

A table to show the number of participants who perceived the ambiguous image as a monkey or as a teapot from both conditions: image presented with animals and image presented with kitchen items

	Perceived as monkey	Perceived as teapot
Presented with animals	15	10
Presented with kitchen items	5	12

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.

[2]

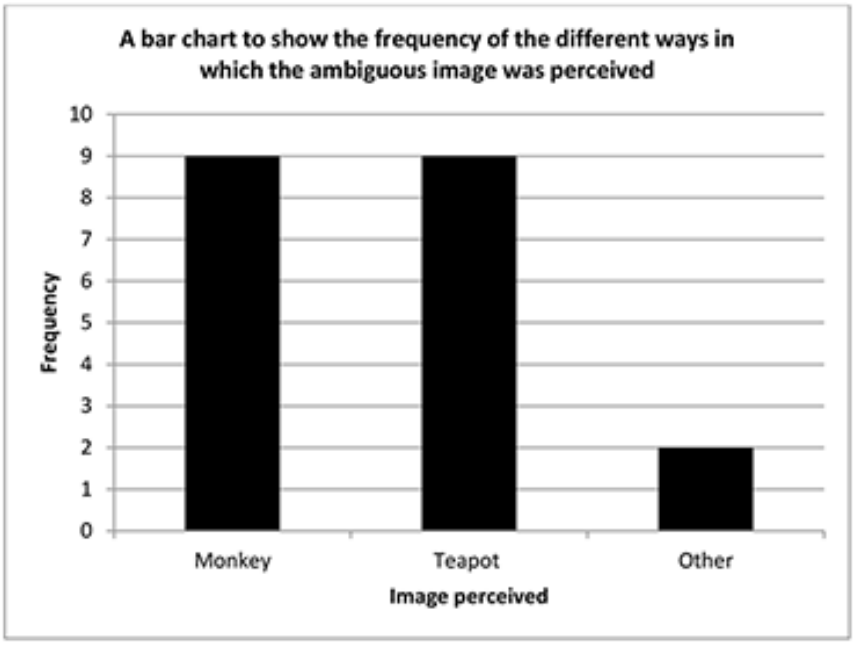
(b). Identify and simplify the ratio of the number of participants who perceived a teapot in the first condition and the number who perceived a teapot in the second condition.

[2]

120.

A psychologist tested the effects of expectations on people's perceptions by carrying out the following experiment. The test item was an ambiguous image – an image that had been purposefully drawn to be perceived in one of two ways – either as a monkey or as a teapot. Participants had to say what they saw after they had viewed the image for one second.

Before carrying out the experiment, the psychologist had checked that the ambiguous image could be perceived in one of two ways. His findings, from this check, are presented in the bar chart below:



Fifty participants were recruited and then randomly allocated into two groups. In one condition, five drawings of other animals were presented, one after the other, before the ambiguous image. Participants had to name each one of these. In the second condition, the set up was the same but five images of kitchen items were used.

The findings from the study are presented below:

A table to show the number of participants who perceived the ambiguous image as a monkey or as a teapot from both conditions: image presented with animals and image presented with kitchen items

	Perceived as monkey	**Perceived as teapot**
Presented with animals	15	10
Presented with kitchen items	5	12

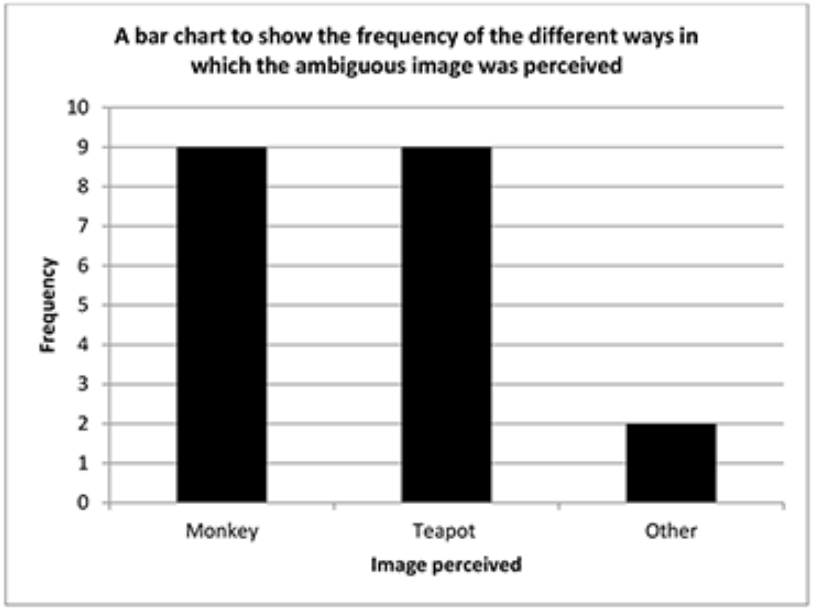
The psychologist analysed the data using the Chi Squared test. Give **two** reasons for this choice of test with reference to the study.

[4]

121.

A psychologist tested the effects of expectations on people's perceptions by carrying out the following experiment. The test item was an ambiguous image – an image that had been purposefully drawn to be perceived in one of two ways – either as a monkey or as a teapot. Participants had to say what they saw after they had viewed the image for one second.

Before carrying out the experiment, the psychologist had checked that the ambiguous image could be perceived in one of two ways. His findings, from this check, are presented in the bar chart below:



Fifty participants were recruited and then randomly allocated into two groups. In one condition, five drawings of other animals were presented, one after the other, before the ambiguous image. Participants had to name each one of these. In the second condition, the set up was the same but five images of kitchen items were used.

The findings from the study are presented below:

A table to show the number of participants who perceived the ambiguous image as a monkey or as a teapot from both conditions: image presented with animals and image presented with kitchen items

	Perceived as monkey	Perceived as teapot
Presented with animals	15	10
Presented with kitchen items	5	12

Explain how the psychologist would determine the appropriate degrees of freedom (df) for this test.

[2]

122. The Chi Squared gave an observed (calculated) value of 3.80

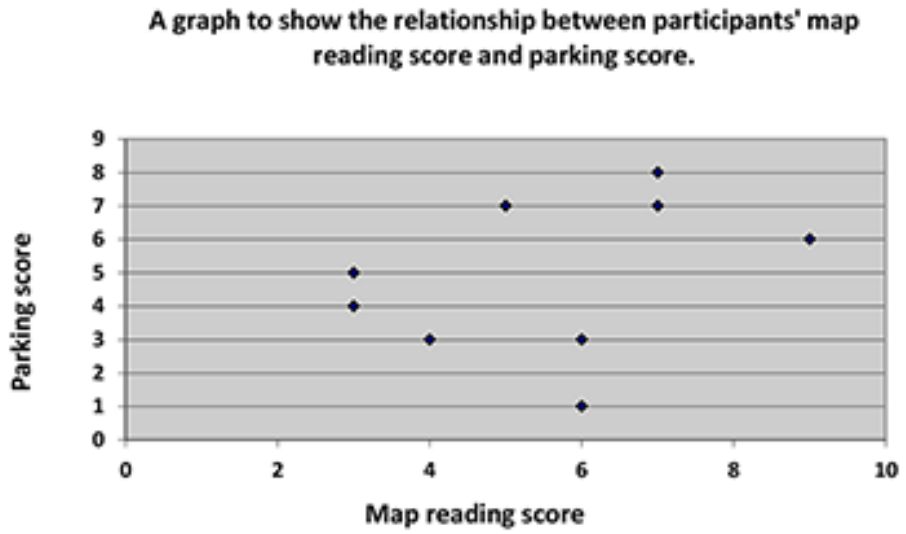
Levels of significance for a one–tailed test

Significance Level	0.05	0.025	0.01
Critical Value	*2.71*	*3.84*	*5.41*

Using the above critical values, explain whether the psychologist has found a significant difference or not.

[4]

123. Look at the following scatter diagram:



Which is the best estimate of the correlation coefficient for the above data?

A 0.7
B 0.4
C 0.3
D 0.8

Your answer ☐ **[1]**

124. What is a weakness of using a mode as a measure of central tendency?

A it can generate a number not in the data set
B it is easily affected by outliers
C it is not suitable for nominal data
D it relies on a score occurring more than once

Your answer ☐ **[1]**

125. Which is a requirement of a parametric test?

A data is at least ordinal level
B mean scores are significantly different
C sample is drawn from a skewed population
D standard deviations are not significantly different

Your answer ☐ **[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?

A to allow for a matched pairs design
B to exclude children who were especially aggressive
C to help to decide on the sex of the role model for each participant
D to measure the change in aggression before and after the experiment

Your answer ☐ **[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?

A 45
B 47.5
C 58
D 76

Your answer ☐ **[1]**

128. What is meant by the term 'socially desirable responses' in psychological research?

A responses which are personal even if they are subjective
B responses which are reliable even if they are invalid
C responses which reflect the participants' wishes even if they are unacceptable to others
D responses which the participants think they ought to give even if they are not true

Your answer ☐ **[1]**

129. Which is an example of qualitative data?

A the diary entries of six patients suffering from schizophrenia
B the length of time each participant spent reading a list of words
C the modal colour chosen by extroverts
D the percentage of respondents who agreed with capital punishment

Your answer ☐ **[1]**

130. What is the probability of a significant result occurring by chance where the significance level is $p \leq 0.025$?

A 2.5% or less

B 25% or less
C at least 97.5%
D less than 2.5%

Your answer ☐ **[1]**

131. Which inferential test should a researcher use to decide whether a correlation is significant?

A Binomial Sign test
B Mann–Whitney U test
C Spearman's Rho test
D Wilcoxon Signed Ranks test

Your answer ☐ **[1]**

132.

Stand to attention. An educational psychologist conducted a study to investigate if getting pupils to stand up when completing some of their work in class increased their performance by making them concentrate more. An independent measures design experiment was used testing a small group of 12 pupils. Half of the class stood up at their desks whilst completing a maths test whilst the others remained seated. The data obtained from the study is presented in the table below.

Score on maths test (max 20) when stood up or sat down when taking the test			
Stood up		Sat down	
participant	score	participant	score
1	18	1	14
2	20	2	8
3	17	3	20
4	15	4	4
5	18	5	15
6	19	6	12

Outline **one** advantage and **one** disadvantage of having quantitative data in this study.

[4]

133(a). The table below displays the data from an experiment investigating the difference in memory ability of a group of young and old people.

Scores in a memory test for words (max 30) by young and old people			
Young people (age 16–25)		Old people (age 65–75)	
participant	score	participant	score
A	26	a	7
B	28	b	25
C	22	c	12
D	30	d	22
E	25	e	12
F	28	f	12

What is the mode for the number of words recalled by the young group?

A 12
B 26
C 27
D 28

Your answer ☐ **[1]**

(b). What would be the value of '*n*' in the formula below to calculate the standard deviation (SD) of the memory scores for the group of old people in this study?

$$SD = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n}}$$

A 6
B 8
C 12
D 15.5

Your answer ☐ **[1]**

(c). What is the mean for the number of words recalled by the old group?

A 12
B 15
C 18
D 26.5

Your answer ☐ **[1]**

(d). What would be the appropriate inferential statistical test to use to analyse the data from this study?

A Chi-square test
B Mann-Whitney U test
C Sign test
D Wilcoxon Signed Ranks test

Your answer ☐ **[1]**

134. What is the symbol for 'much less than'?

A <
B <<
C =
D ==

Your answer ☐ **[1]**

135. What type of reliability is it when there is a consistency in the recordings made by two or more different observers?

A external
B inter-rater
C split half
D test-retest

Your answer ☐ **[1]**

136(a).

What type of distribution is shown in the graph below displaying the results of how people performed in a general knowledge quiz?



A negatively skewed
B normal
C positively skewed
D symmetrical

Your answer ☐ **[1]**

(b). What can be said about people's general knowledge from this graph?

A most people had a high general knowledge score
B most people had a low general knowledge score
C overall, people's general knowledge is about the same
D there was about the same number of people who scored high as scored low

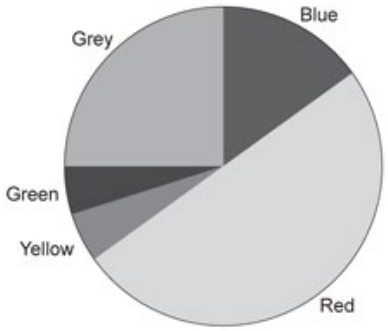
Your answer ☐ **[1]**

137. What type of inferential statistical tests assume that the data used in the analysis are drawn from a normally distributed population?

A parabolic
B parallel
C parametric
D paranormal

Your answer ☐ **[1]**

138. The pie chart below displays the data from a sample of 200 people asked what their favourite colour is. How many people said grey was their favourite colour?



A 25
B 50
C 75
D 90

Your answer ☐ **[1]**

139(a). The table below presents some data from Milgram's study of obedience showing what the maximum shock was that people delivered. All participants continued up to 300 volts, at which stage 5 people dropped out.

Voltage level	Number of participants for whom this was the maximum shock given
300	5
315	4
330	2
345	1
360	1
375	1
390	0
405	0
420	0
435	0
450	26

What percentage of participants gave a final shock of 315 volts?

A 1.6%
B 4%
C 8%
D 10%

Your answer ☐ **[1]**

(b). What was the total percentage of participants who were 'disobedient' (i.e. did not give a shock of 450 v)?

A 5.6%
B 14%
C 35%
D 65%

Your answer ☐ **[1]**

140. Which of these is an ethical consideration when conducting psychological research?

A integration
B integrity
C intercourse
D interrogation

Your answer ☐ **[1]**

141. Which of the following would assess the internal reliability of a questionnaire used by a psychologist in their research?

A comparing responses to questions asking different things
B comparing responses to questions asking the same things
C comparing responses to questions by two different people
D comparing responses to questions from a different test

Your answer ☐ **[1]**

142.

Stand to attention. An educational psychologist conducted a study to investigate if getting pupils to stand up when completing some of their work in class increased their performance by making them concentrate more. An independent measures design experiment was used testing a small group of 12 pupils. Half of the class stood up at their desks whilst completing a maths test whilst the others remained seated. The data obtained from the study is presented in the table below.

Score on maths test (max 20) when stood up or sat down when taking the test			
Stood up		Sat down	
participant	score	participant	score
1	18	1	14
2	20	2	8
3	17	3	20
4	15	4	4
5	18	5	15
6	19	6	12

Identify **two** findings from the data presented in this table.

[4]

143. What is the symbol for 'much greater than'?

A <
B <<
C >
D >>

Your answer **[1]**

144. If a researcher checks to see if participants have answered questions that ask about the same thing in the same or similar way, what is this?

A external reliability
B external validity
C internal reliability
D internal validity

Your answer **[1]**

145. What term refers to whether research relates to everyday life?

A ecological validity
B face validity
C internal validity
D population validity

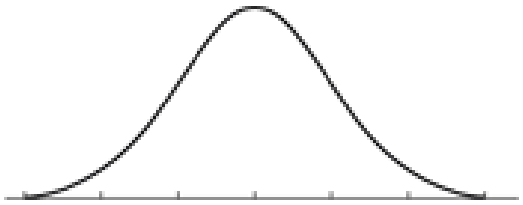
Your answer **[1]**

146. Which of these descriptive statistics would it be possible to use with nominal data?

A mean
B median
C mode
D range

Your answer **[1]**

147. What type of distribution is this?



A bimodal
B negatively skewed
C normal
D positively skewed

Your answer ☐ **[1]**

148. Which of these is not a measure of dispersion?

A mean
B range
C standard deviation
D variance

Your answer ☐ **[1]**

149.

Do like me. The extent to which people mimic each other's behaviour (e.g. fold arms at the same time) is often regarded as an indicator of how well they are getting on together, especially if in a romantic relationship. A psychologist wants to conduct a study using the observation method to investigate this by studying the behaviour of couples having a drink in a bar. Behaviour was observed every 10 minutes during a period from 8 pm to 10 pm.

What does inter-rater reliability refer to in this study?

[3]

150.

Funny research. Laughing can be good for us and make us feel better in many ways but psychologists do not know much about why people laugh or what people find funny or why. To investigate this a psychologist used the self-report method to ask people questions about their experiences of laughing and what they find funny. Some of the data that was collected is presented below.

Do you like it when people play practical jokes on you?	
'yes'	'no'
120	80

Do you laugh with your mouth open?		
	'yes'	'no'
males	63	17
females	38	22

Findings when people were asked which category of jokes do you find the funniest	
Category	**Number of people**
technology	15
health	35
work	10
family	30
news	20
sex	50

What is the ratio of people who like it when practical jokes are played on them to those that do not? Present your answer in its simplest form and show your workings.

[2]

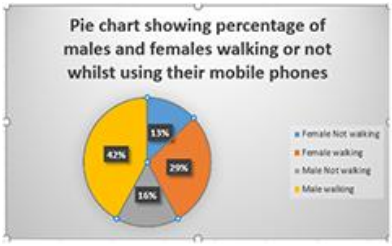
END OF QUESTION PAPER

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,

					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.
	b		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	Max 3 3 2 1 0	-Context = structure/ weight, personality etc **Examiner's Comments** 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.
			Total	**6**	
21			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	Max 3 3 2 1 0	-Context = structure/ weight, personality etc Do not credit weaknesses of correlations that do not apply to quantitative data. **Examiner's Comments** 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

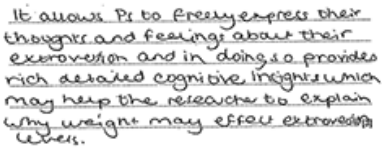
			Brief and/or weak attempt to outline weakness (whether in context or not) The candidate has not provided any creditworthy information		good, contextual examples. A few candidates did not put their responses into the context of the research outlined at the start of Section B or gave very brief responses 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). **<u>Examiner's Comments</u>** 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

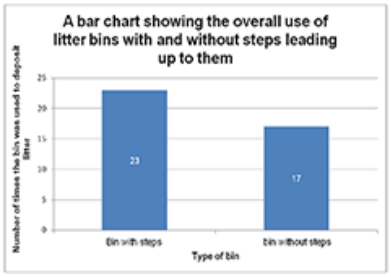
 
It allows Ps to freely express their thoughts and feelings about their extroversion and in doing so provides rich detailed cognitive insights which may help the researcher to explain why weight may effect extroversion levels.

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>

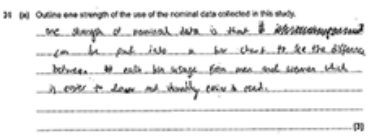
Some candidates did achieve very well for this question and were able to clearly explain what the analysis from the Chi-Square test informed us regarding the use of the two different types of bins in this study. However, many did not achieve full marks due to a lack of clarity. There were several candidates 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**

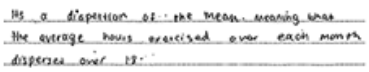
			A Incorrectly accepting the null hypothesis		Answered correctly by most candidates. Some candidates chose option B incorrectly.
			Total	**1**	
36			A when there are a few scores much lower than the rest	1	**Examiner's Comments** A significant number of candidates were unable to give correct response 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 The candidate has not provided any creditworthy information	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.
			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 Clear, detailed response but not in context / **OR** attempt in context Brief and/or weak attempt to outline a conclusion (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 = 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 **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.
			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

					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).
	b		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	**Max 2** **2** **1** **0**	Context = ‘stand’ / ‘standing up’, ‘maths test’, ‘concentration’ etc **Examiner’s Comments** 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.
	c		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	**Max 4**	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

			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) For each calculation of the range (i.e. for males and females) … Range correctly stated with correct workings shown 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	**Max 4** **2** **1** **0**	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 **<u>Examiner's Comments</u>** 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.
	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. Clear outline of conclusion in context Attempt in context Attempt but not in context The candidate has not provided any creditworthy information	**Max 3** **3** **2** **1** **0**	-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) 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: a — O-E: 5 – 8.5 = –3.5 — $(O-E)^2$: $(–3.5)^2$ = 12.25 — $(O-E)^2 \setminus E$: 12.25 / 8.5 = 1.441 Cell: b — O-E: 15 – 11.5 = 3.5 — $(O-E)^2$: $(3.5)^2$ = 12.25 — $(O-E)^2 \setminus E$: 12.25 / 11.5 1.065 Cell: c — O-E: 12 – 8.5 = 3.5 — $(O-E)^2$: $(3.5)^2$ = 12.25 — $(O-E)^2 \setminus E$: 12.25 / 8.5 1.441 Cell: d — O-E: 8 – 11.5 = –3.5 — $(O-E)^2$: $(–3.5)^2$ = 12.25 — $(O-E)^2 \setminus E$: 12.25 / 11.5 1.065 $\sum (O-E)^2 \setminus E$: 1.441 + 1.065 + 1.441 + 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*

<table>
<tr><td></td><td></td><td></td><td>to or exceed the table critical value (3.841) to be significant
<table><tr><td>Correctly written significance statement (calculated value, degrees of freedom and probability level)</td><td>OR written in words rather than a formal statement</td></tr><tr><td>just stating p<0.05</td><td>OR weak and / or brief written response</td></tr></table>The candidate has not provided any creditworthy information</td><td>2

1

0</td><td><u>Examiner’s Comments</u>

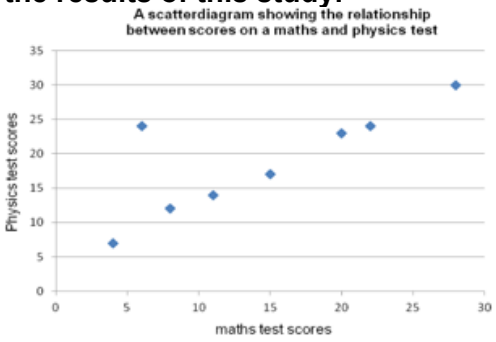
Some candidates in response to this question provided the full, formal statement of statistical significance that would appear in the write-up of a practical report when citing the outcomes of the Chi square test (X^2 = 5.01, df=1, p<0.05). Others interpreted the question as asking for a more general, 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.</td></tr>
<tr><td></td><td></td><td></td><td>Total</td><td>9</td><td></td></tr>
<tr><td>64</td><td>a</td><td></td><td>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.</td><td>2
AO2
c/g</td><td>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.</td></tr>
<tr><td></td><td>b</td><td></td><td>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)’.</td><td>2
AO2
c/g</td><td>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.</td></tr>
</table>

			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

			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.		
			Total	**3**	
78			D	1	
			Total	**1**	
79			B	1	
			Total	**1**	
80			B	1	
			Total	**1**	
81			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	**Max 6** **5-6** **3-4** **1-2** **0**	-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.
			Total	**6**	

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

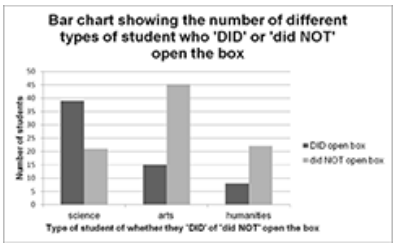
			needs to be calculated. This requires a knowledge of percentages. 37.5% of 40 = 37.5/100 x 40 = 15 12.5% of 40 = 12.5/100 x40 = 5 So the ratio is 15:5, which can be simplified to 3:1	**4**	ratios in response to this questions and show some appropriate workings.
			Correctly stated ratio with all workings 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) **<u>Examiner's Comments</u>** Most candidates were able to select and present two findings from the data table provided.
			Total	**4**	
143			D	1	**<u>Examiner's Comments</u>** There was some confusion here about how to interpret the symbols
			Total	**1**	
144			C	1	**<u>Examiner's Comments</u>** 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	**<u>Examiner's Comments</u>** Mostly correct responses
			Total	**1**	
146			C	1	**<u>Examiner's Comments</u>** Responses to this question indicated some candidates were not familiar with the term 'descriptive statistic'
			Total	**1**	
147			C	1	**<u>Examiner's Comments</u>** 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**	

1(a).

Looks tasty? A psychologist wanted to investigate how expectations influence our perception of food. To do this she conducted an experiment where she presented one group of participants a bag of crisps labelled as 'premium range' in a bright and colourful packet. A different group of participants were presented with the same crisps but labelled as 'budget range' in a plain packet. After eating some of the crisps participants were asked to rate how tasty they were on a scale of 1 ('yuk') to 20 ('yummy').

Ratings of how tasty crisps were (1–20)	
Premium labelled brand	**Budget** labelled brand
14	8
18	3
20	10
8	6
18	18
12	14
10	8
19	4
15	7
16	8
18	8
12	2

Name the appropriate inferential statistical test to analyse the data in this study. Give reasons for your answer.

[4]

(b). Explain how you would find the critical value to compare the calculated value to after conducting this test.

[2]

(c). Outline **one** conclusion that could be made about this study if $p<0.05$ appeared in the significance statement after conducting this test.

[4]

2(a).

Looks tasty? A psychologist wanted to investigate how expectations influence our perception of food. To do this she conducted an experiment where she presented one group of participants a bag of crisps labelled as 'premium range' in a bright and colourful packet. A different group of participants were presented with the same crisps but labelled as 'budget range' in a plain packet. After eating some of the crisps participants were asked to rate how tasty they were on a scale of 1 ('yuk') to 20 ('yummy').

Ratings of how tasty crisps were (1–20)	
Premium labelled brand	**Budget** labelled brand
14	8
18	3
20	10
8	6
18	18

12	14
10	8
19	4
15	7
16	8
18	8
12	2

The range and standard deviation are both measures of dispersion. Outline **one** way that they are different.

[2]

(b). Outline **two** conclusions that can be made about this experiment from the calculation of the range.

[4]

3. Which inferential statistical test requires at least ordinal level data to examine if there is a difference between the data collected from different people in each condition of an experiment?

A Chi-square
B Mann-Whitney U
C Spearman's Rho
D Wilcoxon Signed Ranks

Your answer ☐ **[1]**

4. What is the name given to data that is collected directly by researchers themselves, rather than obtaining it from another person or source?

A direct
B nominal
C primary
D secondary

Your answer ☐ **[1]**

5. Which of the following is **not** related to external validity?

A the extent to which findings can be applied to other places
B the extent to which findings can be obtained at other times
C the extent to which findings can be related to other people not involved in the research
D the extent to which findings can be repeated by testing the same people in the same way

Your answer ☐ **[1]**

6. Below is some data from three people in a study that recorded the reaction time in seconds taken to solve a puzzle when background noise was being played:

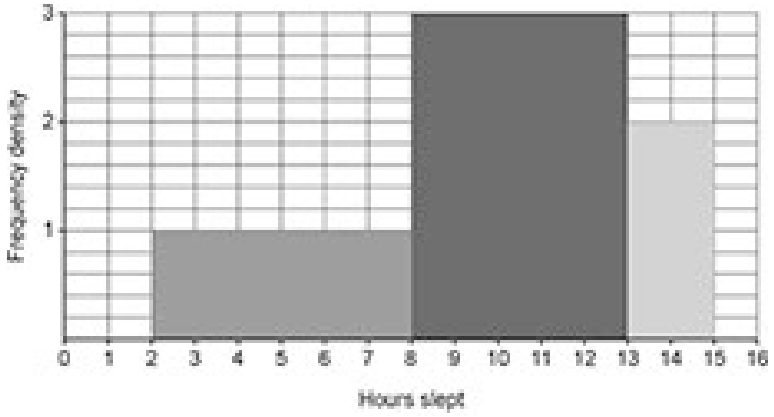
25.89, 16.21, 20.92 ~ 26, 16, 21

What does the symbol (**~**) in between these sets of data mean?

A approximately equal to
B equal to
C proportional to
D related to

Your answer ☐ **[1]**

7(a). A survey was conducted asking a group of students how much they slept each night.



What type of display is used to present the data in the image above?

A bar-chart
B histogram
C line graph
D scatter diagram

Your answer ☐ **[1]**

(b). How many students reported sleeping between 8 and 12 hours?

A 3
B 12
C 15
D 25

Your answer ☐ **[1]**

(c). What was the overall total sample size in this survey?

A 15
B 20
C 25
D 30

Your answer ☐ **[1]**

8.

TV treats. People often eat snack foods, such as crisps, nuts and chocolate whilst watching TV. It may be that the more TV that is watched, the more snack foods are eaten. To investigate this a psychologist wants to use the correlation technique to see if there is a relationship between the amount of TV watched and snack foods eaten.

Name the type of graph that would be used to display the data from a correlation analysis.

_____ **[1]**

9.

TV treats. People often eat snack foods, such as crisps, nuts and chocolate whilst watching TV. It may be that the more TV that is watched, the more snack foods are eaten. To investigate this a psychologist wants to use the correlation technique to see if there is a relationship between the amount of TV watched and snack foods eaten.

Explain how you could reduce the possibility of social desirability in this study.

_____ **[4]**

10.

TV treats. People often eat snack foods, such as crisps, nuts and chocolate whilst watching TV. It may be that the more TV that is watched, the more snack foods are eaten. To investigate this a psychologist wants to use the correlation technique to see if there is a relationship between the amount of TV watched and snack foods eaten.

Explain what the term 'criterion validity' refers to in this study.

[3]

11(a).

Looks tasty? A psychologist wanted to investigate how expectations influence our perception of food. To do this she conducted an experiment where she presented one group of participants a bag of crisps labelled as 'premium range' in a bright and colourful packet. A different group of participants were presented with the same crisps but labelled as 'budget range' in a plain packet. After eating some of the crisps participants were asked to rate how tasty they were on a scale of 1 ('yuk') to 20 ('yummy').

Ratings of how tasty crisps were (1–20)	
Premium labelled brand	**Budget** labelled brand
14	8
18	3
20	10
8	6
18	18
12	14
10	8
19	4
15	7
16	8
18	8
12	2

Explain what quantitative data is.

[2]

(b). Outline **one** advantage of having quantitative data rather than qualitative data in this study.

[3]

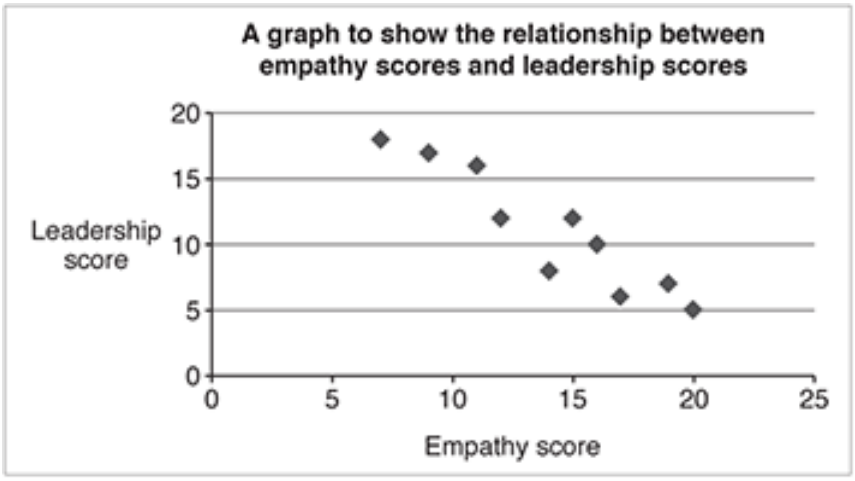
12. A psychologist has designed a new psychometric test to measure extroversion. She gives 30 participants both her test and an established personality test, and compares their scores on the two tests.

What is she trying to establish by doing this?

A concurrent validity
B criterion validity
C ecological validity
D external validity

Your answer ☐ **[1]**

13. Look at the following scatter diagram.



Which is the best estimate of the correlation coefficient for the above data?

A –0.7
B –0.2
C +0.2
D +0.7

Your answer ☐ **[1]**

14. A binomial sign test is correctly used to assess the significance of a set of results.

What level of data and type of design are necessary to carry out this test?

A nominal level data using a repeated measures design
B nominal level data using an independent measures design
C ordinal level data using a repeated measures design
D ordinal level data using an independent measures design

Your answer ☐ **[1]**

15. Which of these is associated with the term ‘demand characteristics’?

A changes in participants’ behaviour as a consequence of knowing they are being studied
B compulsory things that the participants must do
C instructions given to the participants by the researcher
D personality traits of the participants

Your answer ☐ **[1]**

16. Which of these would produce interval level data?

A estimates of the number of sweets in a jar
B ratings of attractiveness of people with pupils of different sizes
C reaction time to respond to visual and auditory stimuli recorded using a stopwatch
D the number of people who hold the door open for the person behind them when entering a shop

Your answer ☐ **[1]**

17.

Furry friends. Some research suggests that animals can help people feel less lonely. To investigate this further a psychologist wants to conduct an experiment to study the effect of caring for pets on elderly people living in one large residential home called 'The Oaklands' which has a total of 125 men and women aged between 60 and 95.

Identify **two** ethical considerations in this study and explain how you would address them.

[6]

18.

A psychologist conducted an observation study to investigate how people react when walking towards each other along a corridor in a building where lots of different office staff work. To do this, she asked for permission to set up hidden cameras for a day to record people's encounters along one of the main corridors connecting one part of the building to another. The results from the study are shown in the table below.

Table 1: Different behaviours observed as two people approach each other in a corridor

Avoids eye contact	Smiles	Says hello	Makes eye contact	Nods	total
80	48	40	24	8	**200**

What level of data is obtained in this study?

---------- **[1]**

19.

A psychologist conducted an observation study to investigate how people react when walking towards each other along a corridor in a building where lots of different office staff work. To do this, she asked for permission to set up hidden cameras for a day to record people's encounters along one of the main corridors connecting one part of the building to another. The results from the study are shown in the table below.

Table 1: Different behaviours observed as two people approach each other in a corridor

Avoids eye contact	Smiles	Says hello	Makes eye contact	Nods	total
80	48	40	24	8	**200**

What is the ratio of the number of times people avoided eye contact to the number of times people said hello?

---------- **[2]**

20.

A psychologist conducted an observation study to investigate how people react when walking towards each other along a corridor in a building where lots of different office staff work. To do this, she asked for permission to set up hidden cameras for a day to record people's encounters along one of the main corridors connecting one part of the building to another. The results from the study are shown in the table below.

Table 1: Different behaviours observed as two people approach each other in a corridor

Avoids eye contact	Smiles	Says hello	Makes eye contact	Nods	total
80	48	40	24	8	**200**

What is the mode for the type of behaviour observed by people as they approach each other in the corridor?

---------- **[1]**

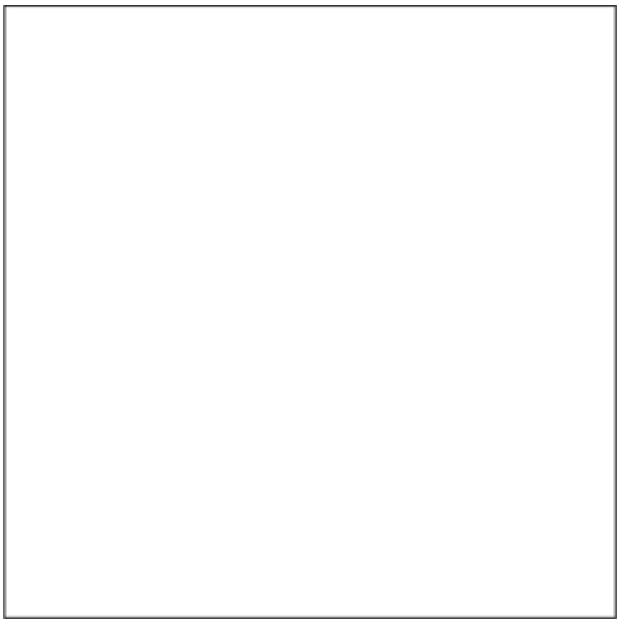
21.

A psychologist conducted an observation study to investigate how people react when walking towards each other along a corridor in a building where lots of different office staff work. To do this, she asked for permission to set up hidden cameras for a day to record people's encounters along one of the main corridors connecting one part of the building to another. The results from the study are shown in the table below.

Table 1: Different behaviours observed as two people approach each other in a corridor

Avoids eye contact	Smiles	Says hello	Makes eye contact	Nods	total
80	48	40	24	8	**200**

Sketch a pie chart to show the data recorded in this study.



[4]

22.

A psychologist conducted an observation study to investigate how people react when walking towards each other along a corridor in a building where lots of different office staff work. To do this, she asked for permission to set up hidden cameras for a day to record people's encounters along one of the main corridors connecting one part of the building to another. The results from the study are shown in the table below.

Table 1: Different behaviours observed as two people approach each other in a corridor

Avoids eye contact	Smiles	Says hello	Makes eye contact	Nods	total
80	48	40	24	8	**200**

Outline **two** conclusions that can be made from the data collected in this study.

[6]

23.

A psychologist conducted an observation study to investigate how people react when walking towards each other along a corridor in a building where lots of different office staff work. To do this, she asked for permission to set up hidden cameras for a day to record people's encounters along one of the main corridors connecting one part of the building to another. The results from the study are shown in the table below.

Table 1: Different behaviours observed as two people approach each other in a corridor

Avoids eye contact	Smiles	Says hello	Makes eye contact	Nods	total
80	48	40	24	8	**200**

In a follow-up study conducted a month later that recorded 310 behaviours in total, the researchers noted that people avoided eye contact 125 times.

Write this as a percentage to two significant figures. Show your working.

[3]

END OF QUESTION PAPER

Mark scheme

Question			Answer/Indicative content	Marks	Guidance
1	a		The appropriate inferential statistical test is the Mann Whitney U test. This is because … (i) It is a test of the difference between two conditions (and the study was investigating the difference in ratings for premium and budget crisps) (ii) It is a test that is used with independent measures designs (and the experiment had different people rating the premium crisps compared to rating the budget crisps), and (iii) It is a test that requires ordinal level data (ratings of the tastiness of crisps on a scale 1 to 20 is ordinal because the outcomes can be ranked) The candidate has not provided any creditworthy information Appropriate test named and justified with more than one clear reason in context Appropriate test named and justified with one clear reason in context Appropriate test named and justified, but not in context Appropriate test named and attempt to justify why (whether in context or not) The candidate has not provided any creditworthy information	Max 4 0 4 3 2 1 0	-Context = crisps, premium and / or budget brand and tasty / tastiness -Context needs to be expressed in relation to justifying choice of test (just saying as a standard lead sentence … 'In this study about taste and crisps …. 'is not acceptable for context here) -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 (e.g. saying nominal rather than ordinal data) **Examiner's Comments** Many candidates struggled with this question and seemed to make a guess from any of the names of inferential statistical tests they could remember. This highlights the importance of making sure knowledge of the criteria for using the five inferential statistical tests in the specification is covered adequately. The best responses correctly identified the test then provided two justifications in context of the research presented.
	b		It would be obtained from a table of critical values using the number of participants in each condition (12) to look up the appropriate figure to use Clear explanation of how to find the critical value Attempt to explain of how to find the critical value The candidate has not provided any creditworthy information	Max 2 2 1 0	-1 mark if just stating something like … 'use table of critical values' -Reference to tables of critical values for the wrong test (e.g. Chi square) = zero -Reference to number of participants alone is not creditworthy **Examiner's Comments**

					There was much confusion here regarded what the term 'critical value' refers to, with some candidates incorrectly assuming it was something obtained directly from the calculation of the inferential statistical test itself (e.g. as in 'expected values' from the Chi square test). Worst still, some discussed aspects of descriptive statistics, such as the mean or standard deviation. This highlights the importance of the need to actually perform some calculations using the inferential tests and gain familiarity with the use of tables of critical values and their role in significance testing.
	c		**2 marks for each conclusion** In this study $p<0.05$ would mean that there is a less than 5% probability that null hypothesis (which states there would be no difference in how premium and budget brand crisps tasted) was true. Therefore, we can conclude that people regard premium brand crisps as being tastier than budget brand crisps. This means that things other than actual taste of crisps can influence our perception of what they are like. Things such as the appearance of the packaging of the crisps and the labels used to describe them. Clear and detailed conclusion outlined in context with correct reference to both the null and alternative hypothesis Clear and detailed conclusion outlined in context with correct reference to *either* the null or alternative hypothesis Clear and detailed conclusion, but not outlined in context **OR** attempt to outline conclusion in context Brief and / or weak attempt to outline conclusion, whether in context or not The candidate has not provided any creditworthy information	Max 4 4 3 2 1 0	-Context = crisps, premium and / or budget brand and tasty / tastiness -For 4 marks must include reference to rejecting the null **and** accepting the alternative hypothesis in context -Reference to alternative and null hypotheses can be implicit – e.g. stating that there is a significant difference between the ratings of the two different brands of crisps (this is creditworthy as H_1) **Examiner's Comments** Candidates who struggled with the previous two related questions also found this question about probability levels and significance testing difficult to comprehend (with some not attempting an answer at all). Once again, the importance of actually carrying out some calculations using the inferential tests is highlighted in order to become aware of what happens after the answer from such tests is obtained that allows hypotheses to be accepted or rejected.

			Total	**10**	
2	a		The range only compares the highest and lowest value, subtracting one from the other, whereas the standard deviation compares each individual score with the mean. Clear outline of one way the range and standard deviation are different Attempt to outline of one way the range and standard deviation are different The candidate has not provided any creditworthy information	Max 2 2 1 0	-For two marks some acknowledgement of the fact that SD takes in to account ALL the data collected is required -If just describing how to calculate one of the measures of dispersion, with no comparison of how this is different to the other (or if the point of comparison is incorrect) = zero marks -Accept as a difference the difficulty of calculating the SD compared to the range **<u>Examiner's Comments</u>** This question required knowledge of both the range and the standard deviation. It was not sufficient just to describe what the range is / how it is calculated on its own. Many candidates did not seem to know what the standard deviation was or the basics of how it is calculated in order to access marks for this question. The best responses here outlined how the standard deviation includes all the data collected in its calculation, compared to using just the two extremes like the range.
	b		Examples could include … -People vary a lot in how they rate the tastiness of premium and budget brand crisps -There is more variation in peoples' ratings of the tastiness of the budget crisps than the premium brand crisps. This means some people seem to think they are very tasty, whereas others do not regard them as tasty at all **2 marks for each conclusion** Clear conclusion outlined in context Clear conclusion outlined but not **OR** attempt to outline conclusion in context in context	Max 4 2 1 0	-Context = crisps, premium and / or budget brand and tasty / tastiness -Cap at 1 mark maximum out of 4 overall if only results / findings presented with no attempt to interpret what they suggest and no conclusion. If there is one finding and one clear conclusion in context = 3 marks, or 2 marks if the conclusion is not in context / attempted in context) -Zero marks if findings presented which are incorrect (e.g. claiming that the range for the budget brand crisps was 16 which means people really liked the taste of them etc) -Range for 'Premium brand' … 20 - 8 = 12 (also accept +1 in calculation, = 13)

			The candidate has not provided any creditworthy information		-Range for 'Budget brand' … 18 - 2 = 16 (also accept +1 in calculation, = 17) **Examiner's Comments** It was important for this question to understand that a conclusion is not simply a result / finding from a piece of research – it is the interpretation of a finding. Furthermore, many candidates here confused the range with a measure of central tendency (e.g. the mean), incorrectly claiming that the higher the range the greater the preference was. The best responses here quoted what the range was first then went on to speculate what this implied in terms of how much agreement there was or not amongst participants in the respective conditions of the experiment. This shows how, although simple and straightforward the range is in terms of how it is calculated, it is important to spend time covering what it actually informs us about the outcomes of a piece of research when used in an applied way.
			Total	**6**	
3			B	1	**Examiner's Comments** There were a mixture of responses here to this question requiring knowledge of selection of inferential statistical test which demonstrates the need to become familiar with the criteria for use of each of the five inferential tests on the specification
			Total	**1**	
4			C	1	**Examiner's Comments** Surprisingly, some candidates struggled with this question, with a mixture of response options being selected
			Total	**1**	

5			D	1	**Examiner's Comments** This question proved to be quite difficult for candidates and highlights the need to cover all the different types of validity on the specification
			Total	**1**	
6			A	1	**Examiner's Comments** Mostly correct answers
			Total	**1**	
7	a		A	1	
	b		C	1	
	c		C	1	**Examiner's Comments** All these related questions concerning the understanding and interpretation of data presented in a histogram proved challenging to many candidates and demonstrates the importance of covering graphical displays in detail and not relying upon a transfer of existing knowledge perhaps from level 2 mathematics courses undertaken by students prior to embarking on the AS Psychology course
			Total	**3**	
8			Scatter diagram (accept 'scattergraph' or 'scattergram' also) AO1: 1 mark for correct naming of scatter diagram (or scattergraph or scattergram) Scatter diagram (or scattergraph or scattergram) correctly named The candidate has not provided any creditworthy information	Max 1 1xAO1 mark 1 0	-Also accept 'scattergraph' and 'scattergram' **Examiner's Comments** Mostly correct answers here (with occasional, incorrect references to 'line graphs')
			Total	**1**	
9			For example: keeping participants naïve; anonymous responses;	Max 4	-Context = reference to TV and / or snacks

			inclusion of other, unrelated questions (distractor / filler questions); providing data / completing study outside of a research context etc Clear explanation of how to reduce social desirability in context Clear explanation of how to reduce social desirability, but not in context **OR** attempt in context Attempt to explain how to reduce social desirability but not in context Brief and / or weak attempt to explain how to reduce social desirability whether in context or not The candidate has not provided any creditworthy information	4 3 2 1 0	-The explanation for reducing social desirability can refer to either variable, or both of them together **Examiner's Comments** Some candidates clearly did not understand what the term 'social desirability' refers to (and some confused it with demand characteristics in general). The best responses were ones characterised by providing a definition of the term first, before a detailed discussion in context of an appropriate strategy to reduce social desirability that provided explicit examples.
			Total	**4**	
10			Criterion validity (or 'predictive validity') assesses how well one measure predicts an outcome for another (related) measure. Here, it refers to how well the measures taken to investigate the relationship between the amount of TV viewed and the number of snacks consumed would compare to different measures of the same thing, such as using weight gain instead of the number of snacks consumed. Clear explanation of what criterion validity refers to in context Clear description of what criterion validity refers to but not in context **OR** attempt in context Brief and / or weak attempt to explain what criterion validity refers to, whether in context or not	Max 3 3 2 1 0	Context = reference to TV and / or snacks Accept reference to 'predictive validity' -Award one mark for discussion of validity in general (and cap at this if no explanation of criterion validity specifically) **Examiner's Comments** This proved to be a very challenging question, with very few candidates achieving all 3 marks on offer. This highlights the importance of covering all the many different types of validity on the specification. Those that did demonstrate an understanding (often using the term 'predictive validity', which was perfectly acceptable) sometimes struggled to provide an example in context of the research proposed that would convey a detailed understanding.

			The candidate has not provided any creditworthy information		
			Total	**3**	
11	a		Quantitative data is information about the quantity of something that is expressed in numbers, rather than words Clear explanation of what quantitative data is Attempt to explain what quantitative data is The candidate has not provided any creditworthy information	Max 2 2 1 0	-1 mark if literally just saying 'numbers' without any attempt to explain what is meant by numbers. -Examples of 2 mark responses could include … 'findings', 'data recorded in numbers', or 'the measurement of a variable or aspect of persons behaviour' -'Numbers that are easy to analyse and compare' = 2 marks **Examiner's Comments** Most candidates were able to explain what quantitative data refers to (the clearest responses were those that included an example)
	b		Advantages include … -Able to perform more descriptive statistics (e.g. calculate the mean of the tastiness of each brand of crisp) -More objective -Easier to analyse and present findings -Easier to compare results across conditions Clear and detailed outline of advantage in context Clear outline of advantage, but not in context **OR** attempt in context Brief and / or weak attempt to outline advantage (whether in context or not) The candidate has not provided any creditworthy information	Max 3 3 2 1 0	-Context = crisps, premium and / or budget brand and tasty / tastiness -Accept any reference to study details (e.g. participant numbers) as context -For 3 marks must be some comparison with qualitative data in discussing strength -Cap at 1 mark if there is no reference to qualitative data at all in answer (whether in context or not). However, candidates may refer to 'data being in words', rather than using the 'qualitative', and this IS acceptable. **Examiner's Comments** Most candidates were able to outline advantage of quantitative data. However, this was not always done in the context of the research presented and not always in comparison with qualitative data (which the question asked for). The best responses here began with an outline in general of an advantage of quantitative data, then went on to provide an example of this in the context of the study and making a contrast to why this was an advantage compared to qualitative data.

<table>
<tr><td></td><td></td><td></td><td>Total</td><td>5</td><td></td></tr>
<tr><td>12</td><td></td><td></td><td>A</td><td>1</td><td><u>Examiner's Comments</u>
This proved to be a challenging question and shows the importance of teaching terms and concepts well</td></tr>
<tr><td></td><td></td><td></td><td>Total</td><td>1</td><td></td></tr>
<tr><td>13</td><td></td><td></td><td>A</td><td>1</td><td><u>Examiner's Comments</u>
Most candidates were able to identify this was a negative correlation but some struggled with the strength of the correlation</td></tr>
<tr><td></td><td></td><td></td><td>Total</td><td>1</td><td></td></tr>
<tr><td>14</td><td></td><td></td><td>A</td><td>1</td><td><u>Examiner's Comments</u>
Generally quite poorly answered and shows the need to prepare candidates to answer questions related to choice of inferential statistical tests</td></tr>
<tr><td></td><td></td><td></td><td>Total</td><td>1</td><td></td></tr>
<tr><td>15</td><td></td><td></td><td>A</td><td>1</td><td><u>Examiner's Comments</u>
Mostly correct</td></tr>
<tr><td></td><td></td><td></td><td>Total</td><td>1</td><td></td></tr>
<tr><td>16</td><td></td><td></td><td>C</td><td>1</td><td><u>Examiner's Comments</u>
Generally well answered</td></tr>
<tr><td></td><td></td><td></td><td>Total</td><td>1</td><td></td></tr>
<tr><td>17</td><td></td><td></td><td>Ethical issues could include: causing elderly residents distress by discussing loneliness with them; issues of safety around animals; interaction with people they are not familiar with (distress); may not fully understand what to research is about (consent); may become attached to animals and not want it taken away etc
3 marks for each ethical issue
Ethical issue clearly identified and way to address it clearly explained in context
<table><tr><td>Ethical issue clearly identified and way to address it clearly</td><td>OR Ethical issue clearly identified and way to address it clearly</td></tr></table></td><td>Max 6

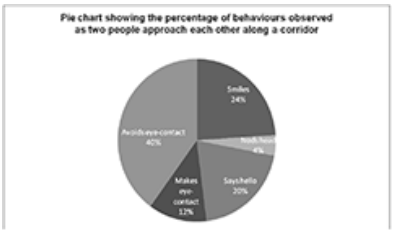
3

2

1</td><td>-Context = pets, loneliness, features of the residents of the home (e.g. elderly)

<u>Examiner's Comments</u>
This question required two appropriate ethical considerations to be identified and addressed. The best responses here were characterised by candidates who presented two separate ethical issues in turn, first by identifying what the ethical issue was and why it was an issue and then going on to explain how it could be addressed. This also needed to be done in context of the research outlined to achieve the highest band marks.</td></tr>
</table>

			explained in context / explained, but not in context Attempt to identify ethical issue and attempt to explain how to address it (whether in context or not) / **OR** clearly identified ethical issue with no suggestion of how to address it The candidate has not provided any creditworthy information	**0**	
			Total	**6**	
18			The level of data obtained in the study is nominal. The candidate has not provided any creditworthy information	**Max 1** **0**	**Examiner's Comments** Although the majority of candidates answered this question correctly some responded with incorrect alternatives revealing that they did not have a good understanding of different levels of data (some incorrectly stated 'ordinal' and some 'interval'). Some also confused levels of data with 'types of data' (primary or secondary).
			Total	**1**	
19			The ratio is 2:1 Correctly stated ratio, simplified to 2:1 Just saying 80:40 without simplifying the answer The candidate has not provided any creditworthy information	**Max 1** **2** **1** **0**	-80:40 is creditworthy for 1 mark **Examiner's Comments** Most candidates demonstrated a good understanding of the concept of ratio in response to this question and went on to present their answer in its simplest form.
			Total	**2**	
20			The mode 'avoids eye-contact' (with 80 occurrences) Correctly stating the mode as 'avoids eye-contact' The candidate has not provided any creditworthy information	**Max 1** **1** **0**	Accept figure '80' as correct answer **Examiner's Comments** Most candidates demonstrated understanding of what the mode is as applied to the data collected from this study.
			Total	**1**	

21			 1 mark is awarded for correctly calculating what proportion of the circle should represent each of the five behavioural categories. 1 mark is awarded for drawing the sectors in proportional size to the data displayed 1 mark is awarded for clear labelling of each sector of the pie chart 1 mark is awarded for a clear and appropriate title 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**	-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 labeling of the pie chart, or separately at the side) Avoids eye contact: 40% = 144 degrees Smiles: 24% = 86.4 degrees Nods head: 4% = 14.4 degrees Says hello: 20% = 72 degrees Makes eye-contact: 12% = 43.2 degrees **Examiner's Comments** In order to achieve full marks on this question candidates needed to sketch a pie chart with sectors drawn in proportion to the amount of responses recorded for each of the five categories of behaviours studied. In order to do this accurately candidates needed to calculate what percentage, or proportion of the circle (in degrees) needed to be used to represent each of the behavioural categories. The answers to these calculations needed to be evident within the labelling of the pie chart or presented separately at the side. Candidates also needed to provide clear labels and an overall title. It was clear that some candidates struggled with this question by the fact that there were several different attempts to draw the sectors of the pie, with lots of crossing and rubbing out of original lines leading to a scruffy response at times. It is worth practicing drawing pie charts using a compass and protractor to get students used to this type of question.
			Total	**4**	
22			Conclusions could include: the most frequent behaviour was to avoid eye-contact, which suggests people do	**6**	-Context = corridor, workplace, office worker etc

<table>
<tr>
<td></td><td></td><td></td>
<td>not like acknowledging each other in a corridor; The lowest frequency of behaviour was ‘nods head’, which suggests people do not like greeting people as they meet in a corridor.

Accept any other appropriate conclusions here.

3 marks for each conclusion
Clear, detailed response in context
<table>
<tr><td>Clear, detailed response but not in context</td><td>OR attempt in context</td></tr>
<tr><td>Attempt, whether in context or not</td><td>OR simply stating a finding</td></tr>
</table>
The candidate has not provided any creditworthy information</td>
<td>3
2

1

0</td>
<td>-Simply referring to the behaviour categories (e.g. ‘smiling’) is not sufficient for context here

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

-For 2 marks could be saying … people avoid each other when approaching in the corridor etc
Do not accept reference to number of people engaging in a certain behavior (as opposed to number of behaviours)

<u>Examiner’s Comments</u>
This question was poorly answered on the whole. Many candidates simply referred to results / findings rather than conclusions (interpretation of findings). Another common misconception was that the total number of recorded behaviours for each category equated to the total number of individual people displaying these behaviours, rather than just the total number of behaviours for each category. If nothing else, this should have been clear from the title of the table, which referred to ‘behaviours’ and not ‘number of people’ exhibiting the behaviours. Candidates who did respond appropriately were able to generate a variety of different conclusions, with the ones achieving the highest marks contextualising their answers (e.g. saying the most common behaviour of avoiding eye contact suggests either people were shy or simply very busy and preoccupied with their work to acknowledge other people as they passed them in the corridor).</td>
</tr>
<tr>
<td></td><td></td><td></td><td>Total</td><td>6</td><td></td>
</tr>
<tr>
<td>23</td><td></td><td></td><td></td><td>Max 3</td><td>-Expression of the percentage in the initial stage of the calculation can be</td>
</tr>
</table>

<table>
<tr>
<td></td><td></td><td></td>
<td>
125 written as a percentage of 310 is
125/310 x 100 = 40.3225806
Expressed to two significant figures this is 40
<table>
<tr><td>AO2 marks</td><td>1 mark is awarded for application of knowledge and understanding of how to convert raw data in to a percentage and a further mark for the application of knowledge and understanding of what significant figures are applied to the data in this study</td></tr>
</table>
Percentage correctly calculated and written to two significant figures
Percentage correctly identified but not written to two significant figures
<table>
<tr><td>Just showing workings (calculations) i.e. 125/310 (x 100)</td><td>OR just stating correct answer without any workings</td></tr>
</table>
The candidate has not provided any creditworthy information
</td>
<td>
3xAO2

3

2

1

0
</td>
<td>
written to any number of decimal places

<u>Examiner’s Comments</u>
There was some confusion in this question with decimal places and significant figures. Most candidates successfully performed the calculation to express 125 as a percentage of 310, but then presented the answer to two decimal places rather than two significant figures. It is worth practicing different techniques for producing estimates of data to demonstrate the difference between decimal places and significant figures here.
</td>
</tr>
<tr>
<td></td><td></td><td></td><td>Total</td><td>3</td><td></td>
</tr>
</table>