**1**

A psychologist was interested in testing a new treatment for people with eating disorders. She put up adverts in several London clinics to recruit participants. Thirty people came forward and they were all given a structured interview by a trained therapist. The therapist then calculated a numerical score for each participant as a measure of their current functioning, where 50 indicates excellent, healthy functioning and zero indicates failure to function adequately. The psychologist then randomly allocated half the participants to a treatment group and half to a no-treatment group. After eight weeks, each participant was re-assessed using a structured interview conducted by the same trained therapist, and given a new numerical score. The trained therapist did not know which participants had been in either group.

For each participant, the psychologist calculated an improvement score by subtracting the score at the start of the study from the score after eight weeks. The greater the number, the better the improvement.

**Median and range of improvement scores for the treatment group and for the no-treatment group**

| | **Treatment group** | **No–treatment group** |
|---|---|---|
| Median | 10.9 | 2.7 |
| Range | 2.1 | 0.8 |

(a) With reference to the data in the table above, outline what the findings of this investigation seem to show about the effectiveness of the treatment.

**(2)**

(b) The psychologist used a statistical test to find out whether there was a significant difference in improvement between the 'treatment' and 'no-treatment' groups. She found a significant difference at the 5% level for a one-tailed test ( $p \leq 0.05$).

Identify an appropriate statistical test for analysing the participants' scores.
Explain why it would be a suitable test to use in this study.

**(4)**

(c) What is the likelihood of the psychologist having made a Type 1 error in this study? Explain your answer.

**(2)**

(d) The psychologist assumed that improvements in the treatment group were a direct result of the new type of treatment. Suggest **two** other reasons why people in the treatment group might have improved.

**(4)**

(e) The psychologist could have used self-report questionnaires to assess the participants instead of using interviews with the therapist. Explain **one** advantage and **one** disadvantage of using self-report questionnaires in this study rather than interviews.

**(4)**

(f) The psychologist needed to obtain informed consent from her participants. Write a brief consent form which would be suitable for this study. You should include some details of what participants could expect to happen in the study and how they would be protected.

**(5)**

(g) What is meant by reliability? Explain how the reliability of the scores in this study could be checked.

**(4)**

(h) The psychologist noticed that female and male participants seemed to have responded rather differently to the treatment.

She decided to test the following hypothesis:

> Female patients with an eating disorder will show greater improvement in their symptoms after treatment with the new therapy than male patients.

She used a new set of participants and, this time, used self-report questionnaires instead of interviews with a therapist.

Imagine that you are the psychologist and are writing up the report of the study. Write an appropriate methods section which includes reasonable detail of design, participants, materials and procedure. Make sure that there is enough detail to allow another researcher to carry out this study in the future.

**(10)**

**(Total 35 marks)**

**2** Read the text below and answer the questions that follow.

A psychologist is using the observational method to look at verbal aggression in a group of children with behavioural difficulties. Pairs of observers watch a single child in the class for a period of one hour and note the number of verbally aggressive acts within ten-minute time intervals. After seeing the first set of ratings, the psychologist becomes concerned about the quality of inter-rater reliability. The tally chart for the two observers is shown in the table below.

**Table: Observation of one child – number of verbally aggressive acts in ten-minute time intervals**

| Time slots | 0–10 | 11–20 | 21–30 | 31–40 | 41–50 | 51–60 |
|---|---|---|---|---|---|---|
| **Observer A** | 2 | 5 | 0 | 6 | 4 | 3 |
| **Observer B** | 4 | 3 | 2 | 1 | 6 | 5 |

(a) Use the data in the Table above to sketch a scattergram. Label the axes and give the scattergram a title.

**(4)**

(b) Using the data in the Table above, explain why the psychologist is concerned about inter-rater reliability.

**(4)**

(c) Identify an appropriate statistical test to check the inter-rater reliability of these two observers. Explain why this is an appropriate test.

**(3)**

(d) If the psychologist does find low reliability, what could she do to improve inter-rater reliability before proceeding with the observational research?

**(4)**

**(Total 15 marks)**

**3** Read the item and then answer the questions that follow.

The results of the study are given in **the table** below.

**Mean number of verbal errors and standard deviations for both conditions**

| | **Condition A (believed audience of 5 listeners)** | **Condition B (believed audience of 100 listeners)** |
|---|---|---|
| **Mean** | 11.1 | 17.2 |
| **Standard deviation** | 1.30 | 3.54 |

(a) What conclusions might the psychologist draw from the data in **the table**? Refer to the means **and** standard deviations in your answer.

**(6)**

(b) Read the item and then answer the question that follows.

> The psychologist had initially intended to use the range as a measure of dispersion in this study but found that one person in **Condition A** had made an exceptionally low number of verbal errors.

Explain how using the standard deviation rather than the range in this situation, would improve the study.

**(3)**

(c) Name an appropriate statistical test that could be used to analyse the number of verbal errors in **the table** above. Explain why the test you have chosen would be a suitable test in this case.

**(4)**

(d) The psychologist found the results were significant at $p<0.05$. What is meant by 'the results were significant at $p<0.05$'?

_____

_____

_____

_____

**(2)**

(e) Briefly explain **one** method the psychologist could use to check the validity of the data she collected in this study.

_____

_____

_____

_____

**(2)**

**(Total 17 marks)**

**4**

(a) The psychologist was also interested in the effects of a restricted diet on memory functioning and he expected memory to become impaired. The psychologist's hypothesis was that participants' scores on a memory test are lower after a restricted diet than before a restricted diet. He gave the volunteers a memory test when they first arrived in the research unit and a similar test at the end of the four-week period. He recorded the memory scores on both tests and analysed them using the Wilcoxon signed ranks test. He set his significance level at 5%.

His calculated value was $T = 53$.

State whether the hypothesis for this study is directional or non–directional.

**(1)**

(b) **Table: Extract from table of critical values from the Wilcoxon signed ranks test**

| Level of significance for a one-tailed test | 0.05 | 0.025 |
|---|---|---|
| Level of significance for a two-tailed test | 0.1 | 0.05 |
| *N* | *T*≤ | |
| 19 | 53 | 46 |
| 20 | 60 | 52 |
| 21 | 67 | 58 |
| 22 | 75 | 65 |

Calculated *T* must be equal to or less than the critical value (table value) for significance at the level shown

Using the table above, state whether or not the psychologist's result was significant. Explain your answer.

**(3)**

**(Total 4 marks)**

**5**

A teacher has worked in the same primary school for two years. While chatting to the children, she is concerned to find that the majority of them come to school without having eaten a healthy breakfast. In her opinion, children who eat 'a decent breakfast' learn to read more quickly and are better behaved than children who do not. She now wants to set up a pre-school breakfast club for the children so that they can all have this beneficial start to the day. The local authority is not willing to spend money on this project purely on the basis of the teacher's opinion and insists on having scientific evidence for the claimed benefits of eating a healthy breakfast.

(a) Explain why the teacher's personal opinion cannot be accepted as scientific evidence. Refer to some of the major features of science in your answer.

A psychologist at the local university agrees to carry out a study to investigate the claim that eating a healthy breakfast improves reading skills. He has access to 400 five-year-old children from 10 local schools, and decides to use 100 children (50 in the experimental group and 50 in the control group). Since the children are so young, he needs to obtain parental consent for them to take part in his study.

**(6)**

(b) The psychologist used a random sampling method. Explain how he could have obtained his sample using this method.

**(3)**

(c) Explain limitations of using random sampling in this study.

**(3)**

(d) Explain why it is important to operationalise the independent variable and the dependent variable in this study and suggest how the psychologist might do this.

**(5)**

(e) The psychologist used a Mann-Whitney test to analyse the data. Give **two** reasons why he chose this test.

**(2)**

(f) He could have used a matched pairs design. Explain why this design would have been more difficult to use in this study.

**(2)**

(g) Other than parental consent, identify **one** ethical issue raised in this study and explain how the psychologist might address it.

**(2)**

(h) The psychologist asks some of his students to conduct a separate observational study at the same time on the same group of children. The aim of this observational study is to test the idea that eating a healthy breakfast affects playground behaviour.

Design an observational study to investigate the effects of a healthy breakfast on playground behaviour. Include in your answer sufficient detail to allow for reasonable replication of the study. You should state the hypothesis you are setting out to test.

In your answer, refer to:

- an appropriate method of investigation
- materials/apparatus and procedure.

Justify your design decisions.

**(12)**

**(Total 35 marks)**

**6** Psychological research suggests an association between birth order and certain abilities. For example, first-born children are often logical in their thinking whereas later-born children tend to be more creative. A psychologist wonders whether this might mean that birth order is associated with different career choices. She decides to investigate and asks 50 artists and 65 lawyers whether they were the first-born child in the family or not.

(a) Write a non-directional hypothesis for this study.

**(2)**

(b) Identify an appropriate sampling method for this study and explain how the psychologist might have obtained such a sample

The psychologist found the following results:

- 20 of the 50 artists were first-born children
- 35 of the 65 lawyers were first-born children.

She analysed her data using a statistical test and calculated a value of $\chi^2$ = 2.27. She then looked at the relevant table to see whether this value was statistically significant. An extract from the table is provided below.

**Table: Critical values of $\chi^2$**

| Level of significance for a one-tailed test | | | | |
|---|---|---|---|---|
| | 0.10 | 0.05 | 0.025 | 0.01 |
| **Level of significance for a two-tailed test** | | | | |
| | 0.20 | 0.10 | 0.05 | 0.02 |
| ***df*** | | | | |
| 1 | 1.64 | 2.71 | 3.84 | 5.41 |

Calculated value of $\chi^2$ must be equal to or exceed the table (critical) values for significance at the level shown

**(3)**

(c) Imagine that you are writing the results section of the report on this investigation. Using information from the description of the study above and the relevant information from the statistical table, provide contents suitable for the results section.

You must provide all of the following:

- an appropriately labelled $2 \times 2$ contingency table a sketch of an appropriately labelled bar chart
- identification of the appropriate statistical test with justification for its use
- identification of an appropriate significance level
- a statement of the results of the statistical test in relation to the hypothesis.

**(12)**

**(Total 17 marks)**

**7** A researcher wanted to see whether cognitive behaviour therapy was an effective treatment for depression. Twenty depressed patients who had all recently completed a course of cognitive behaviour therapy were involved in the investigation. From their employment records, the researcher kept a record of the number of absences from work each patient had in the year following their treatment. This was compared with the number of absences from work each patient had in the year prior to their treatment.

Those patients who had fewer absences from work in the year following their treatment than in the year prior to their treatment were classified as ‘improved’ (+). Those patients who had more absences were classified as ‘deteriorated’ (-). Those patients who had the same number of absences were classified as ‘neither’ (0).

The results of the investigation are included in **Table 1** below.

**Table 1**

| Patient | Improved | Deteriorated | Neither |
|---|---|---|---|
| **1** | + | | |
| **2** | | | 0 |
| **3** | | – | |
| **4** | + | | |
| **5** | + | | |
| **6** | + | | |
| **7** | | – | |
| **8** | | – | |
| **9** | | | 0 |
| **10** | + | | |
| **11** | | – | |
| **12** | + | | |
| **13** | + | | |
| **14** | + | | |
| **15** | + | | |
| **16** | | – | |
| **17** | + | | |
| **18** | + | | |
| **19** | + | | |
| **20** | | | 0 |

The researcher decided to use the sign test to analyse the data.

(a) Explain **two** factors that the researcher had to take into account when deciding to use the sign test. Refer to the investigation above in your answer.

**(4)**

(b) Calculate the sign test value of s for the data in **Table 1**. Explain how you reached your answer.

**(2)**

**Table 2: Critical values for the sign test**

| n | 0.005 (one tailed) 0.01 (two tailed) | 0.01 (one tailed) 0.02 (two tailed) | 0.025 (one tailed) 0.05 (two tailed) | 0.05 (one tailed) 0.10 (two tailed) |
|---|---|---|---|---|
| **16** | 2 | 2 | 3 | 4 |
| **17** | 2 | 3 | 4 | 4 |
| **18** | 3 | 3 | 4 | 5 |

**For significance, the value of the less frequent sign is equal to, or less than, the value of the table.**

(c) With reference to the critical values in **Table 2**, explain whether or not the value of s that you calculated in response to **question (b)** is significant at the 0.05 level for a two tailed test.

**(2)**

(d) The investigation above is based on secondary data.

In what ways would the use of primary data have improved this investigation?

**(3)**

(e) Outline the implications of psychological research for the economy. Refer to the investigation above in your answer.

**(5)**

**(Total 16 marks)**

**8** In an observational study, 100 cars were fitted with video cameras to record the driver's behaviour. Two psychologists used content analysis to analyse the data from the films. They found that 75% of accidents involved a lack of attention by the driver. The most common distractions were using a hands-free phone or talking to a passenger. Other distractions included looking at the scenery, smoking, eating, personal grooming and trying to reach something within the car.

(a) What is content analysis?

**(2)**

(b) Explain how the psychologists might have carried out content analysis to analyse the film clips of driver behaviour.

**(4)**

(c) Explain how the two psychologists might have assessed the reliability of their content analysis.

The psychologists then designed an experiment to test the effects of using a hands-free phone on drivers' attention. They recruited a sample of 30 experienced police drivers and asked them to take part in two computer-simulated driving tests. Both tests involved watching a three-minute film of a road. Participants were instructed to click the mouse as quickly as possible, when a potential hazard (such as a car pulling out ahead) was spotted.

Each participant completed two computer-simulated driving tests:

- Test A, whilst chatting with one of the psychologists on a hands-free phone
- Test B, in silence, with no distractions.

The order in which they completed the computer tests was counterbalanced.

**(3)**

(d) Explain why the psychologists chose to use a repeated measures design in this experiment.

**(3)**

(e) Identify **one** possible extraneous variable in this experiment. Explain how this variable may have influenced the results of this experiment.

**(3)**

(f) Explain **one or more** ethical issues that the psychologists should have considered in this experiment.

**(4)**

(g) Write a set of standardised instructions that would be suitable to read out to participants, before they carry out Test A, chatting on a hands-free phone.

The computer simulator measured two aspects of driver behaviour:

- the number of hazards detected by each driver
- the time taken to respond to each hazard, in seconds.

The mean scores for each of these measures is shown in the table below.

**Table to show the mean number of hazards detected and mean reaction times in seconds for Test A and Test B**

| **Mean scores** | **Test A: with hands-free phone** | **Test B: in silence** |
|---|---|---|
| **Number of hazards detected** | 26.0 | 23.0 |
| **Reaction time in seconds** | 0.45 | 0.27 |

The psychologists then used an inferential statistical test to assess whether there was a difference in the two conditions.

**(5)**

(h) Identify an appropriate statistical test to analyse the difference in the number of hazards detected in the two conditions of this experiment. Explain why this test of difference would be appropriate.

They found no significant difference in the number of hazards detected ($p > 0.05$), but there was a significant difference in reaction times ($p . 0.01$).

**(3)**

(i) Explain why the psychologists did not think that they had made a Type 1 error in relation to the difference in reaction times.

**(2)**

(j) Replication is one feature of the scientific method. The psychologists decided to replicate this experiment using a larger sample of 250 inexperienced drivers.

Explain why replication of this study would be useful.

**(3)**

**(Total 32 marks)**

**9**

The psychologists then wanted to see whether the use of diagrams in medical consultations would affect recall of medical information.

In a laboratory experiment involving a medical consultation role-play, participants were randomly allocated to one of two conditions. In Condition A, a doctor used diagrams to present to each participant a series of facts about high blood pressure. In Condition B, the same doctor presented the same series of facts about high blood pressure to each participant but without the use of diagrams.

At the end of the consultation, participants were tested on their recall of facts about high blood pressure. Each participant was given a score out of ten for the number of facts recalled.

(a) In this case, the psychologists decided to use a laboratory experiment rather than a field experiment. Discuss advantages of carrying out this experiment in a laboratory.

**(4)**

(b) Identify an appropriate statistical test that the psychologists could use to analyse the data from the follow-up study. Give **one** reason why this test is appropriate.

**(2)**

**(Total 6 marks)**

**10**

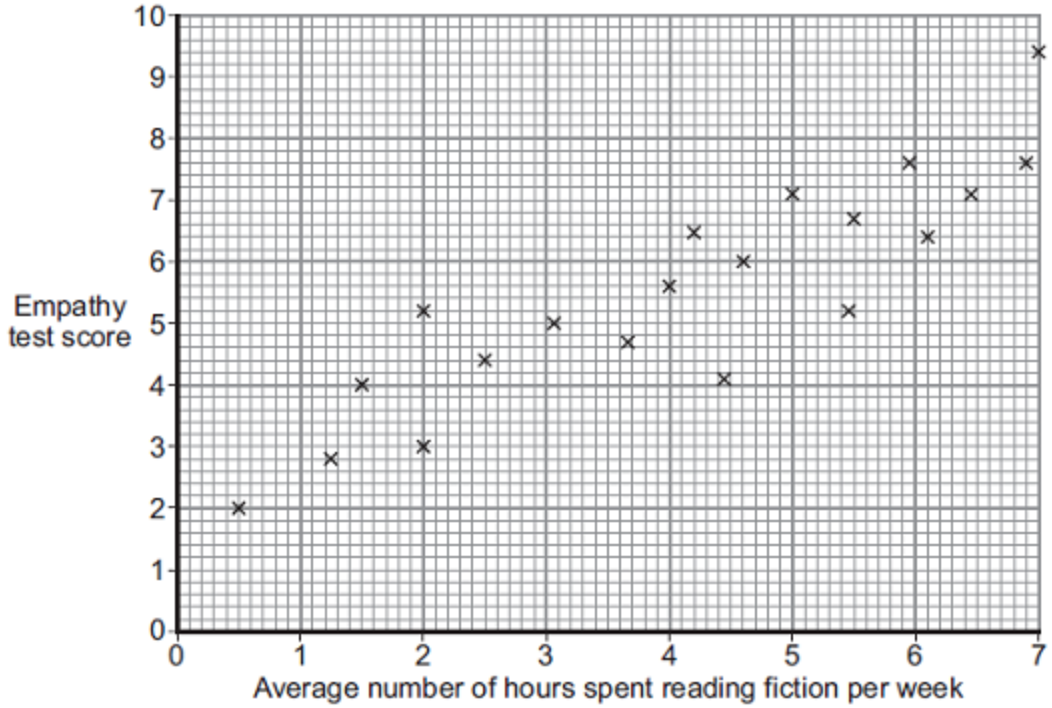
A student teacher was interested in the relationship between empathy (consideration and feelings for others) and the time spent reading fiction. She decided to investigate whether or not such a relationship was present in children.

The student teacher designed her own questionnaire to measure empathy in 8-year-old children. The higher the score achieved, the greater the empathy. Twenty children, all from one school, took part. Each child completed the questionnaire individually.

The student teacher designed another questionnaire to measure ‘time spent reading fiction’. Each child was given this questionnaire to take home and complete with his or her parents over a four-week period. ‘Time spent reading fiction’ included the time spent by parents reading to the child as well as the time the child spent reading independently. Using the responses to this questionnaire, the student teacher calculated how much time per week, on average, each child spent reading fiction.

The data obtained are shown in the graph below.

**Scattergram of children’s scores on a test of empathy and the average number of hours spent reading fiction per week.**




(a) Outline the relationship between empathy and the average number of hours spent reading fiction per week shown in the graph above.

**(1)**

(b) Name an appropriate test to determine whether or not there is a significant relationship between the two variables in the graph above. Justify your answer with reference to levels of measurement.

**(2)**

The student teacher decided to use a two-tailed test.

(c) Outline **one** way in which the student teacher could have assessed the validity of the empathy questionnaire.

**(2)**

(d) Apart from the issue of validity, identify and briefly explain **one** methodological limitation of the study.

**(2)**

(e) Explain why it was appropriate for the student teacher to use a correlation study rather than an experiment.

**(3)**

(f) The student teacher noticed that some students on her course commented that they were better able to recall information if they could read the information rather than listen to it in lectures.

Design an experiment to test the following hypothesis:

'People who are given written information will recall more than people who hear information in spoken form.'

In your answer, you should refer to the following and justify your design decisions:

- the variables to be considered
- the experimental design to be used
- the sample
- relevant materials
- an outline of the proposed procedure.

**(8)**

**(Total 18 marks)**

**11**

Two psychologists investigated the relationship between age and recall of medical advice. Previous research had shown that recall of medical advice tended to be poorer in older patients. The study was conducted at a doctor's surgery and involved a sample of 30 patients aged between 18 and 78 years. They all saw the same doctor, who made notes of the advice that she gave during the consultation.

One of the psychologists interviewed each of the patients individually, immediately after they had seen the doctor. The psychologist asked each patient a set of questions about what the doctor had said about their diagnosis and treatment. The patients' responses were recorded and then typed out. Working independently the psychologists compared each typed account with the doctor's written notes in order to rate the accuracy of the accounts on a scale of 1 – 10. A high rating indicated that the patient's recall was very accurate and a low rating indicated that the patient's recall was very inaccurate.

(a) The psychologists decided to propose a directional hypothesis. Why was a directional hypothesis appropriate in this case?

**(1)**

(b) Write a suitable directional hypothesis for this investigation.

**(3)**

(c) The psychologists were careful to consider the issue of reliability during the study. What is meant by reliability?

**(1)**

(d) Explain how the psychologists might have assessed the reliability of their ratings.

**(3)**

(e) This study collected both qualitative and quantitative data. From the description of the study above, identify the qualitative data and the quantitative data.

The psychologists used Spearman's rho to analyse the data from their investigation. They chose to use the 0.05 level of significance. The result gave a correlation coefficient of −0.52.

**(2)**

(f) Give **two** reasons why the psychologists used Spearman's rho to analyse the data.

**(2)**

(g) Using the table below, state whether the result is significant or not significant and explain why.

**(2)**

**Extract from a table of critical values of Spearman's rho ($r_S$)**

| **Level of significance for a one-tailed test** | | |
|---|---|---|
| | 0.05 | 0.01 |
| **Level of significance for a two-tailed test** | | |
| | 0.10 | 0.02 |
| N=29 | 0.312 | 0.433 |
| 30 | 0.306 | 0.425 |
| 31 | 0.301 | 0.418 |

Calculated $r_s$ must equal or exceed the table (critical) value for significance at the level shown.

(h) Explain what is meant by a Type 1 error.

**(2)**

(i) Use the information in the table above to explain why the psychologists did not think that they had made a Type 1 error in this case.

**(3)**

**(Total 19 marks)**

**12** A maths teacher wondered whether there was a relationship between mathematical ability and musical ability. She decided to test this out on the GCSE students in the school. From 210 students, she randomly selected 10 and gave each of them two tests. She used part of a GCSE exam paper to test their mathematical ability. The higher the mark, the better the mathematical ability. She could not find a musical ability test so she devised her own. She asked each student to sing a song of their choice. She then rated their performance on a scale of 1–10, where 1 is completely tuneless and 10 is in perfect tune.

(a) Suggest a suitable non-directional hypothesis for this study.

**(3)**

(b) Why might the measure of musical ability used by the teacher lack validity?

**(3)**

(c) Explain how the teacher could have checked the reliability of the mathematical ability test.

**(3)**

(d) Explain why the teacher chose to use a random sample in this study.

The results of the study are given in the table below.

**Mathematical ability test scores and musical ability ratings for 10 students**

| Student | Mathematical ability test score | Musical ability rating |
|---|---|---|
| 1 | 10 | 10 |
| 2 | 2 | 9 |
| 3 | 9 | 3 |
| 4 | 6 | 6 |
| 5 | 3 | 9 |
| 6 | 10 | 2 |
| 7 | 2 | 1 |
| 8 | 1 | 8 |
| 9 | 8 | 4 |
| 10 | 4 | 7 |

**(2)**

(e) In your answer book, sketch a graph to show the data in the table above. Give the graph an appropriate title and label the axes.

**(3)**

(f) Discuss what the data in the table above and the graph that you have sketched seem to show about the relationship between mathematical ability and musical ability.

**(3)**

(g) The teacher noticed that most of the students who were rated highly on musical ability were left-handed. The teacher is aware that her previous definition of musical ability lacked validity.

Design a study to test whether there is a difference in the musical ability of left-handed students and right-handed students. You have access to a sixth form of 200 students.

You should:

- identify the design that you would use
- explain an appropriate sampling method and justify your choice
- describe the procedure that you would use, including details of how you would assess musical ability
- write a suitable debrief for these participants.

**(10)**

(h) In your answer book, draw a table to show how you would record your results. Identify an appropriate statistical test to analyse the data that you would collect. Justify your choice.

**(3)**

**(Total 30 marks)**

**13** A study was carried out to test the effectiveness of a new anger management programme. The programme had been designed by a team of psychologists working in a young offenders' institution.

Fifteen male offenders aged 17– 21 years took part in the programme. An anger score for each offender was obtained before the start of the programme. This score was based on a questionnaire designed by the psychologists. The questionnaire had 10 items. The maximum score was 50; the higher the score, the greater the level of anger.

The month-long programme of anger management involved 8 two-hour sessions.

Throughout the programme, the offenders were told to keep a diary of situations that made them angry and to record their anger in these situations. After the programme had ended, they were told to continue to keep their diary.

Two weeks later, after the programme had ended, a second anger score was obtained for each offender. The same questionnaire was used.

The data obtained are shown in **Table 1** below.

**Table 1: Median anger scores and the ranges before and after the programme**

| | **Before** | **After** |
|---|---|---|
| **Median** | **35** | **24** |
| **Range** | **15** | **17** |

(a) Explain why measures of dispersion are often used in addition to measures of central tendency to summarise data. Refer to the results of this study in your answer.

**(2)**

(b) A Wilcoxon signed ranks test was used to test for a significant difference between the anger scores at the start of the programme and after the programme had ended.

The calculated value of *T* was found to be 22.

**Table 2: Critical values of *T***

| **Level of significance for two-tailed test** | 0.1 | 0.05 | 0.02 |
|---|---|---|---|
| **Critical value of *T* (when *N* = 15)** | 30 | 25 | 19 |

*T must be equal to or less than the critical value to be significant.*

Using **Table 2** above, explain whether or not the result is significant.

**(2)**

(c) Explain why the psychologists decided to use a Wilcoxon signed ranks test to analyse the data.

**(3)**

(d) Explain **two** possible reasons for asking each offender to keep a diary.

**(4)**

(e) An independent researcher reviewed the design of the study and noted that there was no control group.

Explain how having a control group could have improved this study.

**(3)**

(f) The independent researcher was also concerned that the psychologists had not checked the reliability and validity of the questionnaire used to measure the level of anger.

Outline how the psychologists could check the reliability **and** the validity of the questionnaire.

**(5)**

**(Total 19 marks)**

**14** Some studies have suggested that there may be a relationship between intelligence and happiness. To investigate this claim, a psychologist used a standardised test to measure intelligence in a sample of 30 children aged 11 years, who were chosen from a local secondary school. He also asked the children to complete a self-report questionnaire designed to measure happiness. The score from the intelligence test was correlated with the score from the happiness questionnaire. The psychologist used a Spearman's rho test to analyse the data. He found that the correlation between intelligence and happiness at age 11 was +0.42.

(a) Write an operationalised non-directional hypothesis for this study.

**(2)**

(b) Identify an alternative method which could have been used to collect data about happiness in this study. Explain why this method might be better than using a questionnaire.

**(4)**

(c) A Spearman's rho test was used to analyse the data. Give **two** reasons why this test was used.

**(2)**

**Extract from table of critical values from Spearman's rho($r_s$) test**

| N (number of participants) | Level of significance for a two-tailed test | |
|---|---|---|
| | **0.10** | **0.05** |
| | **Level of significance for a one-tailed test** | |
| | **0.05** | **0.025** |
| 29 | 0.312 | 0.368 |
| 30 | 0.306 | 0.362 |
| 31 | 0.301 | 0.356 |

Calculated $r_s$ must equal or exceed the table (critical) value for significance at the level shown.

(d) The psychologist used a non-directional hypothesis. Using the table above, state whether or not the correlation between intelligence and happiness at age 11 (+0.42) was significant. Explain your answer.

**(3)**

(e) Five years later, the same young people were asked to complete the intelligence test and the happiness questionnaire for a second time. This time the correlation was –0.29.

With reference to **both** correlation scores, outline what these findings seem to show about the link between intelligence and happiness.

**(4)**

**(Total 15 marks)**

**1**

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

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

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

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

(a) **AO2 / 3 = 2**

One mark for one brief finding and a further mark for appropriate elaboration or for two brief findings or one mark for a slightly muddled answer.
On average, the treatment group showed greater improvement after the treatment than the no-treatment group. The average improvement score for the no-treatment group was very low suggesting that the treatment gains for the treatment group were not simply a result of the passage of time.

There was some variation in both groups as shown by the ranges but it was wider in the treatment group. The low range in the no-treatment group suggests that most people in this group had similar low improvement scores.

(b) **AO1 = 1, AO2 / 3 = 3**

One mark for identification of a suitable test and 3 further marks for an appropriate justification.The specification only requires knowledge of non-parametric tests. However, if a candidate names an independent t-test and justifies its use, this is perfectly acceptable. It is likely that most candidates will identify a non-parametric test. The most appropriate test is the Mann-Whitney and the justifications for its use are:

- independent groups design
- at least ordinal data
- differences.

(c) **AO2 / 3 = 2**

One mark for correctly identifying the likelihood and one further mark for an appropriate explanation or one mark for a slightly muddled answer.

The likelihood of making a Type 1 error is 5%. A Type 1 error occurs when a researcher claims support for the research hypothesis with a significant statistical test, but in fact, the variations in the scores are due to chance variables. If the level of significance is set at 5%, there will always be a one in twenty chance or less that the results are due to chance rather than to the influence of the independent variable or some other factors.

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

Two marks for each reason. One mark for a basic identification and one further mark for elaboration.

Possible reasons include:

Expectations – the patients might expect the treatment to do them some good and it becomes a self-fulfilling prophesy.

Biased sample – even though the participants were randomly assigned to groups, the treatment group might, by chance have included more people with milder symptoms that were more likely to respond to treatment.

Other support – we do not know what other support/ treatment that the participants might have had over the 8 week therapy period.

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

Two marks for the advantage and two marks for the disadvantage. One mark for simply identifying an advantage / disadvantage and the further mark for elaboration in the context of the study. Answers which are not set in context cannot achieve full marks.

Advantage: Much quicker to administer and to score – could all have been given out at the same time whereas the therapist has to conduct 30 time-consuming interviews; cheaper than interviews, ie in terms of the therapist's time; people might be more comfortable, and, therefore, more honest, if they have to write responses rather than face an interviewer (could work the other way as well – see disadvantages).

Disadvantage: Self-report questionnaires might not yield as accurate data as an interview – questions can limit range of answers and there are no additional cues, eg body language, participants might be less honest on a questionnaire than in a face-to-face interview.

Marks can be awarded for any appropriate advantages / disadvantages.

(f) **AO2 / 3 = 5**

Candidates should demonstrate understanding of some of the requirements of a good consent form. For full marks, it should be succinct, clear and informative.

It is likely to include some of the following information: treatment programme that is noninvasive; requirement to be assessed on current level of functioning; use of a trained therapist to conduct interviews; duration of the programme; requirement for re-assessment at the end of the programme; random allocation to a treatment or no-treatment group.

It should show awareness of ethical considerations, eg

- no pressure to consent – it will not affect any other aspects of their treatment if they choose not to take part
- they can withdraw at any time
- they can withdraw their data from the study
- their data will be kept confidential and anonymous
- they should feel free to ask the researcher any questions at any time
- they will receive a full debrief at the end of the programme.

For full marks, candidates must include a range of both procedural and ethical points.

| **AO2 / 3 Mark Bands (5 marks)** |
|---|
| **5 marks Effective**<br>Consent form demonstrates sound knowledge and understanding of research ethics. |
| **4 – 3 marks Reasonable**<br>Consent form demonstrates reasonable knowledge and understanding of research ethics. |
| **2 marks Basic**<br>Consent form demonstrates basic, superficial knowledge and understanding of research ethics. |
| **1 mark Rudimentary**<br>Consent form is rudimentary demonstrating very limited understanding of research ethics. |
| **0 marks**<br>No creditworthy material is presented. |

(g) **AO1 = 2, AO2 / 3 = 2**

**AO1**: One mark for brief description, eg ’consistency’ and one further mark for elaboration. Reliability refers to consistency over time. If a test, questionnaire, etc, is reliable, people tend to score the same on the test if they take it again soon afterwards.

**AO2 / 3**: One mark for a very brief answer, eg ‘do another test’ or ‘test them again’ or ‘use another interviewer to check’. Two marks for some elaboration.

Reliability could have been checked by administering a valid and reliable questionnaire to the participants as well as interviewing them and then comparing the scores on the two measures. If the interview score was reliable, there would be strong positive correlation between the scores.

The interviews could have been filmed and given to another trained therapist to assess. A strong correlation between the scores given by each therapist would demonstrate reliability.

(h) **AO2 / 3 = 10**

For full marks, the method section should be written clearly, succinctly and in such a way that the study would be replicable. It should be set out in a conventional reporting style, possibly under appropriate headings. Examiners should be mindful that there are now different, but equally acceptable reporting styles. For example, candidates should not be penalised for writing in the first person. The important factor here is whether the study could be replicated.

There should be reasonable detail with regard to:

- design
- participants
- materials
- procedures.

| AO2 / 3 Mark Bands (10 marks) |
|---|
| **10 – 9 marks Effective**<br>Effective method section that demonstrates sound knowledge and understanding of investigation design.<br>The design decisions are appropriate and the description provides accurate detail of the design, participants, materials and procedure of the study.<br>Effective and appropriate report style. |
| **8 – 6 marks Reasonable**<br>The method section demonstrates reasonable knowledge and understanding of investigation design.<br>The design decisions are generally appropriate and the description provides reasonable detail of the design, participants, materials and procedure of the study.<br>Generally appropriate report style. |
| **5 – 3 marks Basic**<br>The method section demonstrates basic knowledge and understanding of investigation design.<br>Some aspects of the design are appropriate. The description provides basic detail of some features of the study or rudimentary outline of the main features.<br>Expression lacks clarity. |
| **2 – 1 mark Rudimentary**<br>The method section demonstrates rudimentary knowledge or understanding of research. The report is weak, muddled or incomplete.<br>Deficiency in expression results in confusion and ambiguity. |
| **0 marks**<br>No creditworthy material is presented. |

**2**

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

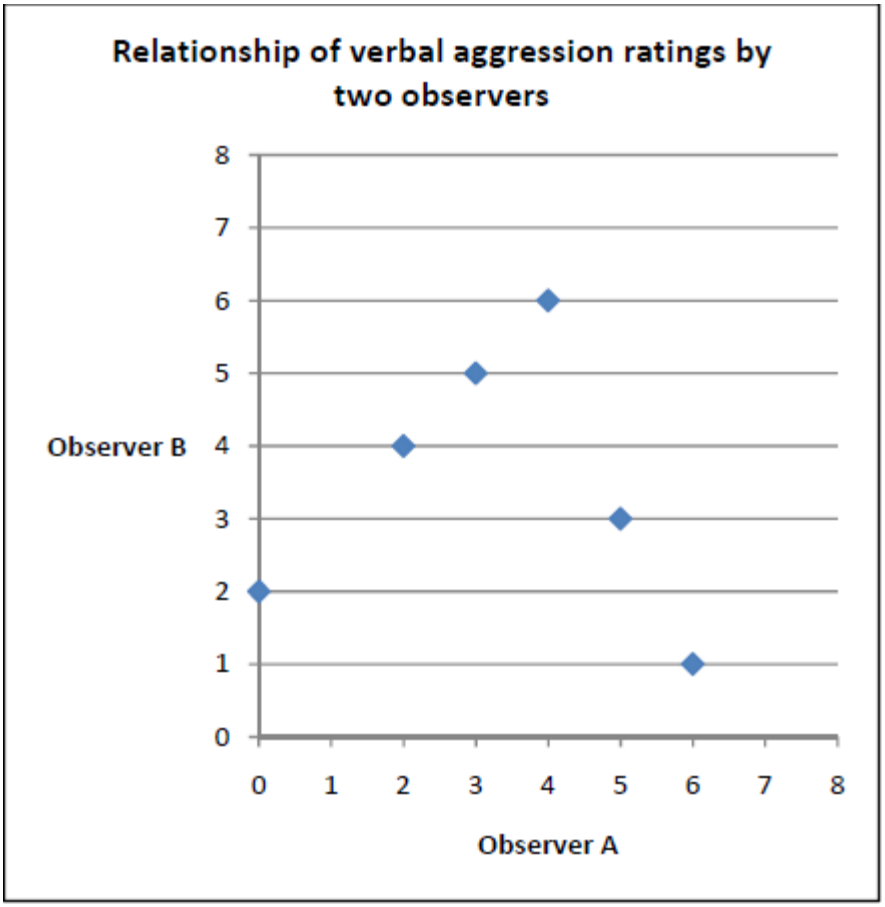
- AO1 knowledge and understanding
- AO2 application (of psychological knowledge)
- AO3 evaluation, analysis, interpretation.

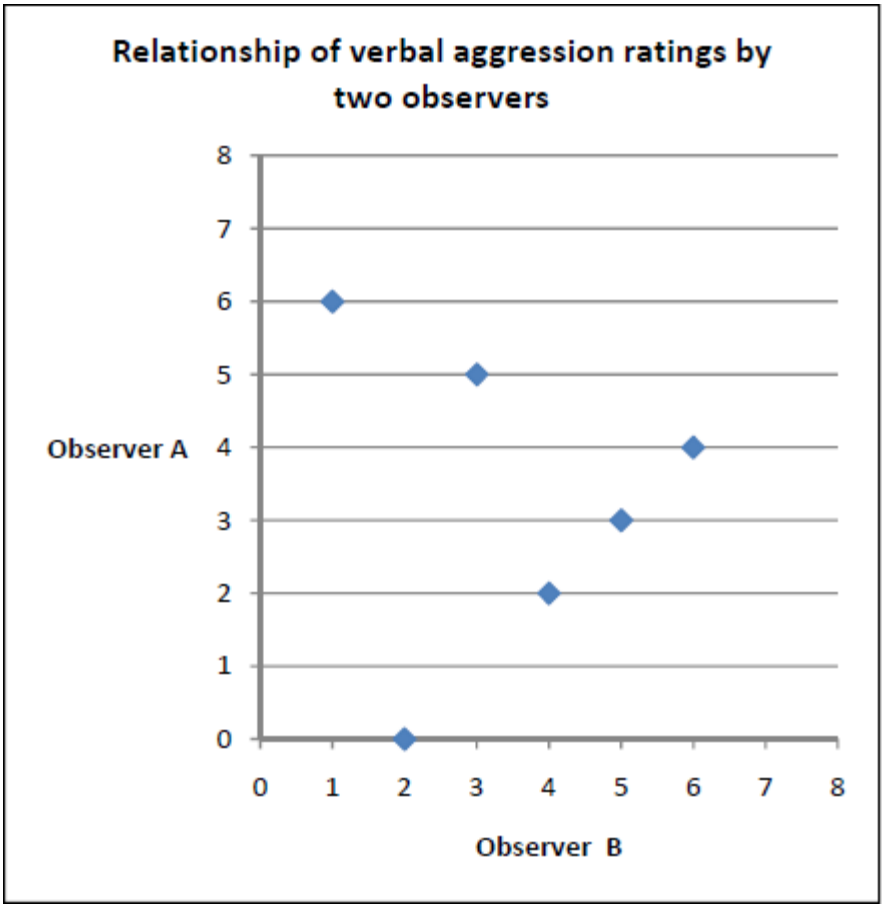
(a) **AO2 / 3 = 4**

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

Each of the examples below is a full mark answer because:

- it is clearly a sketch of a scattergram
- the data are appropriately plotted
- the labels of the axes and the title taken together show full understanding of the nature of the data.






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

For full marks, candidates should give a reasonably detailed explanation eg she is concerned because the observers should both recognise the same types of verbal behaviour as aggressive and you would expect their tallies to be very similar. In this case, the observers disagree in every 10-minute time interval even though they are both watching the same child and should be using the same criteria. In some time slots, there is a really big difference in the number of acts.
This suggests that the observers have interpreted the criteria differently or that, at certain times, one observer was more vigilant then the other (4 marks).

1 mark – 'because the observers do not agree with each other'.
3 further marks for elaboration.
Candidates who simply describe what is meant by inter-rate reliability can gain no marks.

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

1 mark for identifying the appropriate test – Spearman's Rho or Pearson's (with appropriate justification).
2 further marks for explaining why it is appropriate ie the psychologist is testing for a correlation and the data that can be treated as ordinal.
Candidates can gain no marks on this question if their choice of statistical test is inappropriate.

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

1 mark for a very brief answer eg 'better training for the observers'
3 further marks for elaboration.

There is a breadth / depth trade-off here. Candidates can elaborate on one improvement eg explain how the training might be improved or outline several improvements in less detail eg establish clearer criteria for categorising verbal aggression, filming the child so that the observers can practise the categorisation.

**3**

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

| Level | Marks | Description |
|---|---|---|
| 3 | 5 – 6 | Conclusions in respect of both means and standard deviations are presented with clarity. Understanding of the relevance of each statistic is demonstrated. Justifications for each make good use of the values given. |
| 2 | 3 – 4 | Conclusions and justification in respect of both means and standard deviations are relevant, but there is some lack of clarity in both.<br>Or, one is done well and justified appropriately (most usually this will be the mean). |
| 1 | 1 – 2 | One conclusion is drawn or two are partially correct. Any justification is limited. The answer lacks clarity. |
| | 0 | No relevant content. |

**Means**

- Conclusion: when people believe they are presenting to a large audience they are less fluent in their spoken communication than when they believe the audience is small (or vice versa).
- Justification / Application: this is supported by the difference in the mean fluency scores which show more verbal mistakes (on average 6 more mistakes) when the audience is believed to be large (or vice versa).

**Standard deviations**

- Conclusion: performances of participants in Condition A where audience is believed to be small are less varied / dispersed / spread out than in Condition B where audience is believed to be large (or vice versa).
- Justification / Application: lower SD in Condition A suggests that individual performances in Condition A were more similar to each other and / or all quite close to the mean of 11.1.

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

**1 mark** – this would be an improvement because the SD is a measure of dispersion that was less easily distorted by a single extreme score.

**Plus**

**1 mark** – one that takes account of the distance of all the verbal error scores from the mean.

**Plus**

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

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

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

**Plus**

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

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

**OR**

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

- data should be treated as ordinal. Cannot assume interval data because verbal errors cannot be assumed to be of equal size (ie one verbal error is not equivalent to any other verbal error)
- the experimental design is independent groups
- the psychologist is looking for a difference between the two conditions
- SDs are quite different.

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

**2 marks** for a clear and appropriate definition as follows:

This means that there is a less than 5% likelihood that this difference would occur if there is no real difference between the conditions **OR** the researchers would have a 95% confidence level.

**1 mark** for a less clear answer which shows some understanding, eg this means the researcher can conclude that the difference was not due to chance.

Accept any other valid answer.

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

**2 marks** for a clear and detailed explanation applied to this study.

**1 mark** for a partial or muddled explanation or one that is only loosely applied to the study.

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

- asking other people if verbal errors are a good measure of verbal fluency (face validity)
- giving participants an alternative / established verbal fluency test and checking to see that the two sets of data are positively correlated (concurrent validity).

**4**

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

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

(a) **AO2 / 3 = 1**

1 mark for correct answer – directional (one-tailed is acceptable).

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

1 mark for correctly stating that the result is significant.
2 further marks for an explanation: the calculated value of $T = 53$ which is less than the value of 60 where N = 20 and $p \leq 0.05$ for a one-tailed test.

If the candidate states that the result is not significant, no marks can be awarded.

**5**

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

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

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

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

(a) **AO2/3 = 6**

Candidates need to show that they understand what differentiates opinion from scientific evidence. They could mention some of the following:

- The teacher has only experienced one school in a particular catchment area so she has only observed a very limited number of 5 year-olds (issues of sampling and replicability).
- She has found out that children do not eat anything nourishing simply by chatting with the children. She has no corroborative evidence from eg parents (issues of objectivity).
- She uses vague phrases such as 'decent breakfast' without being clear what this means (operationalisation).
- She has generated a theory and made predictions based on flimsy evidence.
- She has not used any scientific method to lead to her conclusions eg a carefully controlled experiment, survey or observation.
- She has drawn conclusions about the effects of breakfast without considering other variables which might affect reading skills and behaviour.

| **AO2/3 Mark bands** |
|---|
| **6 marks Effective**<br>Explanation demonstrates sound understanding. Application of knowledge is effective and shows coherent elaboration. Ideas are well structured and expressed clearly and fluently. Consistently effective use of psychological terminology. |
| **5 – 4 marks Reasonable**<br>Explanation demonstrates reasonable understanding. Application of knowledge is reasonably effective and shows some elaboration. Most ideas appropriately structured and expressed clearly. Appropriate use of psychological terminology. |
| **3 – 2 marks Basic**<br>Explanation demonstrates basic, superficial understanding. Application of knowledge is basic. Expression of ideas lacks clarity. Limited use of psychological terminology. |
| **1 mark Rudimentary**<br>Explanation is rudimentary, demonstrating very limited understanding. Application of knowledge is weak, muddled and may be mainly irrelevant. Deficiency in expression of ideas results in confusion and ambiguity. The answer lacks structure, often merely a series of unconnected assertions. |
| **0 marks**<br>No creditworthy material is presented. |

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

In a random sample, every member of the identified population has an equal chance of selection. In this case, the sampling frame consists of the 400 five-year-old children attending ten local schools. In order to obtain a simple random sample, the researcher has to have the names of all 400 children and can then select using one of the following methods:

- **Random number tables** – random number tables are specially devised to meet the following criteria – they contain strings of numbers where each number has the same chance of being selected as any other and each number is independent of the others. Such tables are readily available in statistics text books etc or can be generated by the researcher using a computer program. The researcher assigns each child a number between 1 and 400. He enters the table at any place (he could close his eyes and point with a finger at a starting place) and then moves either horizontally or vertically to produce a string of random numbers. He records all the numbers which correspond to the 400 children until he has recorded a total of 100 non-duplicated numbers.

- **Computer selection** – This is a similar method where the computer does most of the work. A computer can generate an endless string of random numbers ie numbers which have no relationship to one another as a sequence. Each child's name is given a number and a random number generator program is used to produce the required sample size (in this case 100 participants).

- **Manual selection** – Using this method, the researcher has to put each name (or an assigned number) on a separate slip of paper and place them all in a container. The researcher then selects 100 slips from the container. The following conditions could apply: the container should be shaken between each draw; the slips of paper should all be the same size and folded in the same way so that one does not feel different from another; the selector draws 'blind' ie cannot see the actual slips of paper.

A simple definition of a random sample is not credit-worthy since it offers no explanation.\ Similarly, answers which only use the word 'random' as an explanation cannot gain credit eg He would choose 100 participants at random from the children. One mark for a very basic method eg 'he would take names from a hat / computer / random number table'. Two further marks for elaboration.

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

Candidates could focus on:

- Even if a sample is random, it may not be truly representative of the population eg might all come from the same school, or be all boys or all girls.
- Practical limitations eg the time and effort needed to write out 400 slips for the manual method.
- Difficulties of obtaining a truly random sample eg even if the sample is selected randomly, parents might refuse to allow their children to participate.

Any plausible and appropriate answers should be credited.
Up to 2 marks for identification of limitations. For 3 marks, one or more limitations must be explained in reasonable detail.

(d) **AO2/3 = 5**

There are two requirements to this question, **why** operationalising variables is important and **how** to operationalise the IV and the DV. If a candidate only explains **how / why**, maximum 3 marks.

The terms' 'decent breakfast' and 'reading skills' are vague. It is important from the point of view of objectivity, replicability and control of extraneous variables to make sure that these terms are closely defined.

Suggestions as to how the psychologist might do this could include the following:

The researcher needs to specify the exact composition of the breakfast (possibly by doing a pilot study or a literature search to identify the components of breakfast most likely to bring about behavioural / cognitive change). He probably also needs to specify the time at which it is consumed. The researcher needs to use a standard reading test which should be administered to all the participants at the beginning of the study and at the end – the dependent variable is likely to be the improvement score.

(e) **AO2/3 = 2**

Reasons are:

- a test of difference
- data (scores from a reading test) are at least ordinal, this would include ordinal / interval and / or ratio
- independent design.

One mark for each appropriate reason (maximum 2 marks).

(f) **AO2/3 = 2**

It would have been more difficult to use a matched-pairs design because of the number of relevant factors that would need to be controlled (eg gender, intelligence, parental attitudes / income / education, experience of pre-school education, number of siblings in family etc). There is a relatively small pool of children available (ie 400) and it could be difficult to match on all these factors. It would also be very time-consuming; it could be quite expensive to carry out the necessary surveys; it could be quite intrusive collecting such information from parents.

One mark for a basic explanation eg “Because it is difficult to match participants appropriately”.
One further mark for elaboration.

(g) **AO2/3 = 2**

One mark for identifying an appropriate issue and second mark for explaining how it could be addressed.

The most likely issue is confidentiality which could be addressed by ensuring that all scores on reading scales and all personal information are anonymised.
There are also ethical problems involved in denying the control group breakfast although it is more difficult for candidates to suggest a way of addressing this – perhaps to put only those children into the control group who do not eat breakfast anyway, restricting the study length to a short period of time and, if the study results support the hypothesis, to provide free breakfasts to these children for the rest of the academic year.
Parental consent is excluded because it is given in the stem so answers which offer this as an issue cannot gain credit.

(h) **AO3 = 12**

Design should be written clearly, succinctly and with sufficient detail for reasonable replicability.
Candidates will not receive credit for details included in the stimulus material. These include using a random sample of 100 children, gaining parental consent and selection of a Mann Whitney test.
To access marks in the top band candidates must state an appropriate hypothesis in which “playground behaviour” is clearly operationalised. The hypothesis could be directional or non-directional.
Given the wording of the question, a correlational hypothesis is not credit-worthy, however, the rest of the answer should be marked on its merits.
Likely aspects of “playground behaviour” would include activity levels, aggression, cooperative play etc.
An attempt to operationalise “a healthy breakfast” should be credited. However, candidates could assume this had already been done by the psychologist.

As this is an observational study any of the following, together with appropriate justification, would be credit-worthy:
Is the observation covert or overt?
Where are observers positioned? (In playground, watching from window?)
Is a video recording of the children used? How will this be analysed (eg content analysis)?
Do the students who observe know what the children ate for breakfast?
At what times of day does the observation take place?
How many children are observed? (Candidates could justify using a smaller sub-sample of the 100 children in the original study).
How long does each observation last?
Will the observers use a behavioural check list / tally chart?
Will more than one observer observe each child? If so, what training will be given and what checks for inter-observer reliability will take place?
Reference to time sampling or event sampling.

Credit any other relevant material.

| **AO2/3 Mark bands** |
| --- |
| **12 – 10 marks Effective design**<br>A design that demonstrates sound knowledge and understanding of observational research. The selection and application of research techniques is appropriate. The description provides sufficient detail for most aspects of the study to be implemented. Some design decisions are justified effectively. |
| **9 – 7 marks Reasonable design**<br>The design is reasonable and demonstrates knowledge and understanding of some aspects of observational research. The selection and application of research techniques is mostly appropriate. The description provides sufficient detail for some aspects of the study to be implemented. Some design decisions are justified. |
| **6 – 4 marks Basic design**<br>The design is basic and demonstrates limited knowledge and understanding of aspects of observational research. The selection and application of research techniques are sometimes appropriate. Some basic design decisions/features of the study are described but there may be significant omissions, lack of clarity and possibly some implausible suggestions that severely limit implementation. Justifications of the design are limited. |
| **3 – 1 marks Rudimentary design**<br>The design is rudimentary. Design decisions are muddled and or mostly inappropriate and are not justified. Description lacks clarity. The study could not be implemented. |
| **0 marks**<br>No creditworthy material. |

**6**

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

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

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

'There is an association between birth order and choice of career' = 2 marks.
A directional hypothesis is not credit-worthy. Reference to a relationship / correlation cannot gain credit.

Although technically, the psychologist is looking for an association, candidates can gain credit for expressing the hypothesis in terms of a difference eg 'There is a difference in career choice depending on birth order.'

2 marks for a clear hypothesis, 1 mark for a hypothesis which lacks clarity.

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

One mark for identifying a sampling method.
One mark for a brief explanation of how to obtain the sample eg 'by advertising for lawyers or artists to come forward'. One further mark for elaboration eg 'by explaining that adverts would have to be placed in appropriate journals etc to attract these particular categories of participants'

Candidates who identify a sampling method but describe it incorrectly can be awarded 1 mark.

(c) **AO2 / AO3 = 12**

This is a 12 mark question but marks are allocated to each of the required components as follows:

- An appropriately labelled table = 2 marks

1 mark for a table that displays the data in the question.
2 marks for a table which includes data relating to non first-born children. Totals are not required for the 2 marks.

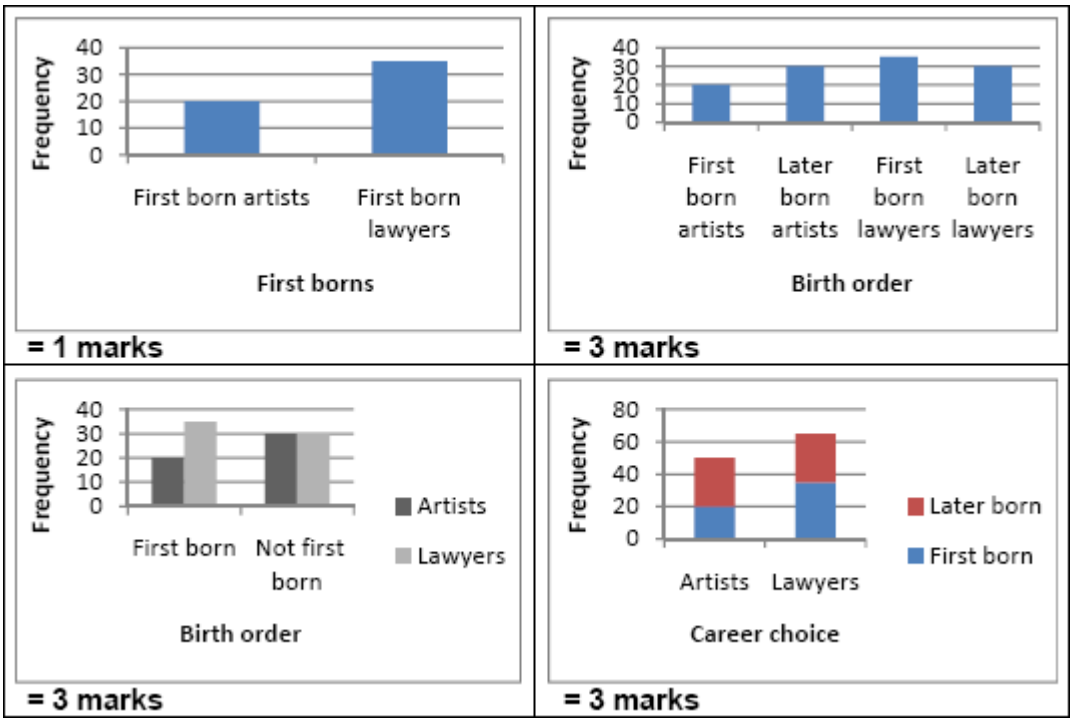
Table: Table to show the career choices of first born and non-first born children

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

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

For 3 marks, candidates need to display the data relating to first born and non-first born career choices on a bar chart. They should label axes correctly and draw the columns to the correct approximate height for a sketch.

For 2 marks, candidates display data as above but labels are missing or lack clarity. For 1 mark, candidates graph the data supplied in the question relating to first born career choices only.




**NB** Labelled axes but no bars = 0 marks.

- identification of appropriate statistical test and justification = 1 + 2 marks

An appropriate test here is the Chi-squared.
Justification gains 2 marks. Any two correct reasons from:

- data are independent
- level of measurement is nominal
- test of association / difference is required.
- identification of appropriate significance level = 1 mark.

The most likely significance level is 5% ($p \leq 0.05$). Candidates are not asked to justify their choice. Candidates who choose a more stringent level can achieve marks but they must then follow this through when they make their statement of results.

Candidates who erroneously report 0.05% or p = 0.5 do not gain credit for level of significance but can achieve credit for the statement of results in relation to the hypothesis.

- a statement of the results of the statistical test in relation to the hypothesis = 3 marks.

For full marks, the candidate should state whether or not they can accept the hypothesis (or they can express this in terms of rejecting the null hypothesis) at a given significance level and refer to the observed and critical values.

Where candidates choose an inappropriate value from the table but interpret that value correctly, they can gain 2 marks.

The critical value for $x^2$ (df =1 p 0.05 (two-tailed)) is 3.84. As the observed value of $x^2$ 2.27 is less than the critical value, we cannot reject the null hypothesis. There is not an association between birth order and career choice.

**7**

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

**2 marks** for identifying two factors that are relevant for use of the sign test: nominal/categorical data; test of difference; related design/repeated measures.

**Plus**

**Up to 2 marks** for application of these to the investigation described:

- Nominal data as patients are assigned to one of three categories – 'improved', 'deteriorated' or 'neither'.
- Testing for difference in the number of absences in the year following and prior to treatment.
- Repeated measures as the same patients' work records are compared before and after treatment.

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

**1 mark** for identifying the correct value of s as 5

**Plus**

**1 mark** for explanation/calculation of how this was arrived at:

- The most commonly occurring sign is + (12) and the least frequently occurring sign is – (5). The 0s are disregarded.
- The total for the least frequently occurring sign is the value of s = 5

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

**1 mark** for stating that the value of s (5) is not significant at the 0.05 level.

**Plus**

**1 mark** for explanation:

- The critical value is 4. As the calculated value is higher than/exceeds the critical value, the result is significant not at the 0.05 level.

Accept alternative wording

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

Marks may be awarded for a single point that is expanded/elaborated or more than one point briefly stated.

1 mark only if there is no reference to the investigation described.

**Possible points:**

- Primary data are obtained 'first-hand' from the participants themselves so are likely to lead to greater insight: e.g. into the patients' experience of treatment, whether they found it beneficial, negative, etc.
- Secondary data, such as time off work, may not be a valid measure of improvement in symptoms of depression. Primary data are more authentic and provide more than a surface understanding: e.g. participants may have taken time off work for reasons not related to their depression.
- The content of the data is more likely to match the researcher's needs and objectives because questions, assessment tools, etc. can be specifically tailored: e.g. an interview may produce more valid data than a list of absences.

(e) **[AO1 = 3 AO2 = 2]**

| Level | Marks | Description |
|---|---|---|
| 3 | 4 – 5 | Knowledge of the implications of psychological research for the economy is clear. Application to the investigation described is effective. The answer is coherent with effective use of terminology. |
| 2 | 2 – 3 | Some knowledge of the implications of psychological research for the economy is present but there is a lack of detail/clarity. Application to the investigation described is limited or absent. Terminology is used appropriately on occasion. |
| 1 | 1 | An implication of psychological research for the economy is briefly stated. |
| | 0 | No relevant content. |

**AO1 – possible content:**

- Psychological research may lead to improvements in psychological health/treatment programmes which may mean that people manage their health better and take less time off work.
- Absence from work costs the economy an estimated 15 billion a year annually and much of this absence is due to 'mild' mental illness: e.g. stress, anxiety.
- Psychological research may lead to better ways of managing people whilst they are at work to improve productivity: e.g. research into motivation and workplace stress.
- 'Cutting-edge' scientific research may encourage investment from overseas companies into this country.

Credit other relevant points/implications, including examples not linked to psychopathology.

**AO2 – application**

- If research (such as the investigation described) suggests that depressives are better able to manage their condition following CBT and return to work, then it may benefit the economy to make treatment more widely available, improve funding, etc.
- Psychological research such as this plays an important role in sustaining a healthy workforce and reducing absenteeism.

Credit other relevant application points.

**8**

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

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

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

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

(a) **AO1 = 2**

Content analysis is a technique for analysing qualitative data of various kinds. Data can be placed into categories and counted (quantitative) or can be analysed in themes (qualitative).

Award 1 mark for a brief statement and a further mark for elaboration.

(b) **AO3 = 4**

- The psychologist could have begun by watching some of the film clips of driver behaviour.
- This would enable the psychologist to identify potential categories which emerged from the data of the different types of distractions seen in the film.
- Such categories / themes might include: passenger distractions, gadget distractions, etc.
- The psychologists would then have watched the films again and counted the number of examples which fell into each category to provide quantitative data.

Credit variations in so far as they explain the process.

Note: maximum 1 mark if no engagement with the stem.

| **AO3 Mark bands** |
|---|
| **4 marks Effective**<br>Effective explanation of the processes involved in content analysis referring to some or all of the above points. |
| **2 – 3 marks Reasonable**<br>Reasonable accurate coverage of the processes involved. |
| **1 mark Basic**<br>Basic identification of the processes involved in content analysis ('watching the films and counting'). |
| **0 marks**<br>No creditworthy material. |

(c) **AO3 = 3**

1 mark for identification of an appropriate way of assessing reliability in this investigation. By far the most likely answers here are inter-rater reliability or test-retest reliability.

2 marks for some explanation / elaboration: 'the two psychologists could carry out content analysis of the films separately and compare their answers' or 'they could re-code the films at a later date and compare the two sets of data'.

3 marks for an accurate and clear explanation which refers to deriving the categories and checking the data. 'The two psychologists could watch the films separately and devise a set of categories. They could compare these and use categories they both agreed on. They could carry out content analysis of the films separately and compare their answers looking for agreement'.

(d) **AO3 = 3**

Candidates can cover one reason explained in detail here or several reasons in less detail.

A repeated measures design was chosen in this experiment:

- to remove the effects of individual differences in reaction times which would occur if an independent groups design was used
- to avoid the potential difficulties involved in matching participants
- to reduce the number of participants required for the experiment.

(e) **AO3 = 3**

This is a repeated measures design and is counter-balanced hence points about order effects and individual differences will not gain credit.

There are a range of potential extraneous variables here including:

- the nature and content of the conversation with the psychologist on the hands-free phone
- interaction between the sex of the psychologist and sex of participant which could influence the type of conversation
- the number of hazards in the computer-based test, hence difficulty of the tests
- the presence of the hands-free headset could have produced distraction.

Award 1 mark for basic identification of a confounding variable and a further 2 marks for elaboration of how this could have affected the dependent variable.

*Example: The chat with the psychologist was not controlled (1 mark) so the difficulty or number of questions could have varied (2 marks). This would influence the DV as more or less attention would be required (3 marks).*

(f) **AO3 = 4**

There are several potential ethical issues here. Candidates can focus on one in detail or several in less detail.

- Protection of participants from harm whilst studying the effects of a hands-free phone on driving. Two key issues here are the use of a computer-based test with no risk attached and of an experienced sample of police drivers.
- Informed consent: Participants should be given full information about the nature of both tasks before deciding whether or not to participate.
- Debriefing: A full debriefing should take place at the end of the experiment. This should provide feedback on performance and allow participants to ask questions if they wish to.
- Freedom to withdraw: Participants should be made aware of their freedom to withdraw before and during the experiment. They should be made aware of their right to withdraw their data after the experiment.
- Confidentiality: Individuals should not be identified, but should retain anonymity (use of numbers or initials instead of names).

Lists of ethical issues with no elaboration 1 mark.

| **AO3 Mark bands** |
|---|
| **4 marks Sound**<br>An appropriate ethical issue is identified and explained in detail. Material is accurate – or several issues are identified and discussed accurately in less detail. |
| **2 – 3 marks Reasonable**<br>One or more appropriate ethical issues are identified and discussed. The answer is generally accurate. |
| **1 mark Basic**<br>Basic identification of an ethical issue (e.g. ‘right to withdraw’) or very brief answers which lack detail. |
| **0 marks**<br>No creditworthy material. |

(g) **AO3 = 5**

The standardised instructions should include the following information:

*a. You will take part in a simulated driving test which will last for three minutes.*

*b. Your task will be to identify potential hazards on the road ahead.*

*c. When you see a hazard, you should press the mouse button as quickly as possible.*

*d. Whilst you are doing the test, I will chat to you on a mobile phone and I would like you to reply using the hands-free mobile phone headset.*

*e. Do you have any questions?*

For full marks, the instructions should adopt an appropriate formal tone. Instructions which are not suitable to be read out should be awarded a maximum mark of 2.

| **AO3 Marks bands Standardised instructions** |
|---|
| **5 marks Effective**<br>The standardised instructions provide accurate detail of the procedure in a clear and concise form and participants' understanding is checked. |
| **4 – 3 marks Reasonable**<br>The standardised instructions provide sufficient detail of the procedure in a reasonably clear form. |
| **2 marks Basic**<br>The standardised instructions provide some details of the procedure though these may not be clear. |
| **1 mark Rudimentary**<br>The standardised instructions provide few details of the procedure and may be muddled and or inaccurate. Omissions in the instructions compromise the procedure. |
| **0 marks**<br>No creditworthy material is presented. |

(h) **AO3 = 3**

Students are required to identify an appropriate test and are asked to justify their choice.

Award 1 mark for identification of the Wilcoxon (signed ranks) test. Candidates could receive credit for Sign test or related t test. Note that reasons / justification must be correct for the test supplied.

If an incorrect test is identified **no marks** can be awarded.

Award 1 mark for basic statement of a reason, and a further mark for elaboration, within the context of the experiment or a further reason.

e.g. for Wilcoxon test:

- A repeated measures design was used (1 mark) as drivers take part in both the hands-free phone and non-phone (silent) conditions (1 mark).
- A repeated measures design was used (1 mark) and the data can be treated as ordinal (1 mark).

Test of difference cannot gain credit.

(i) **AO3 = 2**

Students are told that the difference in reaction times was significant at the $p \leq 0.01$ level.

Award 1 mark for a basic understanding of this ('the result is highly significant') and a further mark for elaboration e.g. identifying that the probability of a Type 1 error here is less than 1 / 100.

(j) **AO3 = 3**

Replication is an important tool in the scientific method. It allows scientists to check findings and ensure that they are robust. In this study, replication is important, as the original sample is small (30 people) and specific (experienced police drivers). For this reason, replication on a larger sample will be used to check if findings apply outside this specific group.

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

**9**

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

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

(a) **AO2 / AO3 = 4**

Up to four marks are awarded for discussing advantage(s) of using a laboratory experiment in this case.

The most likely advantages of the laboratory setting in this experiment include:

- Control over extraneous variables. The lab setting meant that extraneous variables could be minimised. In this experiment, outside factors such as waiting time, noise and stress (which would be difficult to control in a field experiment) were removed.
- Ethical issues. In this case, the testing of memory in a field experiment would have involved ethical issues including deception of patients or withholding of information.

Candidates may also refer to other advantages of the laboratory setting such as replicability. These can receive full credit if they contextualised within the scenario.

Award four marks for an answer which provides accurate and detailed discussion of relevant advantage(s) with a clear link to the scenario.

Award two or three marks for an answer which includes discussion of relevant advantage(s), with some reference to the scenario.

Award one mark only for an answer which merely identifies one or more relevant advantage(s) of a laboratory experiment appropriate to this scenario.

Advantages of laboratory experiments which are not relevant to this study cannot gain any credit eg use of technical equipment.

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

- One mark for correctly identifying the Mann Whitney U test or independent t test.
- One mark awarded for an accurate reason for choice (for Mann Whitney these are: test of difference, independent groups design / independent data or data which can be treated at an ordinal level).

**10**

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

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

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

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

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

One mark for answers either:

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

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

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

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

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

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

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

1 mark for a knowledge of a way (not just naming a type of validity) and 2nd mark for explaining how this would be implemented in this case. Most likely answers will address face validity or concurrent validity, but accept any other way such as construct validity, content validity, criterion validity and predictive validity.

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

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

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

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

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

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

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

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

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

**Mark bands**

| | |
|---|---|
| **8 – 7 marks** | **Very good answers**<br>All 5 points well addressed and some sound justification.<br>Answer shows sound knowledge and understanding and an ability to design an appropriate experiment. The proposal is coherent and feasible, and includes details of all the essential elements of the chosen design. Information allows for clear understanding of the proposed design. There may be some minor omission(s) at the bottom of the band. |
| **6 – 5 marks** | **Good answers**<br>3 or 4 points well addressed and some justification.<br>The design shows knowledge and understanding and some ability to design an appropriate experiment. The proposal is feasible but may lack the clarity and coherence of the top band.<br>There may be some inaccuracies and omissions. |
| **4 – 3 marks** | **Average to weak answers**<br>At least 3 points are addressed and attempt at justification.<br>The answer shows some knowledge and understanding but detail of the proposal may lack clarity.<br>There are inaccuracies and omissions. |
| **2 – 1 marks** | **Poor answers**<br>1-2 points are addressed.<br>There must be some relevant material. The experimental method may not be obvious. There may be substantial confusion, inaccuracy and / or irrelevance. |
| **0 marks** | **No relevant content** |

**11**

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

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

(a) **AO2 / AO3 = 1**

One mark for an accurate reason: The decision to use a directional hypothesis was based on findings of previous research which pointed to an effect in a particular direction ie memory is poorer with age.

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

A suitable directional hypothesis would be 'There is a negative correlation (relationship) between age and recall accuracy rating'.

- 3 marks for a fully operationalised hypothesis as above
- 2 marks for a directional correlational hypothesis that identifies age and recall as the two variables but is not fully operationalised
- 1 mark for a directional hypothesis where the variables are not identified (‘there will be a negative correlation’) or where the hypothesis lacks clarity.

Award zero marks for a non-directional or null hypothesis or any hypothesis predicting a difference or association.

(c) **AO1 = 1**

One mark for an accurate definition: The extent to which results or procedures are consistent or simply 'consistency'.

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

One mark for identification of a way of ensuring reliability. By far the most likely answer here is inter-rater reliability.

Two marks for some explanation/elaboration: using two separate psychologists and comparing them.

Three marks for an accurate and clear explanation: using two separate psychologists to rate the typed accounts for accuracy and comparing / correlating the ratings to see how similar they are.

Candidates could make a case for test retest which would involve the same psychologist re-examining the ratings after a period of time.

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

Award one mark for correct identification of one of each type of data.

- Qualitative data: the patient’s responses, the typed accounts, the doctor’s notes.
- Quantitative data: the ratings of recall accuracy on a scale of 1 – 10, ages of patients.

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

One mark for each accurate reason given:

- the researchers are testing for a correlation or a relationship between two variables.
- the data is to be treated as ordinal because the recall accuracy is in the form of ratings.

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

One mark for stating that the result is significant.

Second mark for explaining that -.52 exceeds .306 ($p \leq 0.05$, n=30 for a one-tailed test).

(h) **AO1 = 2**

One mark for a brief or muddled answer which hints at rejecting HO / accepting the H1 in error.

Two marks for explaining the term: where the researcher rejects the null hypothesis (or accepts the research / alternative hypothesis) when in fact the effect is due to chance – often referred to as an error of optimists.

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

3 marks for a clear explanation which is based on comparison of the calculated value of rs with the critical value at the 0.01 level of significance and indicates competence in use of statistical tables as follows:

- A Type 1 error is unlikely because the calculated value of rs (-0.52) exceeds the critical table value at both the 0.05 and 0.01 level for a one-tailed test.
- The chance of a Type1 error occurring is therefore less than 1%.
- This means that the researchers can be 99% certain that the results obtained are not due to chance.

Award one mark for a brief explanation (it is significant at 0.01).
Award two further marks for an explanation which refers to two of the above points.

Award one mark for stating that the obtained value (-0.52) exceeds the critical value (0.306) by a reasonable margin.

**12**

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

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

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

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

The main issue is that the teacher has made up her own test:

- This involved subjective judgement on the part of the teacher who rates the students' musical ability. Her judgement may not reflect real differences in musical ability and is likely to differ from other people's judgement and / or any absolute criteria for tunefulness.
- Lack of reliability in rating musical ability would compromise the validity of the measure.
- As the students can choose the song they will sing, the rating of ability could reflect the teacher liking / dislike of the song rather than the student's ability.
- The rating may be invalid as the students selected songs which varied in difficulty so the tunefulness reflected the difficulty of the song not the students' ability.
- Operationalising musical ability as tuneful singing is a very narrow measure. Someone can have musical ability such as playing an instrument which would not be reflected by this measure.

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

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

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

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

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

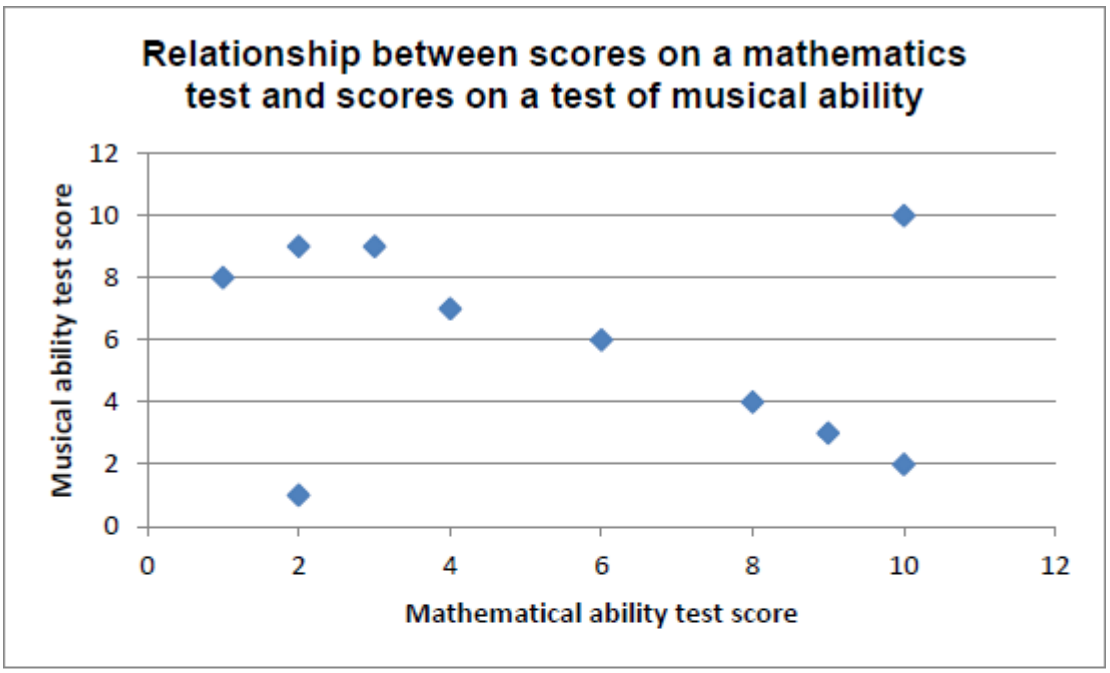
The teacher chose to use a random sample because it would probably be more representative of the whole GCSE group than if she had used an opportunity or volunteer sample. Candidates could also say that she had ready access to her target population making it convenient for her to select a random sample.

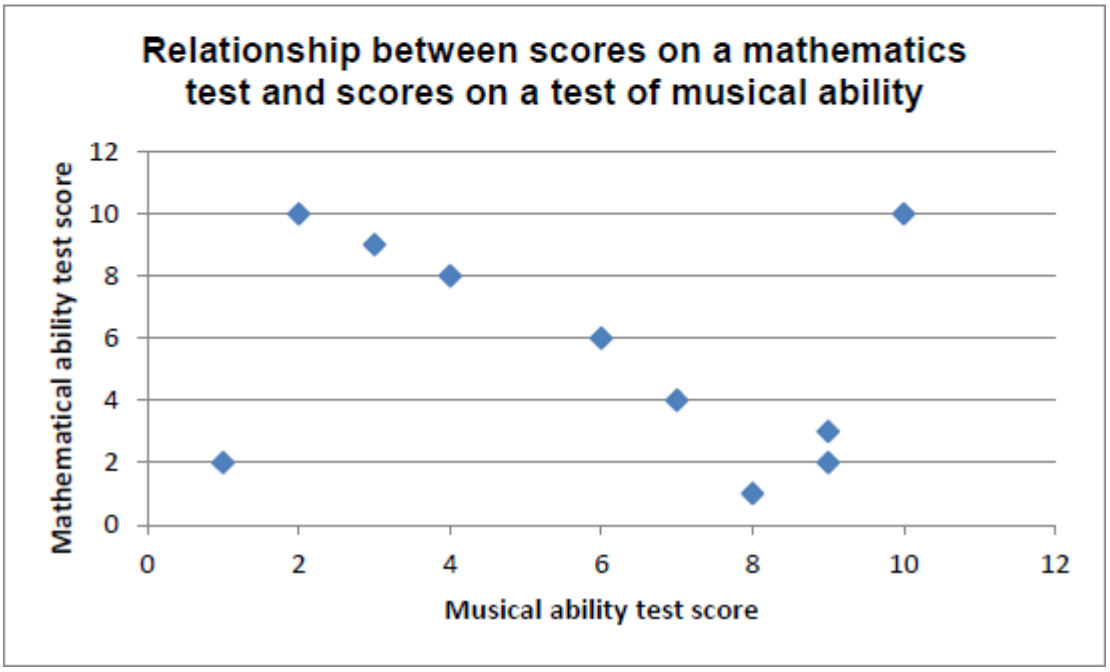
No credit for definition of a random sample.
1 mark for a brief or muddled reason (it is not biased).
2 marks for a reason that clearly points to an advantage of random sampling. This could be achieved through a comparison with another method (it is less likely to be biased than a volunteer sample).

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

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

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



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

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

- The graph seems to show a negative correlation between mathematical and musical ability.
- This means that high scorers in mathematical ability tend to achieve low scores on musical ability and vice versa.
- The presence of two strong outliers, means that the actual correlation is very weak and closer to zero.
- Comment on the small sample size which limits the conclusions that could be drawn.
- Credit can be achieved for plausible interpretations of the strength of the correlation which are justified (ie looks moderate to strong or the outliers make it weak in practice) or those based on rough calculations (around -0.2).

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

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

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

Design – 1 mark

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

Sampling – 2 marks

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

Procedure and assessment of musical ability – 4 marks

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

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

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

Debrief – 3 marks

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

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

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

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

**Musical ability scores:**

| Participant number | Left handed | Right handed |
|---|---|---|
| **1** | | |
| **2** | | |
| **3** | | |

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

**13**

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

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

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

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

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

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

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

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

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

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

Up to two marks for each reason and explanation. Likely points: as an aid to memory; a qualitative measure to supplement the quantitative data collected; to check the validity of the questionnaire; part of the therapeutic process / increased self-awareness.
Accept other valid reasons.
One mark for an appropriate reason and one mark for an explanation of the reason.

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

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

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

Accept ‘scientific’ and ‘validity’ only if justified.

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

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

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

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

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

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

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

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

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

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

**14**

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

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

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

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

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

Award 2 marks for an appropriate non-directional hypothesis which is operationalised. ‘There is a relationship between happiness scores on a questionnaire and intelligence test scores’.

Award 1 mark for a non-directional hypothesis which is not fully operationalised or lacks clarity (‘there is a relationship between happiness and intelligence’).
Award no marks for a null or directional hypothesis, or one that predicts a difference / link / association / connection.

(b) **AO2/AO3 = 4**

An interview is the most likely answer. An interview would be a more appropriate method than a questionnaire as it enables questions to be clarified and responses to be probed, thus overcoming the main disadvantages of questionnaires.

Students could also make a case for the analysis of diaries/written materials as a way of collecting data about happiness. These would generally overcome the problems of social desirability and demand characteristics inherent in questionnaires. Students could also make a case for the use of observation.

Award one mark for identifying an appropriate method. Award up to three further marks for an explanation of why this method would be better than a questionnaire.

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

Award 1 mark each for any two of the following reasons:

- Study is looking for a correlation (relationship)
- Suitable for pairs of scores
- The data type obtained is ordinal, at least ordinal or interval level
- Linear relationship between scores.

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

Students should state that the obtained value of + 0.42 exceeds the critical value for a twotailed test (.362) for N = 30. The results are therefore statistically significant ($p \leq 0.05$) Award 2 marks for a student who supplies two pieces of information. Award 1 mark for a student who states that the results are significant but does not provide an explanation OR the student who states results are significant but uses incorrect values from the table. Award 0 marks for students who argue that results are not significant.

(e) **AO2/AO3 = 4**

This question requires students to interpret a further correlation co-efficient (this time demonstrating a non-significant negative correlation) and put both findings together. For full marks, answers should cover the two key bullet points below:

- At age 11, there is a significant positive correlation between happiness and intelligence, demonstrating that more intelligent children tend to be happier
- At age 16, the correlation is not statistically significant.

Students may also make the point that there may be a weak tendency for more intelligent teenagers to be less happy at 16 years of age, although this is not statistically significant. Students may also refer to the contradiction in the results or provide an overall conclusion.

| **AO2 / AO3 Mark bands** |
|---|
| **4 marks Effective**<br>Effective analysis and understanding.<br>The answer includes the findings of the two studies which are expressed clearly and fluently with appropriate reference to intelligence and happiness. Effective use of statistical terminology. |
| **3 marks Reasonable**<br>Reasonable analysis and understanding. The answer is generally focussed and includes reference to both of the key findings which are reasonably clear. There is reasonable use of statistical terminology. |
| **2 marks Basic**<br>Basic, superficial understanding. The answer is sometimes focussed OR covers only one of the key conclusions. Expression of ideas lacks clarity. Limited use of statistical terminology. |
| **1 mark Rudimentary**<br>Rudimentary with very limited understanding.<br>The answer is weak, muddled and may be mainly irrelevant.<br>Deficiency in expression of ideas results in confusion and ambiguity. The answer lacks structure, often merely a series of unconnected assertions. |
| **0 marks**<br>No creditworthy material is presented. |