PSYCHOLOGY

Paper 9990/11
Approaches, Issues and Debates

Key messages

Candidates need to know all components of every core study as listed in the syllabus. Questions can be asked about any part of a core study. Centres are reminded to ensure that they are teaching the correct studies, especially when the psychologist has published many replications, for example, Milgram or Bandura.

Candidates need to read the whole question carefully to ensure that their responses are fulfilling the demands of each one. For example, the question may require data, a named issue to be included or relate back to a previous answer. To achieve full marks, these need to be correctly present in their responses. The essay (final question) requires four evaluation points to be in depth (two strengths and two weaknesses) with at least one of these about the named issue. In depth tends to be having two examples of a particular concept or to support an evaluative point. Credit is limited if the named issue is omitted or just described.

Candidates need to be careful about how they are presenting the results of studies. For example, they need to know if the results are about how many participants performed a task correctly or on how many trials the participant was correct. This can have a large impact on the interpretation of results and whether a response can gain credit. In addition, there must never be any value judgements with results. Results are the factual presentation of data without any interpretation. Therefore, using words like 'better' or 'worse' are not appropriate in relation to the presentation of results.

Candidates also need to engage with any stimulus material presented in a question (for example, a novel situation) to ensure they can access all available marks. In addition, when a question refers to 'in this study' the answer requires contextualisation, with an explicit example from that study.

Candidates also need to know the set procedure of studies in the order presented in the original journal article. Questions can be based around just part of a procedure, and the candidate must be able to produce an answer that is directed and concise, rather than writing about the whole of the procedure. This can sometimes mean a candidate may run out of time for other questions.

Candidates should be able to give full definitions of terms and issues and debates listed in the syllabus and provide full assumptions for all four approaches.

In Section B, the similarity/difference question needs to be based around psychological principles instead of basic analysis or elements of the study that are not showing any psychological knowledge.

There is enough time for answers to be planned to ensure that the response given by a candidate is focused on the demands of each question. This is a crucial skill to develop as some candidates appear to have good knowledge of a study but do not apply this effectively to the question(s) set.

There was a common misconception about a study that restricted the performance of candidates:

- The study by Milgram is not a laboratory experiment**.** There is no IV or DV so arguments about this for **Question 2(b)** could not be awarded credit.
- The study by Perry *et al* was a simulation so no participant had a real ball, stranger etc. invading their personal space.
- The study by Hölzel *et al* is not a laboratory experiment. Participants had to visit a laboratory to have their grey matter measured but the study did not take place in a laboratory. This was presented in **Question 9(b)** but could not be given credit.
- In the study by Bandura *et al* consent was not taken from parents, and the children did not watch a video of the models being aggressive. The models were live.

© 2025

General comments

The marks achieved by the candidates sitting this examination covered a wide spread of possible marks. However, three-quarters of the candidates scored 30 marks or below (and one-quarter scored 15 marks or below). Some candidates provided a range of excellent answers to many of the questions and could explain psychological terminology well, providing evidence that they were prepared for the examination.

Stronger overall responses followed the demands of each question with explicit use of psychological terminology and logical, well-planned answers in evidence. Appropriate examples were used from studies when the question expected it and there was evidence of candidates being able to apply their knowledge to real-world behaviours in terms of what and how.

There were several blank responses in this series (most items had blank responses). As positive marking is used, candidates should attempt all questions even if they are unsure of the response they are providing.

Performance on **Questions 4, 5, 6, 8**, and Section B correlated the strongest with a candidate's overall score.

Comments on specific questions

Question 1

(a) Most of the candidates could identify one of the toys from the wheeled category with popular choices being the wagon or truck.

(b) Stronger responses could clearly define the behaviour of 'drag' as used in the study. Many candidates could communicate the idea of it involving moving the toy along the ground. However, there were a significant amount of candidates who presented tautological answers by stating that to drag 'is to drag' which cannot be given any credit as it does not show understanding. For definitions, candidates need to be able to define the term without using the term.

(c) Only a small minority of responses could clearly outline the term. Common choices included it being about how behaviours can be changed/taught to fit into societal norms. These could be awarded maximum credit. However, most responses provided an 'everyday' definition of the term in relation to socialising with friends and family, but this was not answering the question.

Question 2

(a) Most responses could clearly present a different conclusion to the study. Popular choices included people being much more obedient than expected and that people will harm a stranger if an authority figure tells them to. Some candidates provided a result which could not be given any credit and some re-wrote the conclusion already presented in the question. It is important for candidates to carefully read what the question is asking for and to understand the difference between a result and a conclusion (see Key messages at the beginning of this report).

(b) Stronger responses could clearly identify a correct strength of the study and then use an example to explain why. Common choices included reliability, credibility of the set-up linked to validity and the use of quantitative data.

Question 3

(a) A minority of responses could clearly outline the capture technique. The idea of it being spontaneous was a popular idea presented by candidates. However, there were many incorrect responses presented. These included 'going out and capturing the elephants for the study' and 'the act of the elephant capturing water in its trunks'. It is essential for candidates to know the main mechanisms and behaviours operationalised in all Core Studies. This question has the second highest number of blank responses.

(b) A minority of responses could clearly present a core result in relation to the mean success rate over time. This mainly focused on the mean success rate increasing over time. It is essential for candidates to read the question carefully, so the correct result is then presented by them. Common

© 2025

incorrect results for this question included comparing overall success rates of juveniles versus the adult or focusing on the number of trials. Neither of these could be given credit.

Question 4

Most responses could describe one part of the procedure; this was typically naming one of the approaching figures. Stronger responses could clearly describe the part of the procedure outlined in the question, with popular choices being eight entrances, the participant had to imagine being in the centre of the circle, and that they needed to press the spacebar to stop the animation getting any closer. This question had the highest rate of blank responses in this paper (12 per cent). Common errors included claiming that the ball, stranger etc. were real and invaded personal space, or that the name of the protagonist was presented after the fixation point (it was before). Candidates need to know specific procedures from all Core Studies.

Question 5

(a) Stronger responses could clearly identify two other features of the sample. Popular choices included diagnosed with AS/HFA, volunteer sample and mean IQ. It is important for candidates to read the question carefully as some presented 15 and/or male in their response but these had already been given in the question. The question required them to identify two other features.

(b) Stronger responses could focus on a strength of the revised eyes test. Popular choices included the collection of quantitative data and that the test was an improvement from an original version (with reasons). However, there were many responses that focused on a strength of the study by Baron-Cohen *et al.* which was not answering the question. It is important for candidates to read the question carefully.

Question 6

(a) Most responses could outline one role of social factors, and this tended to be about social pressure. Stronger responses focused on the role of authority figures and how children do not want to 'get into trouble'. Many responses presented information from the study itself, but this was not answering the question. It is important to read the question carefully as it was about the background to the study, not the study itself.

(b) Candidates were equally likely to choose Doris or Tony. For Doris, popular choices included consent and harm. Stronger responses could clearly provide evidence from the study to explain why Doris was correct, for example the parents giving consent for their children and that the children were monitored for stress/made to feel comfortable. For Tony, the most popular choice was psychological harm. As with Doris, stronger responses could clearly provide evidence from the study, for example that the children may have felt too much pressure when deciding if the target was in the line-up or not. There were some responses that attempted to focus on deception, but this guideline was not broken in this study.

Question 7

The average score on this question was under 2 marks. Stronger responses could clearly define the explanations and then provide logical examples from the study by Piliavin *et al*. For the individual explanation a popular example was having an empathetic personality. For the situational explanation, a popular example was condition of the victim and it being in a subway car. There were many answers that provided tautological responses, for example the situation explanation is about what situation you find yourself in. Responses like this cannot be awarded credit as it is not clear if the candidate has simply copied the word from the question. Definitions need to be clear and never use the term within the definition.

Question 8

(a) Stronger responses focused on the situation Pavel was in and gave appropriate advice to help reduce his level of stress. Popular advice included enrolling on a mindfulness course, engaging in yoga-type stretching and engaging in mindfulness at home or when walking. For these types of questions, the advice must be manageable and/or practical. There were several responses advising Pavel to go and have a series of MRI scans to check his progress – this is both impractical, but also the question was asking about reducing his stress levels and an MRI cannot measure that. Candidates need to ensure that they are giving meaningful advice. There were many examples when candidates were justifying why they would give this advice, but this was not

© 2025

answering the question – it is simply about giving advice to Pavel. **Question 8(b)** allowed the candidate to focus on 'why'.

(b) Most responses could provide a basic reason as to why the candidate had provided the advice in **Question 8(a)**. Popular choices were the increasing of grey matter in areas linked to stress management and focusing on one aspect of mindfulness that would help Pavel.

Question 9

(a) Most responses could describe around two aspects of how the EEG was used in the study. Popular choices included the placement of the electrodes, how it could identify what stage of sleep participants were in, and how they were tied into a single cord to not restrict movement. Some candidates focused on the narrative of the dreams provided, or the mechanisms about how the participants were woken up, but these were not answering the question set.

(b) Stronger responses could clearly explain two similarities. Popular choices to compare the studies on included the collection of quantitative data, generalisability of the sample and ethics. To improve responses to this type of question, candidates need to choose comparison points that can be developed and explained, using examples from both studies to explain the similarity and/or difference. For example, explaining the experimental nature of both studies would involve explaining that cause and effect can happen in both studies with examples of controls from both studies to allow cause and effect be stronger. However, stating that each study had a different aim does not allow the response to be detailed so will always only achieve Level 1. Candidates need to choose carefully what the comparisons are, ensuring that they are logical and can be explained fully, using examples from both studies. It is also very important to read the question to see what can or cannot be used on the response. In this case, the candidates were told not to refer to brain measurement techniques, yet several candidates did use brain measurement techniques in their responses and were therefore awarded Level 0. In addition, candidates need to know which studies have been placed in the different approaches. There were many examples of candidates choosing studies from the other three approaches and not biological.

There were also some candidates who responded with one similarity and one difference. It is important for candidates to read the rules of the question.

Question 10

The strongest responses evaluated the study by Bandura *et al* in depth and in terms of two strengths and two weaknesses with at least one of these points covering the named issue of ethics. Common choices included ethics, generalisability, reliability, mundane realism, and quantitative data. These strong responses could explain why an element of the study was a strength or a weakness, using specific examples from the study by Bandura *et al* to explicitly support their point. These answers tended to score Level 5 marks. Candidates need to ensure that they follow the demands of the question, covering two strengths and two weaknesses, all in equal depth. Some responses did cover the four evaluation points but were brief or did not use the study by Bandura *et al* as examples, which meant the response scored in the lower bands. Other responses included three evaluation points that were thorough, logical, and well argued with a fourth point that was not in context which meant it could not be given Level 5. Candidates need to know that any description of the study does not gain credit in these types of questions as it is testing their evaluation skills *only*. There were some common errors evident in responses, for example, consent was taken for the children from their parents and that the children watched videos of aggression. In addition, some responses are **still** following a GRAVE approach to this question (Generalisability, Reliability, Application, Validity, Ethics) when the 'A' part is not creditworthy for an AO3 question. A response that fails to have one evaluation point about the named issue can only score Level 3 (6 marks) maximum. There were many responses that briefly outlined strengths and weaknesses with only some being in context which is a Level 2 response. Any response that has no context cannot get above a Level 1 mark. Some responses did not cover the named issue. To improve on this question, candidates need to plan carefully, choosing two strengths and two weaknesses with one of these being the named issue, avoiding real world application where possible. Each strength and weakness should be of equal length with an explanation as to why it is a strength or weakness with examples (plural) from the study to show clear understanding. An evaluation that is in depth tends to have at least two explicit examples from the named study for each evaluative point made. These are the requirements for a Level 5 response. The average response was Level 2 (3 marks) for this cohort.

© 2025

PSYCHOLOGY

Paper 9990/12
Approaches, Issues and Debates

Key messages

Candidates need to know all components of every core study as listed in the syllabus. Questions can be asked about any part of a core study. Centres are reminded to ensure that they are teaching the correct studies, especially when the psychologist has published many replications, for example, Milgram or Bandura.

Candidates need to read the whole question carefully to ensure that their responses are fulfilling the demands of each one. For example, the question may require data, a named issue to be included or relate back to a previous answer. To achieve full marks, these need to be correctly present in their responses. The essay (final question) requires four evaluation points to be in depth (two strengths and two weaknesses) with at least one of these about the named issue. In depth tends to be having two examples of a particular concept or to support an evaluative point. Credit is limited if the named issue is omitted or just described.

Candidates need to be careful about how they are presenting the results of studies. For example, they need to know if the results are about how many participants performed a task correctly or on how many trials the participant was correct. This can have a large impact on the interpretation of results and whether a response can gain credit. In addition, there must never be any value judgements with results. Results are the factual presentation of data without any interpretation. Therefore, using words like 'better' or 'worse' are not appropriate in relation to the presentation of results.

Candidates also need to engage with any stimulus material presented in a question (for example, a novel situation) to ensure they can access all available marks. In addition, when a question refers to 'in this study' the answer requires contextualisation with an explicit example from that study.

Candidates also need to know the set procedure of studies in the order presented in the original journal article. Questions can be based around just part of a procedure, and the candidate must be able to produce an answer that is directed and concise rather than writing about the whole of the procedure. This can sometimes mean a candidate may run out of time for other questions.

Candidates should be able to give full definitions of terms and issues and debates listed in the syllabus and provide full assumptions for all four approaches.

In Section B, the similarity/difference question needs to be based around psychological principles instead of basic analysis or elements of the study that are not showing any psychological knowledge.

There is enough time for answers to be planned to ensure that the response given by a candidate is focused on the demands of each question. This is a crucial skill to develop as some candidates appear to have good knowledge of a study but do not apply this effectively to the question(s) set.

There were common misconceptions about certain studies that restricted the performance of candidates:

1. For the study by Perry *et al.* in relation to ethics: Stating that the participants did not interact with someone later in Experiment 2 meant the lack of deception guideline had been broken; this by itself is not breaking the guideline. However, when linked to the possibility of compromising harm then it is breaking the guideline. Withholding information by itself is not breaking any guideline, it is the potential consequence that is. This is clear in the second example, when candidates claim the participant not knowing if they were given placebo or oxytocin, meant the guideline of deception as broken. This is not breaking the guideline.

2. The study by Perry *et al.* was a simulation so no participant had a real ball, stranger etc. invading their personal space.

© 2025

3. The study by Saavedra and Silverman is not experimental.

4. The study by Pozzulo *et al.* did not involve witnessing a real crime, did not use a real-life line-up for children to look at and it did not take place in a police station.

General comments

There was a wide range of marks achieved by candidates in this examination. However, half of the candidates scored 29 marks or above (and one-quarter scored 19 marks or below). Some candidates provided a range of excellent answers to many of the questions and could explain psychological terminology well, providing evidence that they were prepared for the examination.

Stronger overall responses followed the demands of each question with explicit use of psychological terminology and logical, well-planned answers in evidence. Appropriate examples were used from studies when the question expected it and there was evidence of candidates being able to apply their knowledge to real-world behaviours, in terms of what and how.

There were several blank responses in this series (most items had blank responses). As positive marking is used, candidates should attempt all questions even if they are unsure of the response they are providing.

Performance on questions 4, 7, and Section B correlated the strongest with a candidate's overall score.

Comments on specific questions

Question 1

(a) Most of the candidates could not state the number of participants. Common errors included stating the percentage of participants or stating all participants.

(b) Most candidates could identify two signs of tension, with popular choices being sweating, trembling and stuttering. Some candidates repeated the example in the question. It is essential for candidates to read the question carefully to ensure they are answering the question set. Some candidates listed behaviours that might show signs of tension but were not those recorded in the study by Milgram.

(c) Only a minority of responses could clearly outline what the candidates were asked to predict. These responses tended to focus on predicting the behaviour of actual participants and what percentage would press the 450v switch. However, many responses tended to outline what was predicted rather than what they were asked to predict. That is, outlining that they thought up to 3 per cent might press the switch. This is not answering the question. It is essential for candidates to read the question carefully to ensure they are focusing their responses on the demands of the question.

Question 2

(a) Most responses could clearly present a different conclusion, briefly, to the study. Popular choices included repeating observed aggression when the model is no longer there or that children are more likely to imitate the aggression of a same-sex model. Some candidates provided a result which could not be given any credit and some re-wrote the conclusion already presented in the question. It is essential for candidates to carefully read what the question is asking for and to understand the difference between a result and a conclusion (see Key messages at the beginning of this report).

(b) Stronger responses could clearly identify a correct weakness of the study in relation to the sample of participants and then use an example to explain why. The most popular choice was about the generalisability of the sample due to its size and/or demographics.

© 2025

Question 3

(a) A minority of responses could clearly identify two features. Popular choices included being white and male. Several candidates identified features of the victim even when the question clearly differentiated between the victim and the model. Some candidates also presented information about the sample of participants which was not answering the question. Also, some responses focused on what was expected of the models across conditions and again this was not answering the question. It is essential for candidates to read the question carefully to ensure they are focusing their responses on the demands of the question.

(b) A minority of responses could clearly present a core result in relation to the victim acting drunk before the model helped. This mainly focused on racial differences. It is essential for candidates to read the question carefully, so the correct result is then presented by them. Common incorrect results for this question included focusing on the ill victim or behaviours that happened after the model had intervened.

Question 4

Most responses could describe at least two ways ethical guidelines were followed. This was typically focusing on housing and diet. Stronger responses could clearly describe, using evidence from the study, how housing was very good, how there were measures to reduce stress / harm and how their diet remained unchanged during the study period. A significant number of responses simply identified and then generically described a range of ethical guidelines in relation to animals, but never then linked these to describe how Hassett *et al.* followed each guideline, limiting the credit to one mark overall.

Question 5

(a) Stronger responses could clearly define the factor of 'acting with awareness' as used in the study. A small proportion of candidates could communicate the idea of it being focused on current behaviour and actions, not behaving 'automatically'. However, there were a significant amount of candidates who presented tautological answers by stating that by acting with awareness 'is to be aware of your situation', which cannot be given any credit as it does not show understanding. For definitions, candidates need to be able to define the term without using the term. This question had the joint highest number of blank responses on this paper.

(b) Stronger responses could focus on a strength of the Five Facet Mindfulness Questionnaire. Popular choices included the collection of quantitative data for before and after comparisons, and that it was a standardised test. However, there were many responses that focused on a strength of the study by Hölzel *et al.* which was not answering the question. It is important for candidates to read the question carefully.

Question 6

(a) Most responses could outline one role of biological factors, and this tended to be about social hormones, in particular, oxytocin. Stronger responses focused on the role of oxytocin in social bonding/empathy and explained the role of the amygdala. Many responses presented information from the study itself, but this was not answering the question. It is important to read the question carefully as it was about the background to the study, not the study itself.

(b) For Alexandre, popular choices included (informed) consent and reducing the amount of psychological harm. Stronger responses could then present an example of this from the study. For example, written consent was taken from all participants, and it was a simulation, so there was no direct invasion of personal space. For Derek, the most popular choice was psychological harm, as even though the study was a simulation, participants were still expected to imagine the invasion of personal space. There were several misconceptions presented by candidates linked to their responses for Derek and these are presented in the Key messages section of this report.

Question 7

The average score on this question was under 2 marks. Stronger responses could clearly define the explanations and then provide logical examples from the study by Dement and Kleitman. For the individual explanation, a popular example was the different dreams reported by the participants. For the situational explanation, a popular example was the EEG/laboratory set-up changing the sleeping/dreaming habits of the

© 2025

participants. There were many answers that provided tautological responses, for example the situation explanation is about what situation you find yourself in. Responses like this cannot be awarded credit as it is not clear if the candidate has simply copied the word from the question. Definitions need to be clear and never use the term within the definition. This question had the joint highest number of blank responses on this paper.

Question 8

(a) Stronger responses focused on the situation Carole was in and gave appropriate advice to help her improve her candidates' concentration levels. Popular advice included giving candidates paper with printed shapes on it, allowing them to shade the shapes on the paper and not worrying about how neat it was. For these types of questions, the advice must be manageable and/or practical. There were several responses where the advice to Carole was to doodle. This is not practical as it does not tell Carole explicitly what to do and does not show understanding as 'doodling' was in the question. Candidates need to ensure that they are giving meaningful advice. There were many examples when candidates were justifying why they would give this advice, but this was not answering the question – it is simply about giving advice to Carole. **Question 8(b)** allowed the candidate to focus on 'why'.

(b) Most responses could provide a detailed reason as to why the candidate had provided the advice in **Question 8(a)**. Popular choices were the findings from the study by Andrade showing that recall improved and how doodling can reduce daydreaming/improves concentration, which can then impact on memory recall.

Question 9

(a) Most responses could describe at least three aspects of the participant used in the study. Popular choices included him being male, his age and when his phobia first began. Common information provided that was not answering the question was what the boy did during therapy and the long-term effects on him of that therapy.

(b) Stronger responses could clearly explain one similarity and/or one difference. Popular choices to compare the studies on included the collection of quantitative data, generalisability of the sample and ethics. To improve responses to this type of question, candidates need to choose comparison points that can be developed and explained, using examples from both studies to explain the similarity and/or difference. Stating that each study had a different aim, or both used children, does not allow the response to be detailed so will always only achieve Level 1. Candidates need to choose carefully what the comparisons are, ensuring that they are logical and can be explained fully, using examples from both studies. For this specific comparison, there were a significant number of responses attempting to compare the two studies on their experimental design, which could only be awarded Level 0, as Saavedra is not experimental. It is also very important to read the question to see what can or cannot be used on the response. In this case, the candidates were told not to refer to sampling technique, yet some candidates did in their responses and were therefore awarded Level 0.

Question 10

The strongest responses evaluated the study by Pozzulo *et al* in depth and in terms of two strengths and two weaknesses, with at least one of these points covering the named issue of ecological validity. Common choices included ethics, generalisability, reliability, ecological validity, and quantitative data. These strong responses could explain why an element of the study was a strength or a weakness, using specific examples from the study by Pozzulo *et al.* to explicitly support their point. These answers tended to score Level 5 marks. Candidates need to ensure that they follow the demands of the question, covering two strengths and two weaknesses, all in equal depth. Some responses did cover the four evaluation points but were brief or did not use the study by Pozzulo *et al* as examples, which meant the response scored in the lower bands. Other responses included three evaluation points that were thorough, logical, and well argued with a fourth point that was not in context, which meant it could not be given Level 5. Candidates need to know that any description of the study does not gain credit in these types of questions as it is testing their evaluation skills only. There were some common errors evident in responses, for example, that real crimes were witnessed and children had to go to a real police station to choose the criminal. In addition, some responses are still following a GRAVE approach to this question (Generalisability, Reliability, Application, Validity, Ethics) when the 'A' part is not creditworthy for an AO3 question. A response that fails to have one evaluation point about the named issue can only score Level 3 (6 marks) maximum. There were many responses that briefly

© 2025

outlined strengths and weaknesses, with only some being in context, which is a Level 2 response. Any response that has no context cannot get above a Level 1 mark. Some responses did not cover the named issue. To improve on this question, candidates need to plan carefully, choosing two strengths and two weaknesses, with one of these being the named issue, avoiding real world application where possible. Each strength and weakness should be of equal length with an explanation as to why it is a strength or weakness, with examples (plural) from the study, to show clear understanding. An evaluation that is in depth tends to have at least two explicit examples from the named study for each evaluative point made. These are the requirements for a Level 5 response. The average response was Level 2 (4 marks) for this cohort.

© 2025

PSYCHOLOGY

Paper 9990/13
Approaches, Issues and Debates

Key messages

Candidates need to know all components of every core study as listed in the syllabus. Questions can be asked about any part of a core study. Centres are reminded to ensure that they are teaching the correct studies, especially when the psychologist has published many studies, for example, Milgram or Bandura.

Candidates need to read the whole question carefully to ensure that their responses are fulfilling the demands of each one. For example, the question may require data, a named issue to be included or relate back to a previous answer. To achieve full marks, these need to be correctly present in their responses. The essay (final question) requires four evaluation points to be in depth (two strengths and two weaknesses) with at least one of these about the named issue. In depth tends to be having two examples of a particular concept or to support an evaluative point. Credit is limited if the named issue is omitted or just described.

Candidates need to be careful about how they are presenting the results of studies. For example, they need to know if the results are about how many participants performed a task correctly or on how many trials the participant was correct. This can have a large impact on the interpretation of results and whether a response can gain credit. In addition, there must never be any value judgements with results. Results are the factual presentation of data without any interpretation. Therefore, using words like 'better' or 'worse' are not appropriate in relation to the presentation of results.

Candidates also need to engage with any stimulus material presented in a question (for example, a novel situation) to ensure they can access all available marks. In addition, when a question refers to 'in this study' the answer requires contextualisation with an explicit example from that study.

Candidates also need to know the set procedure of studies in the order presented in the original journal article. Questions can be based around just *part* of a procedure, and the candidate must be able to produce an answer that is directed and concise rather than writing about the whole of the procedure. This can sometimes mean a candidate may run out of time for other questions.

Candidates should be able to give full definitions of terms and issues and debates listed in the syllabus and provide full assumptions for all four approaches.

In ***Section B***, the similarity/difference question needs to be based around psychological principles instead of basic analysis or elements of the study that are not showing any psychological knowledge.

There is enough time for answers to be planned to ensure that the response given by a candidate is focused on the demands of each question. This is a crucial skill to develop as some candidates appear to have good knowledge of a study but do not apply this effectively to the question(s) set.

There were common misconceptions about certain studies that restricted the performance of candidates:

1. The sampling technique used in the study by Andrade is opportunity and not volunteer.
2. The study by Fagen *et al* did not use any classical conditioning.
3. Andrade was unethical and used deception as she did not tell them they would be having to recall names/places. This does not break the ethical guideline of lack of deception. It is an essential feature of the study and participants knew afterwards why it had been done and there is nothing reported about an adverse effect.
4. The study by Hölzel *et al* is not a laboratory experiment. Participants had to visit a laboratory to have their grey matter measured but the study did not take place in a laboratory. This was presented in Question 10 but could not be given credit.

© 2025

General comments

The marks achieved by the candidates sitting this examination covered a wide spread of possible marks. However, half of the candidates scored 31 marks or above (and one-quarter scored 24 marks or below). Some candidates provided a range of excellent answers to many of the questions and could explain psychological terminology well, providing evidence that they were prepared for the examination.

Stronger overall responses followed the demands of each question with explicit use of psychological terminology and logical, well-planned answers in evidence. Appropriate examples were used from studies when the question expected it and there was evidence of candidates being able to apply their knowledge to real-world behaviours in terms of what and how.

There were some blank responses in this series but far fewer than in previous series. As positive marking is used, candidates should attempt all questions even if they are unsure of the response they are providing.

Performance on **Questions 1, 5**, and ***Section B*** correlated the strongest with a candidate's overall score.

Comments on specific questions

Question 1

(a) Most of the candidates could clearly outline how the participants were recruited. Popular choices included it being via opportunity sampling and that they had just completed an unrelated experiment before being asked. A common error was claiming Andrade used volunteer sampling.

(b) Most candidates could identify two features of the material. Common choices included the size of the shapes and that they were circles and squares. It is important for candidates to read the question carefully as some described what the doodling group were expected to do with the paper during the study, which was not answering the question.

(c) Only a very small minority of candidates could identify how recall was measured. Candidates tended to focus on the recall of other types of information or how scores were calculated – neither were answering the question. This item had the joint highest number of blank responses on the paper.

Question 2

(a) Most responses could clearly present a different conclusion, briefly, to the study. Popular choices included about the status of the victim or how diffusion of responsibility was not seen. Some candidates provided a result which could not be given any credit and some re-wrote the conclusion already presented in the question. It is essential for candidates to carefully read what the question is asking for and to understand the difference between a result and a conclusion (see Key messages at the beginning of this report).

(b) Stronger responses could clearly identify a correct weakness of the study in relation to ethics and then use an example to explain why. The most popular choice was psychological harm of witnessing the collapse.

Question 3

(a) A slight majority of responses could clearly identify two features. Popular choices included the troop size, where they were based (Yerkes) and the final sample size used in the data analysis phase. Some candidates also presented information about what the monkeys did in the study, but this was not answering the question. It is essential for candidates to read the question carefully to ensure they are focusing their responses on the demands of the question.

(b) Most responses could clearly present a core result in relation to the toy preferences of female monkeys. This mainly focused on duration or frequency of play across toy type. It is essential for candidates to read the question carefully, so the correct result is then presented by them. Common incorrect results for this question included focusing on male monkey preferences.

© 2025

Question 4

Most responses could describe at least one aspect of how the cartoon foils were chosen. This was typically focusing on them being readily available and having to be of a similar appearance to the target cartoon characters. Stronger responses could clearly describe the entire procedure involved in selecting the cartoon foil. However, many candidates described what the pictures looked like or described how the participants were asked to choose a target, with one of these cartoon foils in the line-up. Both are not answering the question. It is important for candidates to carefully read the question to ensure their response is focused on its demands. This item had the joint highest number of blank responses on the paper.

Question 5

(a) A slight majority of candidates could correctly identify two features of the sample. Popular choices included the sample size, which university they were studying at, and that they were volunteers. Some candidates repeated the 'male' feature already presented in the question and could not be awarded credit for that. For this type of question, candidates cannot repeat a feature or features already presented so it is essential to read questions carefully.

(b) Stronger responses could focus on a weakness of using animations. Popular choices included the lack of realism and that it is not a true measure of personal space invasion. However, there were some responses that focused on a weakness of the study by Perry *et al.* which was not answering the question. It is important for candidates to read the question carefully.

Question 6

(a) Most responses could outline one role of disgust in childhood phobias, and this tended to be about aversive feelings or being repulsed. Stronger responses focused on evaluative learning in relation to anticipated threats and that it can usually stem from a childhood trauma or event. Many responses presented information from the study itself, but this was not answering the question. This item had the joint highest number of blank responses on the paper.

(b) Candidates were much more likely to choose Josh rather than Hannah. For Josh, popular choices included the sample size and type of phobia. Stronger responses could clearly provide evidence from the study to explain why Josh was correct, for example the hierarchy being unique to him or that the ratings were subjective and very difficult to generalise to other people, even those with button phobia. For Hannah, the most popular choice was the use of positive reinforcement from her mother. Like with Josh, stronger responses could clearly provide evidence from the study, for example the Feeling Thermometer itself could be useful with other patients.

Question 7

The average score on this question was over 2 marks. Stronger responses could clearly define the debate and then provide logical examples from the study by Bandura *et al*. For the nature side of the debate, a popular example was the role of testosterone as a way of explaining males being more aggressive. However, some candidates attempted to link gender into the argument which could not be awarded credit for nature as gender is psychological, whereas sex is biological. For the nurture side of the debate, a popular example was the role of social learning through observation and imitation, in aggressive behaviour in this study. There were many responses that provided tautological responses, for example the nature side of the debate is about 'nature'. Responses like this cannot be awarded credit as it is not clear if the candidate has simply copied the word from the question. Definitions need to be clear and never use the term within the definition.

Question 8

(a) Stronger responses focused on the situation Pauline was in and gave appropriate advice to help her candidates learn the new skill. Popular advice included the use of a reward-based system, regular primary reinforcers being used and a consistent way of chaining a sequence of 'skills' together. For these types of questions, the advice must be manageable and/or practical. There were several responses where the advice to Pauline was to simply 'teach them'. This is not practical as it does not tell Pauline explicitly what to do and does not show understanding as Pauline being a teacher was in the questions. Candidates need to ensure that they are giving meaningful advice. There were many examples when candidates were justifying why they would give this advice, but this was not answering the question – it is simply about giving advice to Pauline. **Question 8(b)** allowed the candidate to focus on 'why'. Also, there were some candidates

© 2025

attempting to use social learning and/or classical conditioning which are both not part of the study by Fagen *et al* so cannot be awarded credit.

(b) Most responses could provide a brief reason as to why the candidate had provided the advice in **Question 8(a)**. Popular choices were the findings from the study by Fagen *et al* showing that shaping and rewarding a set of smaller skills can then be 'put together' to learn a full new skill, as in the scenarios for **Question 8(a)**.

Question 9

(a) Most responses could describe at least one aspect of the psychology being investigated. Popular choices included what theory of mind predicts and social sensitivity. Common information provided that was not answering the question, was the findings from the study by Baron-Cohen *et al* which could only be awarded one 'example mark' overall. For this type of question, candidates need to focus on the generic ideas that are being investigated in each of the Core Studies, and these are all listed in the syllabus.

(b) Stronger responses could clearly explain two similarities. Popular choices to compare the studies on included the collection of quantitative data, generalisability of the sample and lack of mundane realism. To improve responses to this type of question, candidates need to choose comparison points that can be developed and explained, using examples from both studies to explain the similarity and/or difference. For example, explaining the experimental nature of both studies would involve discussing that cause and effect can happen in both studies, with examples of controls from both studies to allow cause and effect be stronger. However, stating that each study used humans does not allow the response to be detailed so will always only achieve Level 1. Candidates need to choose carefully what the comparisons are, ensuring that they are logical and can be explained fully, using examples from both studies. Candidates who attempted to show a similarity on ethics tended to struggle, typically due to a misconception of 'deception' in relation to the study by Andrade. See the Key messages section of this report.

There were also some candidates who responded with one similarity and one difference. It is important for candidates to read the rules of the question.

Question 10

The strongest responses evaluated the study by Hölzel *et al* in depth and in terms of two strengths and two weaknesses, with at least one of these points covering the named issue of the use of self-reports. Common choices included ethics, generalisability, reliability, self-report data, and quantitative data. These strong responses could explain why an element of the study was a strength or a weakness using specific examples from the study by Hölzel *et al* to explicitly support their point. These answers tended to score Level 5 marks. Candidates need to ensure that they follow the demands of the question, covering two strengths and two weaknesses, all in equal depth. Some responses did cover the four evaluation points but were brief or did not use the study by Hölzel *et al* as examples, which meant the response scored in the lower bands. Other responses included three evaluation points that were thorough, logical, and well argued with a fourth point that was not in context which meant it could not be given Level 5. Candidates need to know that any description of the study does not gain credit in these types of questions as it is testing their evaluation skills only. There were some common errors evident in responses, for example, that the study was a laboratory experiment. In addition, some responses are still following a GRAVE approach to this question (Generalisability, Reliability, Application, Validity, Ethics) when the 'A' part is not creditworthy for an AO3 question. A response that fails to have one evaluation point about the named issue can only score Level 3 (6 marks) maximum. There were many responses that briefly outlined strengths and weaknesses with only some being in context, which is a Level 2 response. Any response that has no context cannot get above a Level 1 mark. Some responses did not cover the named issue. To improve on this question, candidates need to plan carefully, choosing two strengths and two weaknesses, with one of these being the named issue, avoiding real world application where possible. Each strength and weakness should be of equal length with an explanation as to why it is a strength or weakness with examples (plural) from the study to show clear understanding. An evaluation that is in depth tends to have at least two explicit examples from the named study for each evaluative point made. These are the requirements for a Level 5 response. The average response was just below Level 3 (5 marks) for this cohort.

© 2025

PSYCHOLOGY

Paper 9990/21
Research Methods

Key messages

- Candidates need to ensure that when providing suggestions for a piece of research, that they are specific, and relevant to the research method given. For example, if it is an observational method, then trying to observe internal states such as 'boredom' or 'tiredness' would not gain credit unless they were operationalised, i.e. the number of times they yawn for tiredness. Too often suggestions are vague and cannot be credited.
- Candidates need to ensure that they still have good knowledge of the studies on the specification for Paper 2. There were several gaps in knowledge seen on the paper for this series.
- Candidates need to ensure that they use the information given in the scenario to answer questions. This is not just about candidates repeating information from the scenario in the question but using the information to inform their answer. For example, in **Question 7** the scenario gave candidates the knowledge that the participant does not like people close to them. There is always a reason for this information, therefore it needs to be used to inform an answer about a particular research method, i.e. why a questionnaire is better than an interview in this case. Generic answers which do not use the scenario, or just repeat it, will lose marks.
- Candidates need to ensure that their responses are focused on the questions within the exam paper. There was more than one question where it was clear that candidates had misread the question and provided responses which were not creditworthy.
- Candidates need to ensure that they understand the expectations for different command words used on the paper; for example, describe and explain. For describe, candidates need to ensure that they provide a sufficient number of unique points related to the marks allocated to the question, whereas for explain, candidates need to identify a particular feature/concept/theory and then link their detailed point to the feature/concept/theory they have identified. Often, explain questions had poorer outcomes in terms of marks as a candidate just described.
- It is worth noting that many candidate responses for the 6-mark question showed good knowledge and understanding of measures of central tendency. Candidates need to make sure they provide appropriate examples within their response to ensure they can access all 6 marks, and do not provide tautological answers. For this question in particular, candidates need to make sure that examples were not just numerical examples of how the mean is worked out. This will not gain credit; it needs to be a study example, whether named or one the candidate has 'made up' themselves (as long as relevant).
- It is worth noting, that candidate performance on the 10-mark extended response question asking candidates to describe an interview showed, once again, an improvement from previous series. There were some thoughtful responses which clearly had good understanding of the requirements of an interview but also were very much aware of replicability. For this, candidates should be commended.

General comments

This was the second November series for the new Psychology specification and once again significant improvements across many areas were seen. Paper 21 scripts provided the full range of marks, showing a good level of knowledge and understanding across most areas of the specification. Where performance was limited, it was due to a lack of knowledge of key studies, or a misunderstanding of the demands of the question. This was clear when looking at questions on Saavedra and Silverman (**2(a)** and **(b)**), Holzel *et al*. (**3(a)(b)**) and animal ethical guidelines (**9(c)(ii)** and **9(c)(iii)**). Candidates did show their ability to structure 10-mark responses (**10(a)**) with many able to produce thoughtful procedures which incorporated all the elements required within the question. It is clear that centres have prepared candidates well for the extended responses in the exam. For future series, candidates need to ensure that they have a good understanding of command words, key research method terminology such as validity and reliability, and the studies which have been named on the specification.

© 2025

Comments on specific questions

Question 1

(a) Most candidates achieved the one mark available on this question. Where candidate responses did not achieve the mark, it was not due to an error in the type of observation chosen, rather the candidate only stated it was a structured observation rather than explaining it. Candidates need to remember that an explain question needs more than just identifying the method but also explaining why this method was chosen.

(b) (i) To achieve the two marks, candidate responses needed to ensure that the categories suggested were both observable and measurable. Candidates struggled slightly with this question. Many candidate responses used internal states such as 'boredom' or 'tiredness' which, on their own, cannot be measured. To achieve the marks for this type of category, operationalisation needed to be present, i.e. 'number of times student yawned'.

Where candidates did achieve marks, the most common responses included 'using music whilst studying (or not)' and the use of revision methods such as flashcards/notes, etc. As stated in the key comments above, it is vital that candidates ensure that what they suggest within a research method question can be carried out.

(ii) Unfortunately, the number of marks obtained by candidates on this question is directly related to their performance on **part (b)(i)**, with it being more likely a candidate would get 0 marks if they did not achieve any marks in **part (b)(i)**. Conversely, where candidate did achieve both marks in **part (b)(i)** they were likely to achieve 1 or 2 marks on this part of the question.

Where candidates did achieve one or two marks available, the most common responses included:

- The category is difficult to operationalise properly, so it may be that the researcher will miss some occasions it occurs/will overcount when observing (this would be 1 as no link to the specific suggestion).
- Studying with a group is difficult to operationalise properly as it may be that they are talking about other things or do not work, so it may be that the researcher will miss some occasions it occurs/will overcount when observing (this would get 2 as clear link to suggestion).

Question 2

(a) To achieve both marks on this question, candidates needed to get all three timeframes correct; to get one-mark candidates needed to get two of the time frames correct. If candidates only got one time frame correct, they would achieve 0 marks.

Candidates struggled with this question. More often than not, candidates would suggest that the timeframes were between sessions (such as at the start, after 6 months, and after 12 months) which would have gained a tick for the first time frame, but no others, meaning no marks was achieved.

There were a minority of candidates who were able to achieve at least one mark but only very few were able to achieve the 2 marks available. Candidates need to be aware that any of the finer details from the key studies can be asked in the examination.

(b) The majority of candidates did achieve the mark for this question, however a minority either suggested that the severity rating increased or did not make it explicit that it decreased; with some just talking about the number of buttons seen and used which was not enough for the mark. As long as the candidate indicated clearly that the severity decreased in this time frame, then the mark would be given.

Question 3

(a) This two-mark question asked candidates to outline two inclusion criteria in the Holzel *et al.* study, other than wanting to reduce their stress and being able to have an MRi. For this question,

© 2025

suggesting what candidates could not have was also credited. The most common responses seen were:

- Not taking any medication.
- That they are healthy
- Very few meditation classes in the last year
- Say that they will do all the homework.

Where two marks were not achieved it was often due to errors such as:

- Saying just that they could not have done any mediation classes (this is not correct)
- Saying they need to be of a specific age without stating the ages.

One interesting response is where candidates say that they were all right-handed. Although it is not clear in the journal whether that was a conscious decision, we can assume it was, so this was credited.

(b) How well candidates did on this question was directly related to the responses that they gave in **3(a)**. Unfortunately, if candidates did not achieve any marks previously, it is extremely unlikely they would achieve any in this question.

Where candidates did achieve the marks the most common responses were:

- The idea that taking mindfulness classes prior to the study could result in brain changes/symptom reduction without the mindfulness.
- Younger/older adults may have brain impairments.

The generic response 'to ensure the IV was the reason for the change in the DV' was credited, but only once.

Where candidates did not achieve the two marks it was either due to incorrect responses from **3(a)** being discussed, repeating the same reason for the different inclusion criteria, or giving two very generic responses.

Question 4

(a) Candidate performance on this question was pleasing. There was no need to link the response to the Baron-Cohen study although many candidates did so, and candidates performed well when this was the case. The majority of candidates could give an appropriate strength, although less were able to elaborate enough to get the second mark.

The most common responses seen are:

- Less chance of demand characteristics (1) because they only see one level of the IV so will not find out the aim (0)
- Less chance of order effects such as fatigue/boredom effects (1) so their performance will not worsen due to the repeated nature of the tasks (1).

(b) Candidates found the weakness a little more challenging but still many gained both marks. By far the most common response seen was the idea of participant variables being a confounding variable in the study. The issue came with then providing enough detail, i.e.: That this factor will affect the results (meaning the independent variable is not the only reason for the results). This meant that many candidates achieved one mark when using this response. Candidates also often used a relevant example here to elaborate on their point which was acceptable, and as in the previous question, where candidates did this, they often would get the second mark.

One thing to note is that having to get a bigger sample is creditworthy, even if it is basic. Students found it difficult to elaborate on this to gain the second mark, as they would need to say about funding/difficult to find samples and very few did so.

Question 5

© 2025

The two-mark question asked candidates to sketch a graph to show a negative correlation between time spent watching television and attention span. The majority of candidates achieved the two marks on this question. Where candidates did not achieve the marks, it was due to the wrong type of graph being sketched (such as a bar chart), or a positive correlation being sketched.

Question 6

The majority of candidates could name at least one or two of the measures, and very often could describe how they could be worked out. This meant that most candidates achieved at least 1 mark for this question, with often 2 or 3 being the most common mark achieved. In terms of common errors these tended to be:

- Misidentifying the range and/or standard deviation as measures of central tendency
- Giving numerical answers to elaborate on the way that you work out the measures and seeing these as examples.

Neither of the above could gain credit. For the numerical examples, these are just visual representations of what candidates had often already said. One exception is if the candidates had not written how they could be worked out, then they could gain a mark for a numerical example if correct.

To achieve marks for examples, candidates needed to either create their own piece of research and then talk about how a mean/median/mode could be used, or use any of the key studies from the specification and highlight how measures of central tendency were used. Unfortunately, some candidates did not do this and therefore could only achieve a maximum of 4 marks, which was a shame as many showed very good knowledge and understanding.

Question 7

(a) This one- mark question asked candidates to why Dennis may prefer Professor March to use a questionnaire rather than a face-to-face interview. To achieve the mark, candidates needed to use the scenario to make an appropriate suggestion.

The most important part of the scenario talked about Dennis' need for personal space and therefore a questionnaire would be better as it could; guarantee a greater distance between the researcher and the participant; it would not need to be face to face; the researcher won't need to be too close or so that Dennis does not get distressed at someone a bit closer.

Candidate performance on this question was below what was expected. Candidates tended to focus on the research method itself rather than the characteristics of the participant. This was only creditworthy if there was specific focus on the honesty of Dennis or if he was on his own and not pressured by the researcher, but very few did so. This meant that most candidates did not achieve the mark for this question.

(b) This two-mark question asked candidates to suggest two examples of questions that Professor March could use to collect data about the need Dennis has for a large interpersonal distance. Candidate performance on this question was pleasing with the majority of candidates able to achieve the marks available.

(c) Candidates struggled with this question, as they really did not pick up on the fact that they needed to talk about the unapproachable/need for distance aspect of the study. Again, there was a focus on questionnaires as a research method and specifically the type of data gained. Many candidates talked about only being able to gain quantitative data (which is not true), or they would not be able to explain their answers (which again is not totally correct). Responses which were creditworthy would include:

- The participants do not want to be handed a questionnaire/be given paper for the CIDs etc. (but as they said in the previous question that it allows separation, many did not use this type of response).
- The idea that the separation between Dennis and the researcher will cause an issue as there may be more of a likelihood that they will not give honest answers due the sensitivity of the topic area (personal space) or the researcher cannot observe body language

© 2025

More candidates did opt for the second bullet point, but the detail given about this was minimal and therefore candidates would often get 1 mark for this question, which was disappointing. Candidates need to ensure they use the information in the scenario to answer the question.

© 2025

Question 8

(a) Candidates performed well on this question with the vast majority showing they understood the strengths of participant observation. Where candidate responses did not achieve the two marks available, it was due to the lack of reference to the study, and to play. For example:

- She can get closer to the children so will know in more depth what they are doing (this was 1 mark as no reference to play)
- She can get closer to the children/talk to the children so will know in more depth how they are playing (this was 2 marks as there is a reference to play).

Candidates need to ensure that they make specific reference to the scenario provided to ensure that they gain the full marks for a question.

(b) As in **part (a)**, candidates performed well on this question. Most were able to state an advantage and so gained at least one mark. What is required is a response that is specific about the distance between the researcher and the children, such as being less likely to be distracted by the researcher, so linked to more natural behaviour.

The most common responses seen were:

- The children would not be aware she is there
- She could use a video to record the behaviour of the children
- She coiled use a second observer.

Similar to the previous question, candidates needed the link to play/the study to achieve the second mark, therefore:

- She could use a video to record their behaviour so she will be able to analyse their behaviour accurately (1 mark as no link to play)
- She could use a video to record their behaviour which can be used to accurately analyse the children's play (2 marks as linked to play)

This lack of a link was the most common reason by far, as to why candidates did not achieve the second mark.

(c) Candidate performance on this question was pleasing, with most candidates achieving the two marks available. By far the most common response was the idea of deception. If candidates talked about being misled/lied to, then this was used as words to that effect for deception. Most candidates who identified deception as an issue were able to then go on and provide relevant detail i.e. because she was not a playground assistant/ that this would mean that the children could not give informed consent. There were a few candidates who did not provide the required detail for the second mark, but they were in the minority.

Question 9

(a) The most important aspect of this question is the command word; explain. This meant that if candidates only said independent measures, then they would not gain the mark. The candidate needed to also explain why it was independent measures i.e. they only took part in the noise/no noise condition. Unfortunately, a number of candidates did not explain their choice and therefore did not gain the mark available.

(b) The vast majority of candidate responses were able to achieve at least one mark for this question. Where performance was limited, it was more than often due to the candidates getting the independent variable and the dependent variable the wrong way round. This was the minority of candidates though, with most showing good knowledge of the different terms.

(c) (i) This one-mark question asked candidates to suggest one way that the animals could be reinforced. The majority of candidate responses were able to achieve the mark for this question. Where performance was limited, it was more than often due to the candidates just stating a specific type of

© 2025

reinforcement i.e. positive reinforcement, rather than how this type of reinforcement could be used with the animals.

(ii) Candidates found this question challenging with the majority of candidates not achieving the marks available. It was important that candidates used the information given in the stem to answer this question. Candidates were told that the animals would be going into narrow tunnels, and that they would meet other animals, therefore the choice of animal needed to be appropriate to these bits of information. Therefore:

- The animal needed to be small enough to fit into the tunnels
- The animals needed to be social enough not to be concerned about meeting other animals in the tunnel.

Unfortunately, although candidates did show knowledge of species, they concentrated on generic issues related to species i.e. using the least amount possible/those that were not endangered, which did not relate to the scenario at all; consequently, achieving none of the marks available.

(iii) Candidates performed fairly well on this question; by far the most common responses were:

- That the noise in the study may cause distress to the animals
- That the animals may be distressed by meeting the other animals

Where performance was limited on this question, was either due to candidates using human ethical guidelines such as right to withdraw as a suggestion or just repeating the word ‘harm’.

Question 10

(a) This ten-mark essay question asked candidates to describe how Dr Aster could conduct a study using an interview to investigate people’s feelings about different types of exciting experience. Candidate responses showed the full range of marks, with some nice responses at the higher mark range. Candidates did still find it a challenge to achieve the higher levels (5) but there were a number of Level 3/4 responses.

Many candidates had a sound understanding of interviews and were able to make relevant decisions about the types of questions used and give examples of questions. Where candidates struggled slightly was with explicitly showing how the questions could then be interpreted, and the format of the interview.

A few things to note therefore for this question:

- In terms of format of the interview, often it was quite implicit within candidate responses. For this reason, many candidates would only gain basic credit for that part of the interview but would not achieve a detail mark which would allow them to access the highest levels.
- Candidates showed excellent knowledge of the types of questions used. Detail will come from the understanding that closed will be rating scales or yes and no answers. Detail can also come from the examples of questions that they give.
- In relation to examples of questions, candidates performed well. At least two examples which are relevant to the study were needed to achieve a basic mark for this element of the interview, and if they gave a clear/accurate/relevant type of question, then they can get a detail mark.
- Unfortunately, interpretation is the one element that many candidates did not include, and when they did it was fairly basic. Candidates could have measures of central tendency, graphs etc or analysis of qualitative data.

Candidate responses within the lower mark range were often able to give an appropriate type of question, i.e. open or closed and were able to suggest one or two questions to be used. However, at this level, candidate responses would have significant gaps within the procedure which would mean that it would either not be, or be, difficult to replicate. Candidate responses may also use other research methods which could not be linked to the interview. In this case, the response could still achieve marks for any questions they used as examples etc, but it was inevitable that these responses struggled to get higher than level 2. At this mark range some candidates still talked about ethics and sampling which was not creditworthy.

© 2025

Candidate responses within the higher mark ranges would be able to describe a procedure that would be replicable by other researchers. Most candidates at this level would suggest types of questions and then give relevant and explicit examples of these types of questions. They would provide a detailed way of interpreting the data gained from the specific type of question and be clear what type of interview that they were performing. What differentiated candidate responses at this level was whether they explicitly talked about structured/unstructured interviews and then went on to justify their choice and then use the types of questions which would be relevant for this format. These candidate responses would be examples of those who would achieve high level 4, or even level 5. Very few candidates at this mark range discussed anything about ethics or sampling which was pleasing to see.

It is essential that candidates are prepared for this question and have a clear understanding of the four required features for each method they can be asked about. This will ensure that in future series, candidates are able to achieve the marks at the highest levels.

(b) (i) and (ii)

Candidate performance on these questions was as expected. Most candidates could give an appropriate strength and/or weakness, such as only using closed questions, meaning that they were easier to interpret due to quantitative data (strength), but a lack of ability to know the reasons behind their answers (weakness). Where performance was limited, it was either due to candidates not giving enough detail for the second mark or using a different part of the procedure for the two parts of this question. It is important that candidates read the question properly to ensure that they know whether they have to use the same part of the procedure in both parts of the question, or not.

© 2025

PSYCHOLOGY

Paper 9990/22
Research Methods

Key messages

- Candidates need to ensure that they understand the expectations for different command words used on the paper; for example, if the command word is a state or identify, then often responses can be single or just a few words, i.e. identifying a type of data.
- Candidates need to ensure that they can define/explain key methodological terms within the specification such as standard deviation and informed consent in relation to children.
- Candidates need to ensure that they link their answers to the information given in the scenario if asked to do so. If the question talks about 'this study' then there is an expectation that a response provides an explicit link. Often candidates showed excellent understanding of named issues/studies but then lost marks for giving generic responses.
- There were gaps in knowledge in relation to some of the key studies, specifically Fagen *et al*. Candidates should ensure that they are familiar with all studies given to them on the specification.
- It is worth noting that candidate responses for the higher (6 and 10 mark) tariff questions showed excellent knowledge and understanding of semi-structured and unstructured interviews and experiments. There were many thoughtful responses for the extended essay question and there was an increase in Level 4 and 5 responses seen. Candidates should be commended for their performance on these questions.

General comments

This was the second November series for the new Psychology specification. Paper 22 scripts provided the full range of marks, showing a high level of knowledge and understanding across many areas of the specification. Where performance was limited, it was due to gaps in knowledge of key terms, or a misunderstanding of the demands of the question. This was noticeable when looking at questions on standard deviation (**4(b)(i)** - **4(b)(ii)**) and ways of producing quantitative data (**8(a)** – **8(c)**). Candidate responses also showed some gaps in knowledge when referring to the Fagen *et al*. study (elephant learning), with responses showing some misunderstandings in relation to how a structured observation was used. It is also worth noting, however, that candidate's knowledge of Milgram and Piliavin *et al*. was excellent across the board. Candidates showed good ability when structuring 10-mark responses (**6** and **10(a)**) with many able to produce thoughtful procedures which incorporated all the elements required within the question. Centres have prepared candidates well for the exam, specifically for the extended responses. For future series, candidates need to ensure that they have a good understanding of command words, key research method terminology such as validity and reliability, and the studies which have been named on the specification.

Comments on specific questions

Section A

Question 1

(a) A large proportion of the candidates were able to achieve the mark available for this question. Where candidate responses did not achieve the one mark, it was invariably due to candidates only mentioning about the fact it was an order effect in repeated measures design, but not mentioning about performance getting worse, which meant they were unable to be credited for their response.

© 2025

(b) Many candidates were able to achieve the mark available for this question. Where performance was limited, it was either due to vague answers, i.e. just saying 'the mindfulness tasks' without going on to suggest why this is relevant or stating that it was due to the FFMQ being long, so they got bored. This is not creditworthy as they would not be repeating a task here, just performing a longer task.

However, if candidates talked about the FFMQ that they needed to fill in after the tasks, then this could then be relevant (as it is the second time they will have done it, so will be bored of the questionnaire).

Candidates need to ensure they answer the question fully to ensure they can achieve the marks available.

Question 2

(a) This question was well answered, as candidates' knowledge of this key term was excellent. Where candidates did not achieve the mark, it was due to them mixing up quantitative and qualitative data. This was a small minority of candidates, however.

(b) Candidate responses to this question were pleasing with the many candidates getting at least one mark, with a significant proportion then going on to get two. Candidates who used Piliavin *et al*. achieved both marks more often than those that used Milgram.

A couple of things to note for this question:

- The second identification or detail mark needs to be made from the same study. A few candidates would identify qualitative data for the Milgram study and then go on to do the same for Piliavin *et al*. This often led to candidates only getting one mark as no detail (or second identification) was given for either study.
- Candidates could only get 1 mark maximum for examples of comments. A few candidates would just provide a list of comments made and only achieved 1 mark, even saying that they were comments could have gained the second mark.
- Some candidates mistakenly said that Piliavin *et al*. interviewed the participants which was not correct. It is important that candidates have good knowledge of the methodology used within each study named in the specification.

Question 3

Candidates struggled slightly with this question. Candidates had good knowledge about what a structured observation entailed, and most could give a generic strength. However, few candidates would then link this strength properly to the Fagen *et al*. study and therefore often only got one mark at most.

Types of answers which could have achieved two marks included:

- Operationalised categories allow for replication (generic). Such as knowing that trunk up meant a specific position such as…. (linked explanation)
- Operationalised fixed categories improve the reliability of the study (generic). As observations such as the two trunk positions can be compared to ensure agreement (linked detail)
- Fixed categories mean that researchers do not get distracted by irrelevant elephant behaviours (linked) making the research more valid (as it will be related to the aim).

However, responses such as "operationalised fixed categories improve the reliability of the study as observations can be compared to ensure agreement", would only get 1 mark as it does not mention elephant behaviour at all.

Question 4

(a) Candidates found this question one of the most challenging on the paper, with a number of candidates not achieving any of the marks. Most candidates focused on generalisability which was good, however they did not specify who it would be generalisable to (as in males and females/both genders).

© 2025

Those candidate responses which did link this to gender would then struggle to provide some detail to gain the second mark, i.e. that the two genders may react differently to the situation, so you need both.

Responses that could have achieved two marks could be:

- It means you can generalise to both genders/both males and females (by far the most common), as males and females may react differently to the situation, so you need to measure both to be able to generalise.
- Males and females may differ in terms of their responses due to levels of attention/knowledge of cartoons, which means that it would not be just the independent variable of age affecting the dependent variable.
- If there were more males than females, then this could be a reason why the results were gained (and not just age) as gender would be a confounding factor which is changing the results of the study.

Unfortunately, only a small minority of candidate responses provided this detail and most were therefore limited to 1 mark.

(b) (i) Candidates who had clear knowledge of the standard deviation were able to get 2 marks very quickly, i.e. by saying it was 'a measure of spread around the mean'. However, there were many candidates who did not show sufficient knowledge, with some suggesting it was a measure of central tendency, or it was the range which could not gain any credit. There were several candidates who did not attempt this question which showed a lack of knowledge.

(ii) How well candidates did on this question was directly correlated to the knowledge shown in the previous question. If candidates had shown knowledge that standard deviation is a measure of spread, then they would usually achieve one mark for suggesting that adults ages are more spread out from the mean than child's ages. There were a minority of candidates, however, that did not provide any further detail to achieve the second mark.

For the second mark, candidates could specify the ages of the children and adults in the study, i.e. saying that children had to be between 4 and 7, whereas adults had a wide range of ages they could be. This second mark could also be gained in isolation without the idea of a wider spread and indeed some candidates who had not achieved any marks in **part (b)(i)** were able to get that second mark in isolation.

Question 5

Candidate responses to this question were pleasing, although they did find the examples challenging. It is worth reminding candidates that if they are finding it difficult to remember the key studies, they can use their own examples of a study (either real or made up) if it explicitly shows the features needed for each type of interview. There were also several instances where candidates described structured interviews rather than unstructured, showing the importance of reading the question carefully.

At the lower end of the mark range, candidates usually achieved one or two marks for suggesting that interviews were self-reports, and that open and closed questions could be used. At this mark range, one description was more detailed than the other (usually semi-structured interviews), or only one was present or correct. Examples were often not present at this mark range.

At the higher end of the mark range, candidates produced some thoughtful responses which looked at a variety of features of the different types of interviews, and relevant studies linked to each type were given.

Question 6

(a) (i) Candidate performance on this question was mixed. Although many candidates did achieve the one mark for identifying confidentiality as the ethical reason, many either did not provide an explanation for this reason and/or did not link it back to the study. This was unfortunate as it was

© 2025

often clear that candidates had good knowledge of the specific ethical issue but just did not answer the question fully.

(ii) Candidate performance on this question was pleasing, with the vast majority of candidates able to provide a relevant ethical reason and link it back to the idea of negative life events.

The most common 2-mark responses included:

- Psychological harm (1) as Mr J may feel upset/uncomfortable recalling negative life events (1)
- Privacy (1) as negative life events are more likely to be personal and so he would not want to reveal them (1)
- Right to withdraw (1) as Mr J may become uncomfortable with having to recall events which he may not want to and so he can stop taking part (1).

Where performance was limited, it was either due to the wrong ethical issues being identified (deception and informed consent did appear in some responses and these were not creditworthy), or there was no reference back to the study for the second mark. This was very much the minority of candidates, however.

(b) There were no real issues with this question, with many candidates able to provide an accurate description of triangulation, with many then going on to provide a linked example.

Where performance was limited, it was either due to:

- A misunderstanding of what triangulation is, which would score zero
- Generic responses which did not link back to the study so could only achieve one mark.

One thing worth noting for this question is that triangulation can be within methods as well as between methods. Most candidates would give examples of being able to do different methods such as interviews, observations, etc., and some would discuss using quantitative and qualitative aspects of one method (usually interviews of questionnaires) and this was appropriate.

(c) Candidate performance on this question was pleasing with many getting the majority, if not all, of the marks. Common creditworthy responses included:

- Low generalisability (the most common answer) as only one person is being studied so may not represent the whole of the population
- The researcher may become too attached to the participant and therefore the responses given by the participant may be interpreted according to the researcher bias
- Social desirability/demand characteristics, as the participant is working closely with the researcher and so may guess the aim/want to please the researcher so change behaviours
- distress (often linked to Saavedra and Silverman or the intensity of study)
- confidentiality (as in depth/only one person then may be more identifiable)

One thing to note, time consuming/costly are not creditworthy as weaknesses as often this is similar for other methods.

Where performance was limited, it was usually due to a lack of detail given. Candidates need to ensure that they provide enough detail within each point to be able to achieve the full marks available.

Question 7

(a) Candidate performance on this question was disappointing, with many candidates not providing a way that was measurable. For example, just saying ‘the words used’ is not operationalised/measurable, which meant that they could not achieve the mark.

(b)(i) The most important part to note for this question is that somewhere in their responses candidates need to refer to the need for understanding of what the study entails and agreement to the child taking part in the study to get the two marks.

Therefore, the most common response seen was:

© 2025

- Consent means a participant must understand and agree to a study
  - but as children are too young to understand what will happen in a study the parents must agree for them
  - but children are legally too young to agree to participate so parents must do it for them.

Candidates found this question challenging with the vast majority not talking about both understanding and agreement, which meant that often they were limited to one mark at most. For example, if candidates wrote about the idea of children not being able to consent on their own and so the parents must agree on their behalf, they could have the one mark for linking their answer to parents and children, but they do not have the full explanation needed for the two marks.

(ii) Candidates found this question challenging with many just saying 'Lorna could ask them in a child friendly way whether they want to be part of her study'. Although this is very basic, it could still get a mark, but it does not provide the link needed to achieve the second mark. To achieve both marks, candidates could suggest a generic question on consent and then give some linked detail (which is child friendly) or suggest a specific question which is child friendly and linked to the study.

(c) Candidates were often able to achieve at least one mark for this question for the idea that correlation does not mean causation, However, many of these candidates were then unable to either provide enough detail and/or link to the study to enable them to achieve the second mark. For example:

- Many candidates did discuss the fact there could be something else (3rd variable) causing the results but did not link it back to the study so only got one
- Some candidates would suggest 'that talking to toys doesn't really help vocabulary' – this does not relate to the methodological nature of the question asked and is just an opinion, not based on knowledge
- If candidates said the above but went on and say that there must be something else that caused the results, then they could achieve a mark.

Question 8

(a) Candidates found this question the most challenging on the paper. It is not the most complex of questions, however there was a significant amount of misinterpretation of the demands required. As alluded to above, most candidates did not interpret the question as intended and did not suggest specific questions as required. For example, they would just say they would use rating scales/closed questions. For this reason, many candidates did not achieve any marks. The questions did not have to be on interviews and questionnaires, it could be two different sorts of questions on the questionnaire, for example rating scales and yes/no answers which they then score or just two different questions related to the study using the same type of question and same method (i.e. questionnaire).

(b) (i) Candidates found this question challenging although many candidates were able to get at least one mark. The most common responses included the idea that closed questions are easy to analyse, such as calculating the average/being able to compare between groups, or that participants can fill in a questionnaire without the researcher there, so they are more likely to be honest. This makes questionnaires more valid than interviews which are likely to be face to face so may lead to social desirability.

Unfortunately, there was only a minority of candidates who were able to achieve the two marks available.

© 2025

(ii) Candidates found this question challenging. The biggest issue was that many candidates did not make the connection that both parts have to be for the same method. So, they could not do questionnaires in **(b)(i)** and then interviews in **(b)(ii)** (or closed then open questions, etc.). Where this happens, they would only be credited for one or the other. Unfortunately, this happened often. The most common responses were the idea that questionnaires may be subject to social desirability bias or closed questions are too limited and therefore may not contain the response the participant wants (so does not actually represent their real thoughts).

(c) There was no real issue with this question regarding candidate responses. Where candidate performance was limited, it was due to candidates drawing a bar chart rather than a histogram. Candidates need to be aware of the difference between a bar chart and a histogram, and when they are used, for future series.

Question 9

(a) Candidate responses showed the full range of marks, with some really thoughtful and creative responses seen at the higher mark range. Candidate responses did still find it a challenge to achieve the highest Level (5) but there were a significant number of Level 4 responses which was pleasing.

Many candidates had a sound understanding of field experiments and were able to make relevant decisions about how to operationalise the independent variable and measure the dependent variable. The part which candidates found more challenging was the experimental design. Many would either not specify an experimental design or just name one (and sometimes it was the wrong one for their procedure). To get the detail mark here candidate responses needed to first show that they knew the correct experimental design and then give some detail showing how this would be implemented.

In terms of the features required for a field experiment:

- Most candidate responses at both the lower and higher mark range got a basic tick for knowing the independent variable (IV) was brightness vs dullness. What differentiated between the lower and higher mark ranges was responses that would describe what this would entail, i.e. for brightness condition they will use pinks/oranges/yellows and for dull it will be browns/blacks/grey. This would achieve a detail mark for this feature.

- Candidate responses at both the lower and higher mark range described and operationalised the dependent variable (DV) well. Many went into detail, talking about the amount of eye contact with the advert/giving them a test about that content of the advert/ using cameras and then analysing the eye contact with the advert – these are nice responses and all would be worth the detail mark.

- In terms of controls, candidate responses needed at least two explained clearly, i.e. the same size advert (if on a billboard), the same content on the advert (only colour being different), that the models holding the advert look and dress the same, that the test is the same (if they use it), that they use the same train/same journey for each condition. Candidates at both the lower and higher mark range were able to identify the controls needed. The differentiation came with the clarity of the explanation given for these controls, which was more likely to be present in the higher mark range although not exclusively so.

- The experimental design was the weakest feature of the field experiment, with a lot of the designs used implicitly. At the lower end of the mark range, often candidates would explain in their procedure that they have used independent measures/repeated measures without them stating it, i.e. if they say that they will use the same train journey but different days for each condition. Some candidate responses would explicitly state the correct experimental design and then go on to say that this was because the participants are only doing one condition each. This would be considered detail for this feature.

It is essential that candidates are prepared for this question and have a clear understanding of the four required features for each method they can be asked about. This will ensure that in future series candidates are able to achieve the marks at the highest levels.

© 2025

(b) Most candidates were able to achieve 1 or 2 marks for identifying the strength. The differentiation came with the level of detail/explanation/justification for each strength, i.e. a link to replicability/validity/reliability.

It is worth noting that candidate responses would not get a second mark for just stating 'means its high in validity', they would have to do a little more. For example, the study was conducted on a regular commuter train with normal passengers meaning it is high in ecological validity as would represent real life behaviour'.

Candidates need to ensure that they explain any key terms/issues/debates when used.

© 2025

PSYCHOLOGY

Paper 9990/23
Research Methods

Key messages

- Candidates need to ensure that they understand the expectations for different command words used on the paper; for example, describe and explain. For describe, candidates need to ensure that they provide a sufficient number of unique points related to the marks allocated to the question, whereas for explain, candidates need to identify a particular feature/concept/theory and then link their detailed point to the feature/concept/theory they have identified. Often, explain questions had poorer outcomes in terms of marks as a candidate just described.
- Candidates need to ensure that they can define/outline key terms within the specification such as 'dependent variable' and 'counterbalancing' and know the difference between terms such as open and closed questions. This is also important in questions where research methodology is used. It was clear that in some questions which talked about methods, there was a lack of understanding about what the words meant.
- Candidates need to ensure that they link their answers to the information given in the stem if asked to do so. Often candidates showed good understanding of named issues/studies but then lost marks for giving generic responses.
- Candidates showed some gaps in knowledge about key studies on this paper. For example, the idea of relative difficulty for Fagen *et al*. and the use of the FFMQ in Holzel *et al*. Candidates need to ensure that they know the details for all the key studies, as this is as important for Paper 2 as Paper 1.
- Candidates seem to struggle with questions which ask them about the validity of a piece of research, or a method within a piece of research. For future series it is worth candidates not only knowing the different types of validity named on the specification, but also what may lower validity, and how it could be improved.
- It is worth noting that candidate responses for the 10-mark extended response question showed good knowledge and understanding of an experiment. There were several thoughtful responses for this extended response question, and candidates should be commended for their performance.

General comments

This was the second November series for the new Psychology specification. Paper 23 scripts provided a good range of marks, showing a good level of knowledge and understanding across most areas of the specification. Where performance was limited, it was due to a lack of knowledge of key terms, or elements of the key studies or a misunderstanding of the demands of the question. This was clear when looking at questions on counterbalancing (**Q1(a)**), elements of the Fagen *et al*. study (**Q 3**), and issues with validity (**Q 9**, **Q 8(b)(i)** and **Q 8(b)(ii)**).

Candidate responses showed good knowledge and understanding of some of the key studies, such as when referring to Piliavin *et al*. which meant that many candidate responses were able to achieve most of the marks. Questions which required candidates to comment on the validity of a method, or a particular test, were still challenging for candidates as many responses would use the term without explaining the reasoning why a method was high or low in validity. For future series, candidates need to ensure that they have a good understanding of command words, key research method terminology such as validity and reliability, and the studies which have been named on the specification.

© 2025

Comments on specific questions

Section A

Question 1

(a) Candidate responses to this question were below average. The majority of candidates who achieved the mark for this question did so by using the 'ABBA' framework which was acceptable (one group did A first then B, one group did B first then A).

Where performance was limited, it was due to a misunderstanding of what counterbalancing means, or vague responses which talk about changing the order but do not relate this to the tasks/levels of the independent variable. There were several candidate responses that did not answer this question which also shows a lack of understanding.

(b) This one-mark question asked candidates to outline one reason why counterbalancing is important. There were no real issues with this question with the vast majority of candidates able to suggest that it was important due to the reduction of order effects/fatigue effects, etc. Where candidates did not achieve the mark, it was invariably due to the lack of understanding shown about counterbalancing as seen in part **Q1(a)**. Candidates in these circumstances would often just use generic evaluative terms such as validity/reliability and therefore did not achieve any marks.

Question 2

This one-mark question asked candidates to state whether a given hypothesis was directional, non-directional or null. There were no issues with this question, with very few candidates unable to achieve the mark available.

Question 3

(a) Candidates found this question a challenge with many misunderstanding the requirements. Many would have the right idea of how to answer this question, i.e. that elephants learn at different rates, but often they were fairly vague responses which did not go into detail about individual differences as an issue. This meant that many were only able to achieve 1 mark at most. A two-mark answer would be, therefore:

- Due to individual differences (1) as all elephants have different needs and therefore will learn at different rates (1)

(b) Candidates found this question very challenging, with few responses able to achieve the two marks available. Creditworthy responses could include:

- The total number of offers/cues/session given by the trainer (basic explanation so 1 mark)
- The total number of offers/cues given by the trainer with the average of the totals of all elephants used (detailed explanation)

Unfortunately, candidates focused on the idea of how long it took them/the time taken to be successful at a task which was not accurate and therefore did not achieve any marks. Where candidates did achieve marks, it was for the first bullet point but there were very few who went on to get the second mark for the use of the average.

Question 4

(a) Most candidates were able to achieve at least one mark for this question, usually for the generic idea of quantitative/objective data being produced. Where candidate responses did not achieve the second mark it was invariably due to the lack of link to the FFMQ, rather than a lack of detail:

'One strength is that quantitative data is able to be produced which means that responses to the FFMQ can be compared between participants' – this would only achieve 1 mark as apart from the use of the words FFMQ this could apply to any questionnaire with closed questions.

Where candidates achieved the two marks a common response was:

© 2025

'The FFMQ produced quantitative data (strength)meaning that the total scores before and after the study can be compared to see whether scores differed (linked detail)' – although this still seems quite generic, the important part is the candidate has shown knowledge of the use of the FFMQ in the study by knowing it was completed both before and after.

Candidates need to ensure in future series that if asked a research method question which has been linked to a study, that they make sure that they make a specific link to the use of that method in the study to achieve the full number of marks.

(b) Most candidates were able to achieve at least one mark for this question, usually for the generic idea of a person being more likely to lie/produce socially desirable answers on a questionnaire. Where candidate responses did not achieve the second mark it was invariably due to the lack of link to the FFMQ, rather than a lack of detail:

'One weakness is that questionnaires may produce socially desirable answers which means that the validity of the responses to the FFMQ would be low' – this would only achieve 1 mark as apart from the use of the words FFMQ this could apply to any questionnaire.

Where candidates did achieve the two marks a common response was:

'Participants answering the FFMQ may produce socially desirable answers meaning the scores obtained may not reflect their levels of mindfulness lowering validity (linked detail)' – although this still seems quite generic the important part is the candidate has shown knowledge of the use of the FFMQ in the study by knowing it was measuring mindfulness.

As seen in the previous question, candidates need to ensure in future series that if asked a research method question which has been linked to a study, that they make sure that they make a specific link to the use of that method in the study to achieve the full number of marks.

Question 5

(a) The vast majority of candidates achieved marks for this question, with a majority getting at least two out of three marks. Virtually all candidates at both the higher and lower mark range achieved a mark for the 'drunk versus ill' difference, with many of these identifying the difference of cane vs brown bag.

Where candidate responses did not achieve any marks, it was due to a lack of understanding of the study with responses which linked to age and gender. Responses which said that one person was drunk, or candidate responses which only focused on the drunk or ill person, was not creditworthy.

(b) Candidate responses to this question were as expected, with many able to achieve at least some marks for this question. Common responses for the first mark would be:

- To ensure that confounding variables were minimized
- For the study to be standardised/controlled

Detail would then come from the idea that results would therefore be only due to the independent variable (generic) and/or the idea that other trials would be consistent (linked detail).

For candidates, it was often the third mark that was difficult to achieve. Many explained about the idea that it was due to making sure the results were only due to the IV, with some adding the idea of extraneous/confounding variables being limited but then provided no further detail or alternative explanation for the third mark.

At the higher mark ranges there were more responses which did achieve the third mark. An example response which received three marks was:

'To ensure the study was standardised for example they wore the same slacks, all male and worse no tie (1) by having strict controls to isolate the effect of the IV on the DV (1) this improved the validity of the study as you can be confident that confounding variables did not affect the DV (1) – this was a really nice response but one, unfortunately, that was not seen that often.

© 2025

Question 6

Candidate performance on this question was pleasing, with many achieving the majority of the marks available. Examples of possible creditworthy responses were:

Open questions
- Participants choose how they respond to the question
- Can be more in-depth/detailed
- They need to be interpreted by the researchers/can be subjective
- Examples such as Dement and Kleitman asking participants to describe their dream
- Own examples such as describe or explain questions.

Closed questions
- Fixed/limited responses to a question
- Produce quantitative data
- Do not need to be interpreted so more objective.
- Examples such as Dement and Kleitman asking participants whether they dreamt (yes or no)
- Baron-Cohen used fixed choices for the eyes test
- Specific questions such as 'How many times did you watch television yesterday' (the answer choices needed to be present to achieve the mark here)

At the lower end of the mark range, candidates usually achieved one or two marks for identifying open questions as being more detailed, and closed questions having fixed choices. Examples were often not present at this mark range and when they were it was usually examples of questions not from key studies.

At the higher end of the mark range, candidates produced some thoughtful responses, and the vast majority were able to give detailed explanations of both open and closed questions but still struggled slightly to provide examples (although Dement and Kleitman was the most common example used for both types of questions). These candidate responses often achieved around 4 to 5 marks, with many getting the full 6 marks.

Question 7

(a) (i) Candidate responses were as expected for this question, with many candidates at both the high and low mark range getting at least one mark for this question and many achieving both marks. Where performance was more limited it was due to either:

- Candidates repeating the question and saying about exploring/interacting.
- Vague answers such as saying the soft toys feel different to the others. This was not precise enough for a mark.
- Talking about gender stereotypes (linking it to Hassett *et al*) which was not creditworthy for this question.

(ii) As similar to the previous question, candidates at both the lower and higher mark range were able to achieve at least one mark for this question, with most candidates at the higher mark range getting two marks.

Common responses included:

- The noisy toys may scare them, so they won't go near them
- The soft toys do not do anything so children may see them as boring (compared to noisy toys)
- They may be shy and so will not want to play ball with the other children.

Where performance was limited, it was mainly due to a lack of a link back to the toy in question. For example, they would say 'because they are more introverted' or 'because they do not want to play with other children'. In these cases, candidate responses can be awarded one mark for the explanation (and it is relevant to one of the toys they named in part **7(a)(i)**, but not the link mark as there is no explicit mention of the toy in question.

(b) Candidate performance on this question was pleasing, with many at both higher and lower mark range getting at least one mark, and many achieving both marks.

© 2025

Common ways of answering this question were

- Providing a child friendly statement such as 'whenever you would like to stop playing with the toys you can do and then you can go home' (would be two marks)
- Suggesting a way, it could be child friendly such as.
  - 'Telling the child that they do not have to stay and play with the toys' or
  - 'Telling the children that they can leave the toys at any point and use easy language, so they understand'

Where candidates did not achieve the two marks, it was invariably due to a generic statement such as 'you can tell them they can leave the study in easy language'. Although they have thought about the child friendly nature of the statement, they have not linked it to the study at all (i.e. to the toys) so can only achieve one mark.

(c) Many candidates achieved at least two, if not three, marks. Most of the marks came from the row categories and the column heading, with the other two being less present. There were some thorough tables at both the lower and higher mark range which fully deserved the three marks. A minority of candidates did not achieve any marks for this question, and this was invariably due to candidates drawing a graph rather than a table which was not creditworthy.

Question 8

(a) Candidate performance on this question was pleasing. The vast majority of candidates achieved at least one mark for this question, with many achieving the two marks available. Where candidates did not achieve the two marks it was due to a lack of operationalisation. Very few candidates did not achieve any marks but where they did it was due to candidates mixing up the DV and the IV.

(b) (i) Candidates found this question challenging. Although it was clear that candidates knew what a participant variable was, they struggled to link this to the question asked. Responses which could achieve the two marks include:

- That they may already do a shelf stacker job (so will be even more bored, changing how fast time goes).
- They may be better at timing genetically (so they will be better whether they are bored or not)
- Some people may enjoy watching any type of film more (so time will go more quickly anyway)

If candidates suggested specific disorders such as ADHD, then these could also be credited.

Unfortunately, candidates did not often do the above, and although they would suggest participant variables which could be credited for 1 or 2 marks, there was very few that were able to link this correctly to validity. Where candidates did not achieve any marks, it was due to participant variables which were not relevant such as gender/age or suggested situational variables such as dark/bright screens.

(ii) Candidates found this question challenging. In addition to candidates who misunderstood the question, many could suggest a possible change but did not link it correctly to validity. Common responses which could achieve two marks were:

- Use random allocation to each condition (change) which means that shelf stackers would not all be in one group skewing the data and lowering validity (improving validity)
- Perform a test which will see how good their timing is (change) to ensure the difference is due to the IV and not ability (improving validity).

Candidates still seem to struggle with questions related to validity. For future series, candidates should ensure that they know issues with validity in relation to key studies, but also unseen studies where practice is key.

(c) (i) Candidates found this question challenging, with most candidates not able to achieve the mark available. Most provided suggestions that would not be appropriate such as:

- Doing things they would do in daily life.

© 2025

- Using a black and white room for 30 mins.

Where candidates did achieve the mark, it was usually for suggesting that they should do no activity.

(ii) Unfortunately, as the vast majority of candidate responses did not achieve any marks for **part (c)(i)** then very few achieved marks for **part (c)(ii)**. Responses which may have gained the mark available were:

- Baseline comparison between the exciting/boring groups
- Differences might be due to vision/excitement not boredom
- Removing the film as a confounding variable.

Unfortunately, this was the case for very few candidates.

Question 9

(a) (i) This one-mark question asked candidates to name the type of graph shown. The correct answer for this question was a bar graph. There were very few candidates that did not get the correct answer for this question. Those candidates that did not achieve the mark, suggested that the graph was a histogram.

(ii) Candidate responses to this question were pleasing with many able to give an appropriate conclusion for this study.

Where candidate performance was limited, it was due to candidates giving a result rather than a conclusion, i.e. that there were more positive early experiences than negative experiences in early life (with no reference to later liking).

(b) Candidates found this question challenging; they struggled to suggest a relevant problem and if they did do so, they did not link it to validity. The most common response which achieved one mark was:

- The idea that adults may not remember early experiences.

The second mark could then be gained by:

- They may miss out experiences that they have had.
- They may lie and say experiences that they did not have.

Where candidates did not achieve any marks, it was due to a misunderstanding about the demands of the question, with many candidates talking about social desirability; relating lying to what Daniel may want to hear, rather than the poor memory as an adult, the lack of ecological validity, or personal food preference; none of which were creditworthy.

Question 10

(a) This ten-mark essay question asked candidates to describe how a researcher could conduct a case study to investigate Liam's unusual ability to focus his attention. Candidate responses showed the full range of marks, with some nice responses at the higher mark range. Candidates did still find it a challenge to achieve the highest levels (4 and 5) but there were a large number of Level 3/4 responses which was pleasing to see.

Many candidates had a sound understanding of case studies in general and were able to make relevant decisions about the background information gained on the participant and the methods that they were going to use. However, where candidates struggled slightly was with the information that needed to be collected about the participant, such as age, history of exceptional attention and other information from parents, which would have enabled candidates to gain a detail mark for that element. One of the elements that candidates needed to achieve to get into the higher levels was to suggest how to interpret the data gained. Often this was fairly basic with many just suggesting a graph for quantitative data. Where candidates used qualitative methods, they rarely suggested how to interpret it.

© 2025

Candidate responses within the lower mark range were often able to give basic techniques to be used, i.e. would suggest questionnaire or interview but not who this would be on or the types of questions used. Also at this level, candidate responses would have significant gaps within the procedure which would mean that it would not be replicable. At this level, the details about the participant were often missing and, if present, would be very basic i.e. suggesting finding out the age it started and nothing else. At this mark range some candidates still talked about ethics and sampling which was not creditworthy.

Candidate responses within the higher mark ranges would be able to describe a procedure that would be replicable by other researchers. Most candidates at this level would suggest at least speaking to the participant before the study and finding out information about his ability. Many would also suggest speaking to other people such as family or friends. Candidates at this level would pick at least two techniques or more and would be able to suggest what these techniques should be looking for.

What differentiated candidate responses at this level was whether they explicitly talked about the analysis of the information gained from these techniques. Even at this level some candidates struggled when talking about qualitative data, although some did suggest looking for themes, but for quantitative data they would suggest measures of central tendency, graphs, etc. These candidates' responses would be examples of those who would achieve high level 4. Very few candidates at this mark range discussed anything about ethics or sampling.

It is essential that candidates are prepared for this question and have a clear understanding of the four required features for each method. This will ensure that in future series, candidates are able to achieve the marks at the highest levels.

(b) Candidate performance on this question was mixed. Many candidates would achieve 1 or 2 marks but then fail to link either of their points to validity. Common responses were:

- The use of naturally occurring noises to increase ecological validity.
- The use of quantitative data which does not require subjective interpretation.
- The lack of confounding variables (if study took place in a laboratory).
- The use of triangulation which allow you to compare the data gained.

Once again, candidates' weakness was their ability to link responses to validity. As suggested in the previous questions, candidates would benefit from practice at unseen scenarios in relation to validity issues.

© 2025

PSYCHOLOGY

Paper 9990/31

Specialist Options: Approaches, Issues and Debates

Key messages

Questions 1, 3(a), 5, 7(a)(i), 7 (a)(ii), 9, 11(a), 13(a), 15(a) –

These questions in this exam asked candidates to apply an area of the syllabus (theory, treatment/explanation of a disorder, technique, self-report/measuring techniques etc.) to explain how it can be applied to a particular scenario or context. It is important that candidates are aware of the titles of the bullet points in the syllabus. It would be helpful for candidates to do revision notes with the title of the topic area and bullet point at the top so that they can identify which part of the syllabus these types of questions are referring to. Candidates should also refer directly to the scenario/context in the question in their response.

Questions 3(b), 7(b), 11(b) and 15(b) –

These questions in this exam asked candidates to evaluate with a strength or weakness the theory/treatment/measurement that was outlined in the candidate's response to **part (a)** of the question. In this exam, this type of question also asked the candidate to outline one problem psychologists have investigating a theory with participants. It would be helpful for candidates when doing revision to learn strengths and weaknesses of the theories, techniques, explanations, treatments, etc. they have learned and put these into their revision notes. In addition, candidates should learn problems psychologists might have when investigating what has been outlined in **part (a)**. They should also practice explaining the evaluation point in the context of the question.

Questions 2, 6, 10 and 14 –

Part (a) - These questions in this exam asked the candidate to outline one of the issues and debates, possibly with an example from the syllabus content. This part of the question could also ask the candidate to outline a theory, study, technique/treatment or self-report used by psychologists that is named in the syllabus. The revision technique outlined previously in this report will aid candidates to learn the syllabus material.

Part (b) – This part of the question may ask candidates to explain a strength or a weakness of the issue/debate or the syllabus content outlined in **part (a)**. The question could also ask candidates to explain how a bullet point in the syllabus links to or supports one of the issues or debates. It would also be useful for candidates to write revision notes where they define the issues/debates and prepare a strength and a weakness of each issue and debate to prepare for the **part (b)** of this type of questions. Candidates should also note how the topics covered in the syllabus fit with each of the issues/debates. These questions in this exam were worth 2 marks for each part of the response and therefore a short response is appropriate.

Questions 4(a), 8(a), 12(a) and 16(a)

These questions in this exam came from one or two of the bullet points in the syllabus. This exam asked the candidate to describe a study, theories, two treatments and theories (and/or studies) about making and presenting a diagnosis. For this exam, some of the responses used the incorrect topic area in the syllabus, the description was brief or lacked focus on the topic area (such as how the treatments treat depression rather than mental health problems in general).

It could be useful for candidates to create revision notes with the title of each topic area and these description in the bullet point as the header. Alternatively, candidates could create a mind map and put this information in the centre. Some of the responses were too long for the first option (e.g. **Question 4(a)**).

© 2025

These types of responses did usually receive a high mark. However, this often meant that the candidate did not have enough time to write a complete response for the second option (e.g. **Question 12(a)**). It would be useful for candidates to practice writing an appropriate length response to these types of questions. Teachers could set a word limit or do a timed response at home or in class. These should be done without referencing any notes or the textbook while doing the timed response.

Questions 4(b), 8(b), 12(b) and 16(b)

This question will always ask the candidate to evaluate the studies, theories, characteristics/explanations/treatments of disorders or techniques described in **part (a)** of the question. The response must include at least two evaluation issues, including the named issue, to be considered to have presented a range of issues to achieve the top band. However, most responses that evaluated using two issues in this exam, achieved in the lower bands due to the response being superficial and often with little analysis. Some responses that considered three issues tended to achieve higher marks as these responses were able to demonstrate comprehensive understanding with good supporting examples from the studies, theories, characteristics/explanations/treatments of disorders or techniques described in the **part (a)** of the answer.

The candidate must also provide some form of analysis to access level 3 and above. This could be done by discussing the strengths and weaknesses of the issue being considered, presenting a counterargument to the issue under discussion or comparing the issue between two studies and/or theories. The response needs to explain the comparison/strength/weakness or counterargument with examples from **part (a)** of the question. It was common for responses to state that two theories or treatments for example, were 'similar' or 'in contrast' for an issue without any explanation as to why they could be compared in this way. This is assessed as little analysis. It could be limited analysis if the response gives several contextualised examples to support the evaluation points. A conclusion at the end of each issue would be helpful to show excellent understanding of the issue under discussion. To achieve the requirements of the level 4 and 5 descriptors, it would be best to structure the response by issue rather than by study and/or theory. It would also be ideal for the response to start with the named issue to make sure the answer covers this requirement of the question. A small minority of candidates did not evaluate using the named issue. Quite a few of the answers were structured by study/theory/treatment rather than by the issue which often led the response to be quite superficial and repetitive. Some of the responses did do analysis. Candidates should be aware this question is worth 10 marks and need to include an appropriate amount of information.

General comments

The marks achieved by candidates for this session of the 9990 specification achieved across the full range of the mark band which was very pleasing to see. Some candidates were well prepared for the exam and showed good knowledge, understanding, application and evaluation throughout their responses. A significant number of candidates were not as well prepared and showed limited knowledge and understanding with brief, superficial and sometimes anecdotal responses. These candidates often had limited evaluation and application skills.

Time management for this paper was good for the majority of candidates and most attempted all questions that were required. A number of candidates did not respond to one or more of the questions asked in the option area. A very small number of the candidates attempted to respond to more than two topic areas but often did not attempt all the questions for each option chosen. These responses achieved at the lower end of the mark band.

The questions on clinical were the more popular choice of option, followed by health.

Comments on specific questions

Clinical Psychology

Question 1

The responses to this question covered the full range of marks. The vast majority were able to suggest at least one explanation of why Hugh has OCD and many suggested two. The most common explanation given was that Hugh inherited OCD from his mother. However most received 1 mark for this as responses did not give any research evidence and/or an outline of concordance rates where monozygotic twins have higher

© 2025

concordance than dizygotic twins. This was necessary to achieve 2 marks as this is the detail to explain genetic causes of OCD. In addition, higher mark responses also outlined either behavioural or psychodynamic explanations. Some who attempted behavioural simply suggested it was learned from the mother rather than using psychological terminology (such as through positive or negative reinforcement). This type of response as well as the many anecdotal responses, such as being traumatised as a child, were not creditworthy.

Question 2

(a) Most responses achieved at least 1 mark for this question with a significant number achieving full marks. Many responses gave a definition of culture accurately, referring to shared norms, beliefs and values, which was sufficient to gain 1 mark. To achieve the second mark, candidates needed to show understanding of the term *differences*. This was often done successfully by explaining how norms or beliefs vary between cultures or by giving a clear example of a cultural difference (for example, differences in attitudes towards eye contact or family roles). Weaker responses tended to be too vague, such as stating that cultures mental health, without further explanation or exemplification. Some responses used the wording in the question such as 'it is how cultures are different' which was not creditworthy.

(b) Responses to this question covered the full range of marks available. Many responses achieved 1 mark by giving a general reason why cultural differences may affect mental health, although these often made limited reference to anxiety disorders specifically. Stronger responses explained how anxiety disorders may be viewed differently across cultures, sometimes using examples from different cultural contexts. The strongest responses went on to explain how these differing views could affect diagnosis, for example from the doctor's perspective, or by considering how cultural expectations may influence what an individual reports during a clinical assessment. A small number of responses focused effectively on how anxiety disorders are perceived or accepted within a particular culture. Weaker or non-creditworthy responses tended to state simply that cultures view mental health differently, without further explanation or clear reference to diagnosis of an anxiety disorder.

Question 3

(a) Responses to this question covered the full range of marks available, with many achieving 3 or 4 marks. Strong responses clearly outlined the key stages of systematic desensitisation, including the teaching of relaxation techniques, the construction of a fear hierarchy, and gradual exposure to feared stimuli while practising relaxation. Many of these responses were well contextualised to flying insects, often using appropriate examples of low- and high-fear situations. Some responses made effective use of psychological terminology, such as reciprocal inhibition, and explained that anxiety must be reduced or absent before progressing to the next level of the hierarchy. Weaker responses tended to give a very generic outline of therapy, sometimes gaining only 1 mark for identifying a fear hierarchy or mentioning relaxation without further detail. A small number of responses incorrectly suggested that the phobic stimulus would be paired with some positive thoughts or a reward, which is not creditworthy for systematic desensitisation.

(b) Responses to this question covered the full range of marks. Many responses were able to gain 1 mark by identifying a general weakness of therapy, such as the treatment being stressful or requiring commitment over time. Stronger responses explained why this was a weakness specifically of systematic desensitisation, often by referring to the need for repeated exposure to the feared stimulus and the possibility that this may lead some individuals to discontinue treatment before it is completed. Some responses supported this with examples, including reference to Habiba's fear of flying insects. Other common weaknesses included practical issues such as time, cost, and access to a therapist; however, these often achieved only 1 mark where the explanation was brief or not clearly linked to systematic desensitisation. A large number of responses incorrectly stated that the treatment was unethical or that it would cause psychological harm or make the phobia worse; these were not creditworthy, as systematic desensitisation is a consent-based therapy and distress is managed through relaxation techniques.

© 2025

Question 4

(a) Responses to this question varied and covered the full range of marks available. Strong responses gave clear and accurate descriptions of both MAOI anti-depressants and Ellis's rational emotive behaviour therapy (REBT). It was not necessary for responses to include all of the content listed in the mark scheme; rather, higher-level responses selected relevant information and explained it accurately. For MAOIs, good responses typically explained that monoamine oxidase inhibitors prevent the breakdown of neurotransmitters such as serotonin, noradrenaline and dopamine, resulting in higher levels in the brain and an improvement in depressive symptoms. Some responses also included appropriate reference to effectiveness or side effects. For REBT, stronger responses clearly outlined the ABC (and sometimes ABCDE) model, explaining the role of irrational beliefs in depression and how disputing these beliefs can lead to more rational thinking and improved mood.

Weaker responses often lacked sufficient detail for one or both parts of the question. For MAOIs, some responses stated that they inhibit monoamine oxidase but did not accurately describe the effect on neurotransmitters or confused MAOIs with SSRIs. For REBT, weaker responses often described only the ABC model without explaining how disputing beliefs helps to treat depression or gave a generic outline of CBT or talk therapy, which was not creditworthy. Some responses confused REBT with Beck's cognitive therapy, limiting the level awarded.

(b) Marks for this part of the question covered the full range available, although the most common levels awarded were Level 1 and Level 2. A notable proportion of responses (approximately 13 per cent) did not attempt this question. Stronger responses (Level 3 and above) were clearly structured issue by issue and included evaluation of both MAOI anti-depressants and Ellis's REBT, often covering three or four issues and debates. In addition to the named issue of reductionism versus holism, common issues discussed included determinism versus free will, application to everyday life, and individual versus situational explanations. Better responses contextualised these issues effectively, for example by explaining how MAOIs may be seen as biologically deterministic but involve free will in choosing to take medication, or how REBT requires active engagement and choice by the patient. Some responses were also able to explain how both treatments could be viewed as reductionist while recognising attempts to address more than one aspect of depression.

Weaker responses (Level 1 or Level 2) often failed to apply issues and debates to the treatments, instead giving brief or generic definitions without clear reference to MAOIs or REBT. Many responses simply stated which side of a debate each treatment supported, such as claiming MAOIs are reductionist or REBT is holistic, without further explanation or contextualisation, which was not creditworthy. Some responses focused mainly on strengths and weaknesses of the treatments without linking these to the issues and debates, which limited the level awarded.

Consumer Psychology

Question 5

Responses to this question covered the full range of marks, although many were awarded 1–2 marks. Most responses were able to name or outline at least one appropriate decision-making strategy, most commonly satisficing or referring generally to non-compensatory strategies. Stronger responses achieved full marks by not only identifying a suitable strategy but also explaining why it was appropriate to the scenario, specifically linking their answer to the fact that the website design meant comparisons between chairs were not possible. These responses often correctly explained how Abdul might select a chair based on meeting basic requirements or prioritising a single key feature.

Weaker responses frequently failed to apply the strategy to the context of internet shopping where comparisons between chairs were not possible, limiting the marks awarded. Some responses described compensatory strategies, such as weighing up multiple features, despite the scenario indicating that comparisons could not be made, which was not creditworthy. Other non-creditworthy responses included references to the take-the-best heuristic, high or low self-monitors, anecdotal descriptions of shopping behaviour, or summaries of the scenario without identification of a decision-making strategy. Very few responses demonstrated a clear understanding that non-compensatory strategies are more likely to be used when website design restricts comparison.

© 2025

Question 6

(a) Responses to this question covered the full range of marks. Some responses achieved full marks by giving a clear definition of determinism and linking this directly to a choice heuristic, often explaining how behaviour is influenced by external factors rather than free choice. A small number of responses also referred to internal causes of behaviour leading to behaviour rather than a choice to behave in a particular way. Many responses achieved 1 mark by providing a basic definition of determinism but failing to link it to a choice heuristic, or by outlining what a choice heuristic is without explaining how it demonstrates determinism. Non-creditworthy responses often attempted to define determinism in terms of experiences determining behaviour without explaining the lack of cognitive control or did not make any relevant reference to choice heuristics being deterministic.

(b) Responses to this question were generally weak, with very few achieving any marks. A notable proportion of responses (approximately 10 per cent) did not attempt the question, and only a small number of candidates did the consumer option. A very small minority of responses achieved 1 mark by identifying a strength of determinism, such as the idea that behaviour is caused by external factors and briefly linking this to choice heuristics by implementing appropriate sales strategies. No responses achieved full marks, as responses rarely made a clear link between a strength of deterministic explanations and a specific choice heuristic. Many responses focused on stating that it helps with understanding consumer behaviour. This type of response did not identify or explain a strength of determinism. This highlights a need for greater awareness of the advantages of deterministic explanations, which are widely used in psychology.

Question 7

(a) (i) Responses to this question generally accessed the full range of marks. Many responses correctly identified Zelda as an explorer (or native) and Louis as a tourist (or raider), which gained full marks. The most common correct pairing was explorer for Zelda and tourist for Louis. However, a large number of responses demonstrated misunderstanding of the command word *identify*. Some responses provided extended descriptions or rewrote the scenario in different words rather than naming an appropriate shopper spatial behaviour pattern, which was not creditworthy. Common incorrect responses included identifying either shopper as a specialist or identifying Louis as a native. Some responses justified the behaviour in terms of shopping needs (for example, stating that Zelda was doing her main shop), rather than identifying a recognised spatial behaviour pattern from the syllabus.

(ii) Responses showed a wide range of quality. Many responses gained the first mark by correctly identifying a different spatial behaviour pattern for Frank, with tourist being the most commonly selected. A small number of responses achieved full marks by outlining one or more key features of a recognised shopper type and linking these clearly to Frank's behaviour in the supermarket. Some responses also gained credit by accurately describing a type of trip, such as a short or central trip.

Weaker responses tended to be very brief or vague, for example stating that Frank would 'zig zag' around the store, which did not clearly indicate a recognised spatial behaviour pattern. Common non-creditworthy responses included inventing behaviour patterns not in the syllabus, such as 'on a mission' or 'routine-based' or simply rewriting the scenario for Frank using different words. This was particularly common where a similar approach had been taken in **part (a)(i)**.

(b) Responses to this question covered a range of marks. Stronger responses were able to explain a clear problem faced by psychologists when investigating shoppers' spatial movement patterns and link this explicitly to shopper behaviour. Common creditworthy responses included explaining that shoppers may not behave consistently across different stores or shopping goals, making classification difficult, or that shoppers may change their behaviour if they are aware they are being observed. Some responses also appropriately referred to ethical issues, such as lack of consent when observing shoppers.

Weaker responses often identified a problem without any explanation or without clear reference to shoppers' spatial movement patterns, which limited marks to 1 or resulted in no credit. A large proportion of responses did not address the question in terms of a problem faced by the investigator. Non-creditworthy responses included incorrect issues such as ecological validity or cultural differences, or simply naming a problem (for example, 'demand characteristics' or 'unethical') with no further elaboration.

© 2025

Question 8

(a) Responses to this question covered the full range of marks available, although a notable proportion of responses (approximately 17 per cent) did not attempt the question. Strong responses demonstrated good knowledge of both the Yale model of communication and a study investigating types of advertising media. For the Yale model, Level 3 responses typically identified several of the five features and provided a brief outline of each, clearly linked to advertising. Weaker descriptions (Level 1 or Level 2) tended to identify fewer features, often focusing only on the channel or the audience, or simply listing features without explanation.

For the study on types of advertising media, the majority of responses that gained credit referred to Ciceri et al. Stronger responses outlined key features of the procedure, including the type of advertising media used and the task or measurement involved, and included at least one relevant result. Some responses included more detail than was required, such as sample characteristics or multiple results, although this was not necessary to achieve higher levels. Weaker responses often included inaccurate or incomplete procedural detail, omitted results, or failed to describe how advertising media were investigated. Non-creditworthy responses included outlining different types of advertising without reference to a study or describing studies that investigated only one type of advertising medium, such as Auty and Lewis or Fischer, which did not meet the requirements of the question.

(b) Marks for this part of the question covered the full range available, although the most common levels awarded were Level 1 and Level 2. A notable proportion of responses (approximately 27 per cent) did not attempt this question, which is likely to be linked to the weaker performance on **part (a)**. Stronger responses (Level 3 and above) were clearly structured issue by issue and included evaluation of both the Yale model of communication and a study on types of advertising media, most commonly Ciceri et al. In addition to the named issue of objective and subjective data, higher-level responses often discussed generalisability, ecological validity and application to everyday life. Some responses were able to explain effectively how Ciceri et al. included objective measures such as eye-tracking, EEG and recognition data, while also considering whether these fully captured subjective experiences such as frustration. Others applied the named issue to the Yale model by discussing the subjective nature of factors such as audience characteristics or message effectiveness.

Weaker responses (Level 1 or Level 2) often failed to apply issues and debates to the content in **part (a)**, beyond simply identifying the Yale model of communication or Ciceri et al. Some responses gave brief or generic definitions of issues such as objectivity or validity without contextualisation. Other non-creditworthy responses simply stated which side of a debate each approach supported, for example claiming that Ciceri et al. is reductionist or that the Yale model is holistic, without explanation or reference to advertising or the study, which did not receive credit.

Health Psychology

Question 9

Responses to this question covered the full range of marks available. Most responses were able to suggest two reasons for Chloe's unrealistic optimism, although it was common for responses to gain only 1 mark per reason where suggestions were outlined briefly. Popular reasons included Chloe's lack of personal or family experience of serious illness, and perceptions of personal control over health outcomes. Stronger responses achieved 2 marks per reason by extending these ideas with appropriate context or examples, such as explaining specific health behaviours Chloe believes will reduce her risk of illness or increase her life expectancy. Some responses also referred to stereotypes about who is most likely to become ill or to beliefs encouraged by family experiences.

Weaker or non-creditworthy responses often rewrote the scenario in different words, for example restating that Chloe believes other students are more likely to become ill than she is. Other responses explained why Chloe's beliefs are unrealistic, rather than giving reasons for her unrealistic optimism, which did not address the question.

© 2025

Question 10

(a) Responses to this question covered a range of marks, with many achieving full marks. Well-prepared responses clearly outlined both sides of the debate, defining individual explanations in terms of internal or dispositional factors (such as personality or innate characteristics) and situational explanations in terms of external or environmental influences. These responses gained both marks by addressing each explanation explicitly. Candidates were generally more successful in defining situational explanations than individual explanations.

Weaker responses often referred to the wording of the question without further explanation, for example stating that an individual explanation refers to the individual, which was not creditworthy. Some responses outlined only one side of the debate, limiting the marks awarded.

(b) Responses to this question were generally weak, with most achieving 0 or 1 mark. A small number achieved full marks by outlining a clear strategy for improving health, such as fear arousal or providing information, and then explaining why this strategy has a situational explanation. These stronger responses focused on the role of external factors, explaining how the strategy creates an environment that encourages healthy behaviour and does not depend on individual personality or disposition.

Many responses achieved only 1 mark by outlining a strategy for improving health but failing to explain why it is situational. A significant number of responses did not outline any relevant health strategy from the syllabus and therefore could not explain why it had a situational explanation. Very few responses referred to strategies for managing stress, despite these being included in the syllabus. Responses that identified a health strategy on its own, without explanation, did not receive credit.

Question 11

(a) Responses to this question covered a range of marks, with many achieving full marks. Strong responses were able to suggest two different measures of pain and provide clear detail about how each measure works. The most common measures described were the clinical interview, visual analogue scales, and the UAB pain behaviour scale. Good responses often named the measure explicitly, such as the McGill Pain Questionnaire, and included one or more relevant features explaining how pain is assessed.

Responses that achieved 1 mark typically did so by naming or briefly outlining a pain measure without providing sufficient detail about how it operates, or by describing the procedure without clearly identifying the measure used. Weaker or non-creditworthy responses included vague references to self-report or observation without linking these to a recognised pain measure from the syllabus or included inaccurate details about how questionnaires or observational scales are scored. It should be noted that biological measures such as blood or urine tests, or scans such as fMRI, do not measure pain and are not creditworthy.

(b) Responses to this question covered the full range of marks available. Stronger responses achieved full marks by clearly identifying a weakness of a pain measure and linking this explicitly to the specific measure suggested in **part (a)**. Common creditworthy weaknesses included lack of depth, differences in how individuals interpret scales or descriptors, the subjective nature of pain, and difficulty accurately observing pain behaviour when pain is not outwardly expressed.

Responses that achieved only 1 mark were often brief and failed to refer clearly to the pain measure suggested in **part (a)**, limiting development of the weakness. Non-creditworthy responses commonly identified only the type of data (for example stating that the measure used quantitative data) without explaining why this was a weakness. Other non-creditworthy responses included discussing weaknesses of pain treatments rather than pain measures or giving a weakness of a biological measure where such a measure had been incorrectly suggested in **part (a)**.

© 2025

Question 12

(a) Responses to this question covered the full range of marks available, although overall performance was weak and a notable proportion of responses (approximately 12 per cent) did not attempt the question. Where responses did gain marks, these were most commonly in Level 1 or Level 2. The vast majority of responses were able to define false positive and false negative diagnoses accurately, and many gave appropriate health-related examples of each. Some stronger responses also referred to difficulties with patient disclosure, including reference to research such as Robinson and West, explaining that patients may disclose more information in non-face-to-face formats, particularly when information is sensitive or embarrassing.

Stronger responses were relatively rare but typically included relevant research evidence linked directly to making or presenting a diagnosis. A small number of responses outlined Savage and Armstrong, which was credited, as this study relates to practitioner behaviour following diagnosis and sits within the same syllabus topic area. However, many responses lacked focus on diagnosis and instead included anecdotal descriptions of what a diagnosis is, reasons why doctors may make mistakes, or suggestions for improving communication. These responses often contained lengthy writing but did not address the specific syllabus content required and therefore received no credit.

(b) Marks for this part of the question covered the full range available, although overall performance was very weak and a substantial proportion of responses (approximately 22 per cent) did not attempt the question. A very small number of responses achieved Level 3 by structuring their evaluation issue by issue and applying it directly to the content in **part (a)**. These responses addressed the named issue of validity, most commonly discussing ecological validity and generalisability, for example by considering the extent to which findings based on specific patient groups (such as those diagnosed with skin cancer) could be applied more broadly.

Many responses achieved Level 1 or Level 2. These responses often demonstrated limited understanding of the requirements of the question and failed to evaluate the material in **part (a)** beyond general statements.

Common issues and debates selected included application to everyday life and generalisations from findings, but these were frequently described without sufficient contextualisation. The named issue of validity was often poorly addressed, with responses either defining validity in general terms or failing to apply it to research into making or presenting a diagnosis. As with **part (a)**, responses that had selected incorrect research were unable to access credit here, which further limited performance.

Organisational Psychology

Question 13

Responses to this question covered the full range of marks available, with a number of strong responses achieving full marks. Stronger responses demonstrated a clear understanding of the stages of group development and were able to apply these effectively to the context of the school council. These responses typically outlined three or more stages and linked most stages to relevant behaviours such as establishing rules, conflict between members, developing shared norms, and working cooperatively to make decisions about issues such as the school menu or recycling facilities.

Weaker responses often showed partial knowledge of group development but struggled to accurately describe the stages or match the correct stage name to the behaviours involved. Some responses listed stage names with little or no explanation, while others gave generic descriptions of group work without reference to specific stages or the school council context. These responses were typically awarded 1 or 2 marks. A small number of responses confused the stages entirely or gave anecdotal accounts of group decision-making that did not clearly reflect the group development model and were therefore not creditworthy.

© 2025

Question 14

(a) Most responses were able to achieve at least 1 mark by giving an accurate definition of the nomothetic approach, typically referring to the aim of identifying general laws of behaviour. Some responses were able to gain the second mark by applying this approach to a need theory of motivation. Stronger responses clearly linked the nomothetic approach to a named theory, most commonly Maslow's hierarchy of needs. Weaker responses often stopped after defining the nomothetic approach and did not include an example from a need theory, limiting them to 1 mark. Some responses focused on the quantitative nature of nomothetic research which made it difficult to link to a need theory. A small number of responses outlined Maslow's or McClelland's theory accurately but failed to link this to the idea of general laws of behaviour, which again restricted the mark awarded.

(b) Performance on this question was generally weak, with relatively few strong responses seen. A small number of responses achieved full marks by clearly outlining how an idiographic approach could be used to investigate motivation at work and linking this explicitly to a need theory. These responses often described methods such as interviews and explained how these would explore an individual employee's experiences of needs. Where this link was made, full marks were awarded. Some responses achieved only 1 mark by outlining an idiographic method, most commonly interviews, without linking this to a need theory of motivation. Other responses attempted to describe an idiographic approach but did not apply it to motivation at work or instead referred to general job satisfaction or workplace behaviour. In line with the mark scheme, questionnaires were only creditworthy where candidates made it clear that open-ended questions would be used. Responses that described need theories without reference to an idiographic method, or that confused idiographic and nomothetic approaches, were not creditworthy.

Question 15

(a) Responses to this question covered a range of marks, with some good answers achieving full marks. Stronger responses clearly identified two different extrinsic motivators and applied them effectively to the context of the carpet factory. The most common extrinsic motivators used were bonus payments and performance-related pay, with responses explaining how linking financial rewards to output could increase motivation and productivity. A small number of responses also referred to pay increases, which were creditworthy when clearly distinguished from other reward systems. Weaker responses often failed to apply their suggestions to the factory context, limiting the marks awarded. Some responses named an extrinsic motivator without explaining how it would operate in practice, resulting in 1 mark. A common error was to describe the same extrinsic motivator twice using slightly different wording, which restricted the total marks available. Other responses confused extrinsic and intrinsic motivation or gave general statements about motivating workers without reference to a reward system, which were not creditworthy.

(b) Performance on this question was mixed. Some responses were able to achieve full marks by identifying a clear strength of an extrinsic motivator other than increased production and explaining why this benefit would occur. Stronger responses referred to increased job satisfaction, a greater sense of fairness, or increased commitment to the factory, and explained how financial rewards could make workers feel valued or invested in the organisation. These responses demonstrated clear understanding and achieved 2 marks. Many responses achieved only 1 mark because the strength was identified but not sufficiently explained. These responses often stated that workers would be more satisfied or motivated without explaining why the suggested reward system would lead to this outcome. A common error was to focus on increased productivity, which was explicitly excluded by the question and therefore not creditworthy. Some responses also made unsupported claims, such as suggesting that bonus payments would automatically increase loyalty to the factory, without explanation. Overall, relatively few responses developed their points fully enough to access both marks.

© 2025

Question 16

(a) Responses to this question covered a range of marks, with some good descriptions achieving Level 3. Stronger responses demonstrated clear understanding of the difference between rapid rotation and slow rotation shiftwork, often referring to the frequency of shift changes and the impact on circadian rhythms. These responses accurately explained why rapid rotation is more disruptive and why slow rotation allows greater physiological adjustment. In addition, better responses included a clear outline of a relevant study, most commonly Gold et al., with brief procedural details and at least one appropriate finding relating to sleep disruption, accidents, or health outcomes.

Weaker responses often gave a very general outline of shiftwork without clearly distinguishing between rapid and slow rotation, or they defined one type without comparison to the other. Some responses failed to identify the types by name, which limited the level awarded. A number of responses either did not attempt the study element or gave a very brief or vague description without sufficient detail of the procedure or results, restricting them to Level 1 or Level 2. In some cases, responses appeared incomplete, suggesting that candidates may have run out of time.

(b) The marks for this question tended to be between level 1 and level 3. A small number of candidates produced strong evaluative responses to this question. These responses addressed the named issue of quantitative and qualitative data effectively and linked their discussion clearly to rapid and slow rotation shiftwork and to research such as Gold et al. The majority of research cited in this area, including Gold et al., relies heavily on quantitative data, for example questionnaire responses and accident rates. Strong responses recognised that quantitative data allows for easier comparison, the calculation of percentages and means, and the identification of patterns across large samples.

Other commonly used evaluative issues included application to everyday life and the use of questionnaires. In addition, weaker (Level 1 or Level 2) responses lacked evaluation points that provided any depth of explanation. Each point tended to be very brief, and reference was frequently made to the content in **part (a)** but this was merely identifying which side of the debate the theory was on or whether it did have applications or quantitative data but only occasionally was this explained. Sometimes analysis might be given but this was also just identified and not explained. Approximately 10 per cent of candidates did not attempt this question.

© 2025

PSYCHOLOGY

Paper 9990/32

Specialist Options: Approaches, Issues and Debates

Key messages

Questions 1, 3(a), 5, 7(a), 9, 11(a), 13(a), 15(a)

These questions in this exam asked candidates to apply an area of the syllabus (theory, treatment/symptoms/explanation of a disorder, self-report/measuring techniques etc.) to explain how it can be applied to a particular scenario or context. It is important that candidates are aware of the titles of the bullet points in the syllabus. It would be helpful for candidates to do revision notes with the title of the topic area and bullet point at the top so that they can identify which part of the syllabus these types of questions are referring to. Candidates should also refer directly to the scenario/context in the question in their response.

Questions 3(b), 7(b), 11(b) and 15(b)

These questions in this exam asked candidates to evaluate with a weakness of the theory/treatment/measurement that was outlined in the candidate's response to **part (a)** of the question. In this exam, this type of question also asked the candidate to outline one problem psychologists have when investigating a theory with participants. It would be helpful for candidates when doing revision to learn strengths and weaknesses of the theories, techniques, explanations, treatments, etc. they have learned and put these into their revision notes. In addition, candidates should learn problems psychologists might have when investigating what has been outlined in **part (a)**. They should also practice explaining the evaluation point in the context of the question.

Questions 2, 6, 10 and 14

Part (a)

These questions in this exam asked the candidate to outline one of the issues and debates, sometimes with an example from the syllabus content. This part of the question could also ask the candidate to outline a theory, explanation, technique/treatment or self-report used by psychologists that is named in the syllabus. The revision technique outlined previously in this report will aid candidates to learn the syllabus material.

Part (b)

This part of the question may ask candidates to explain a strength or a weakness of the issue/debate or the syllabus content outlined in **part (a)**. The question could also ask candidates to explain how a bullet point in the syllabus links to or supports one of the issues or debates. It would also be useful for candidates to write revision notes where they define the issues/debates and prepare a strength and a weakness of each issue and debate to prepare for the **part (b)** of this type of questions. Candidates should also note how the topics covered in the syllabus fit with each of the issues/debates. These questions in this exam were worth 2 marks for each part of the response and therefore a short response is appropriate.

Questions 4(a), 8(a), 12(a) and 16(a)

These questions in this exam came from one or two of the bullet points in the syllabus. This exam asked the candidate to describe a study, two treatments, three explanations and two theories (and/or studies). For this exam, some of the responses used the incorrect topic area in the syllabus, the description was brief or lacked focus on the topic area (such as how the treatments treat anxiety disorders and fear-related disorders rather than mental health problems in general).

© 2025

It could be useful for candidates to create revision notes with the title of each topic area and the description in the bullet point as the header. Alternatively, candidates could create a mind map and put this information in the centre. Some of the responses were too long for the first option (e.g. **Question 4(a)**). These types of responses did usually receive a high mark. However, this often meant that the candidate did not have enough time to write a complete response for the second option (e.g. **Question 12(a)**). It would be useful for candidates to practice writing an appropriate length response to these types of questions. Teachers could set a word limit or do a timed response at home or in class. These should be done without referencing any notes or the textbook while doing the timed response.

Questions 4(b), 8(b), 12(b) and 16(b)

This question will always ask the candidate to evaluate the studies, theories, explanations or treatments of disorders or techniques described in **part (a)** of the question. The response must include at least two evaluation issues, including the named issue, to be considered to have presented a range of issues to achieve the top band. However, most responses that evaluated using two issues in this exam, achieved in the lower bands due to the response being superficial and often with little analysis. Some responses that considered three issues tended to achieve higher marks as these responses were able to demonstrate comprehensive understanding with good supporting examples from the studies, theories, characteristics/explanations/treatments of disorders or techniques described in the **part (a)** of the answer.

The candidate must also provide some form of analysis to access Level 3 and above. This could be done by discussing the strengths and weaknesses of the issue being considered, presenting a counterargument to the issue under discussion or comparing the issue between two studies and/or theories. The response needs to explain the comparison/strength/weakness or counterargument with examples from **part (a)** of the question. It was common for responses to state that two theories or treatments for example, were 'similar' or 'in contrast' for an issue without any explanation as to why they could be compared in this way. This is assessed as little analysis. It could be limited analysis if the response gives several contextualised examples to support the evaluation points. A conclusion at the end of each issue would be helpful to show excellent understanding of the issue under discussion. To achieve the requirements of the level 4 and 5 descriptors, it would be best to structure the response by issue rather than by study and/or theory. It would also be ideal for the response to start with the named issue to make sure the answer covers this requirement of the question.

A small minority of candidates did not evaluate using the named issue. Quite a few of the answers were structured by study/theory/treatment rather than by the issue which often led the response to be quite superficial and repetitive. Some of the responses did do analysis. Candidates should be aware this question is worth 10 marks and need to include an appropriate amount of information.

General comments

The marks achieved by candidates for this session of the 9990 specification achieved across the full range of the mark band which was very pleasing to see. Some candidates were well prepared for the exam and showed good knowledge, understanding, application and evaluation throughout their responses. A significant number of candidates were not as well prepared and showed limited knowledge and understanding with brief, superficial and sometimes anecdotal responses. These candidates often had limited evaluation and application skills.

Time management for this paper was good for the majority of candidates and most attempted all questions that were required. A number of candidates did not respond to one or more of the questions asked in the option area. A very small number of the candidates attempted to respond to more than two topic areas but often did not attempt all the questions for each option chosen. These responses achieved at the lower end of the mark band.

The questions on clinical were the more popular choice of option, followed by health.

© 2025

Comments on specific questions

Clinical Psychology

Question 1

The responses to this question covered the full range of marks. Many were able to suggest at least one difference between depressive disorder (unipolar) and bipolar disorder, most commonly referring to the presence of manic episodes in bipolar disorder. However, a large number of responses only described symptoms of bipolar disorder without making an explicit comparison with unipolar depression, which limited these answers to 1 mark. Higher-mark responses clearly contrasted the two disorders, for example by explaining that bipolar disorder involves fluctuations between mania and depression, whereas unipolar depression involves a persistent low mood with no manic episodes. Another common difference outlined was the difference in symptoms such as feeling very elated/reduced need for sleep during the manic phase whereas unipolar often has an increased need for sleep which remains consistent. Weaker responses either listed symptoms without linking them to the correct disorder or confused the two conditions, and a small number referred to symptoms of other disorders, which was not creditworthy.

Question 2

(a) Most responses achieved at least 1 mark for this question by giving a basic definition of learned helplessness, usually referring to a perceived lack of control or people believing they cannot change their situation. A significant number were also able to gain full marks by clearly linking learned helplessness to depressive disorder, explaining that when individuals believe their actions will not improve outcomes they stop trying, which can lead to withdrawal, low motivation and/or depressed mood. Some responses effectively used Seligman's dog study to illustrate learned helplessness and then applied this to depression. Weaker responses tended to describe symptoms of depression rather than the explanation itself or defined learned helplessness without explaining how it accounts for depressive disorder. A small number simply reused wording from the question, such as stating that people 'cannot help themselves,' which was not creditworthy.

(b) Responses to this question covered the full range of marks. Many responses achieved 1 mark by showing a basic understanding of what is meant by an individual explanation, often by referring to thoughts, perceptions or personal experiences. Stronger responses clearly linked this to learned helplessness, explaining that it is the individual's own perception of lack of control, rather than the situation itself, that leads to depressive feelings. These responses often contrasted two people in the same situation responding differently, which demonstrated clear understanding and achieved full marks. A common response that did not receive credit was to outline the situational explanation, frequently focusing on environmental factors or life events rather than the individual's interpretation of these experiences.

Question 3

(a) Many responses were able to gain at least 2 marks by identifying a relevant reason why atypical antipsychotics may be suitable, most commonly that they have fewer or less severe side effects than typical antipsychotics or that they treat both positive and negative symptoms of schizophrenia. Stronger responses achieved 3–4 marks by developing these reasons with accurate detail, such as naming specific side effects associated with typical antipsychotics (for example tardive dyskinesia) or giving examples of negative symptoms that atypical antipsychotics may help to alleviate. Some responses also explained why this made atypical antipsychotics a better choice than typical antipsychotics. Some responses stated reasons very briefly, for example simply saying that atypical antipsychotics are 'more effective', without explanation, or gave reasons that applied equally to typical antipsychotics, such as blocking dopamine. Neither of these types of responses were creditworthy. A small number of responses contained inaccuracies about how quickly atypical antipsychotics work or confused the differences between typical and atypical drugs which did not receive credit.

(b) Responses to this question covered the full range of marks available. The most common weakness identified was the presence of side effects, with stronger responses gaining full marks by naming specific side effects such as nausea and drowsiness. Others explained how side effects could lead to poor adherence or relapse if medication is stopped. Some responses also explained that antipsychotics are a passive treatment or that tolerance can develop over time, leading to breakthrough symptoms. Weaker responses tended to identify a weakness without further

© 2025

explanation, often achieving only 1 mark. Non-creditworthy responses included references to cost, claims that medication cures schizophrenia, or vague statements about not addressing the root cause without further elaboration.

Question 4

(a) Responses to this question covered the full range of marks. Level 3 responses were able to describe both cognitive-behavioural therapy (CBT) and applied tension clearly and accurately, often outlining several components of CBT such as identifying faulty thinking and cognitive restructuring and how this can help treat fear-related disorders (with a phobia commonly used as an example). The best responses also demonstrated good understanding of applied tension by explaining how muscle tensing is used to increase blood pressure and prevent fainting in blood-injection-injury phobia, with some effectively referencing the Chapman and DeLapp study to support their description. Many responses achieved Level 2 marks by giving generally accurate but less detailed descriptions, often focusing more strongly on CBT than applied tension. Weaker responses commonly confused applied tension with relaxation techniques, gave very brief outlines such as simply stating that muscles are tensed, or described unrelated therapies such as systematic desensitisation. Some responses also applied CBT to other disorders which limited credit.

(b) Marks for this part of the question covered the full range available, although the most common levels awarded were Level 1 and Level 2. A notable proportion of responses (approximately 7 per cent) did not attempt this question.

Stronger responses (Level 3) were clearly structured issue by issue and evaluated both CBT and applied tension, using the named issue of generalisability. These responses typically explained that CBT is highly generalisable as it can be applied to a wide range of anxiety and fear-related disorders, whereas applied tension is more limited as it is specifically designed for blood-injection-injury phobia, often supporting this with reference to the Chapman and DeLapp case study. Higher-level responses also discussed additional issues such as idiographic versus nomothetic approaches, nature versus nurture, and determinism versus free will, with effective contextualisation to both treatments.

Many responses achieved Level 1 or Level 2 by identifying relevant issues but offering limited analysis or failing to apply them consistently to both treatments. Weaker responses tended to provide generic evaluations, define issues without linking them to CBT or applied tension, or focus on strengths and weaknesses without reference to any relevant issues or debates.

Consumer Psychology

Question 5

Many responses were able to suggest at least one appropriate effect of a pleasant odour on shoppers, most commonly that it would make shoppers feel happier or more satisfied, leading them to spend longer in the mall or spend more money. Stronger responses achieved full marks by clearly identifying two distinct effects and linking each to shopper behaviour, for example explaining that a pleasant odour could increase pleasure or arousal, resulting in longer time in store, more store visits, or higher spending/more purchases. Weaker responses often repeated the same effect twice, only described the odour as pleasant (which was already given in the question), or failed to explain how the odour would influence shopper behaviour. Some responses mentioned only one effect or did not make the link to shoppers' actions clear, which limited the marks awarded.

Question 6

(a) Responses achieved full marks by giving a clear definition of the nomothetic approach, typically referring to the use of quantitative data or establishing general laws of behaviour, and then linking this to an appropriate example from the effect of sound on consumer behaviour. Commonly, responses referred accurately to findings from studies such as North et al., explaining how classical music influenced spending behaviour, or to the Woods et al. study on how sound and noise affect the perception of food taste. Some responses achieved only 1 mark by providing an accurate definition of the nomothetic approach but failing to link it clearly to the effect of sound on consumer behaviour, or by describing an effect of sound without making it clear how this illustrates a nomothetic approach. Non-creditworthy responses included vague or inaccurate definitions of

© 2025

nomothetic, confusion with idiographic approaches, or invented or unsupported examples that were not clearly related to sound and consumer behaviour.

(b) Many responses achieved 1 mark by identifying a general strength of the nomothetic approach, most commonly that it allows findings to be generalised to larger populations. Stronger responses achieved full marks by clearly explaining why this is a strength, for example by linking large sample sizes to increased generalisability, predictive value, or the scientific credibility of psychology. These responses often explained how establishing general laws can make psychological findings more useful in practice. Weaker responses tended to be vague, such as stating that the nomothetic approach can be applied to the real world without further explanation. Some responses incorrectly identified the use of quantitative data as a strength in itself which was not creditworthy, rather than explaining that this allows for comparisons to be made. 12 per cent of candidates who did the consumer option did not attempt this question.

Question 7

(a) Many responses demonstrated a sound understanding that food names can influence customer choice, with common suggestions including using more descriptive, evocative, or familiar names to make dishes sound more appealing. Stronger responses (3–4 marks) clearly applied this knowledge, often giving specific examples of food names and explaining how these would affect customers' perceptions or behaviour, such as being more likely to order a dish or perceive it as tastier or higher quality. These responses were sometimes supported by reference to research such as Lockyer or Wansink et al. Many responses achieved 1–2 marks by giving a basic suggestion, such as making food names sound more attractive, but lacked detailed explanation or clear context. A common non-creditworthy response focused on unrelated aspects of menu design (such as layout or fonts).

(b) Responses to this question covered a range of marks. Stronger responses were able to clearly explain one problem with carrying out research into the effect of food names on menu item choice and explicitly link this problem to how food names influence consumer behaviour. Common creditworthy answers included issues of generalisability, for example explaining that findings from one type of restaurant, culture, or cuisine may not apply to others, or that individual differences in taste preferences could make it difficult to isolate the effect of the food name itself. Other good responses referred to methodological problems such as demand characteristics or difficulties collecting valid self-report data from diners. Non-creditworthy responses often identified a problem in name only, such as stating 'individual differences' or 'social desirability bias' with no further elaboration.

Question 8

(a) Responses to this question covered the full range of marks available, although a notable proportion of candidates (approximately 13 per cent) did not attempt the question. This was the weakest-performing of the four 6–mark questions on the paper. Level 3 responses demonstrated clear knowledge and understanding of compensatory, non-compensatory and partially compensatory decision-making strategies, accurately using psychological terminology such as equal weight, weighted additive, satisficing, elimination by aspects and lexicographic strategies, often supported with relevant consumer examples. Many candidates achieved Level 1 marks by outlining decision-making strategies in general terms or by providing examples (such as buying cars or phones) without clearly identifying or explaining the specific types of strategies involved. Weaker responses frequently lacked accurate terminology, failed to distinguish between the different strategies, or described consumer decision-making very broadly without reference to compensatory or non-compensatory processes. A small number of responses were largely irrelevant which was not creditworthy, indicating limited understanding of consumer decision-making strategies.

(b) Responses to this question covered the full range of marks, although many were awarded Levels 1 or 2. Stronger responses evaluated the consumer decision-making strategies using relevant issues and debates and applied these effectively to everyday contexts. Good answers addressed the named issue of application to everyday life in an evaluative way, for example by explaining how understanding compensatory or non-compensatory strategies could help businesses design websites or product comparisons to match consumer decision processes. These responses often went on to consider limitations (which provided analysis), such as whether consumers consistently use the same strategy across different types of purchases or situations. Other common issues included determinism versus free will and idiographic versus nomothetic approaches.

© 2025

Weaker responses tended to describe the strategies or give everyday examples without evaluation, or listed issues and debates without applying them to the strategies. Non-creditworthy responses simply stated which side of a debate each strategy supported, for example, just stating that one or more of the strategies has good or poor applications to everyday life, without explanation of why this is the case.

Health Psychology

Question 9

Responses to this question covered the full range of marks available. Some responses did suggest at least one appropriate prompt, most commonly automated text messages or phone calls, and these often gained 1 mark where the idea was outlined briefly. Stronger responses achieved full marks by giving two clearly different prompts and describing them in concrete detail, such as specifying the wording of a text message, the time it would be sent, or how a message or sticker could be placed on medication packaging to act as a daily reminder. These responses demonstrated a clear understanding of how prompts can be used in a practical way. Weaker responses often identified a correct type of prompt but failed to add any further detail, limiting marks. Some non-creditworthy responses suggested techniques that were not prompts, such as fear arousal, monitoring medication levels, or aspects of practitioner behaviour, which did not address the question.

Question 10

(a) Many responses were able to gain at least 1 mark by outlining the free-will side of the debate, often explaining that behaviour is the result of personal choice or individual control. However, understanding of determinism was generally weaker. A common error was to suggest that determinism means behaviour is partly controlled or that individuals have limited control, which does not accurately reflect the deterministic position that behaviour is caused by internal or external forces beyond conscious control. Full mark responses clearly distinguished between the two sides of the debate, accurately explaining that determinism proposes behaviour is governed by biological, environmental or psychological factors, whereas free will suggests individuals actively choose how they behave. Weaker responses often described only one side of the debate, most commonly free will, or gave vague statements such as behaviour being ‘influenced’ by factors without explaining the lack of control central to determinism.

(b) Responses to this question varied in quality, with many achieving 1 mark by identifying an appropriate way to manage stress, such as biofeedback, imagery, or stress inoculation training. 2-mark responses clearly explained why free will is important in the chosen technique, for example by emphasising that the individual must actively choose to engage with the method, control their responses, or practise the technique to reduce stress. Biofeedback and imagery were the most used examples in higher-mark responses, often linked effectively to personal control. Weaker responses tended to discuss free will in a general sense without linking it to a specific stress management technique or described stress avoidance strategies (such as leaving a situation) which were not creditworthy.

Question 11

(a) Responses to this question covered a wide range of marks, although overall performance was weak, and a notable proportion of candidates (approximately 17 per cent) did not attempt the question. Stronger responses demonstrated a clear understanding of the General Adaptation Syndrome (GAS) model and were able to describe the alarm reaction, resistance and exhaustion stages accurately while applying them directly to Lydia’s situation. These responses effectively linked Lydia’s new job and caring responsibilities to prolonged stress exposure, explaining symptoms such as irritability, anxiety and fatigue using appropriate physiological and psychological features of each stage. Some responses were able to gain 1–2 marks by outlining one or two stages of the GAS model, most commonly the alarm or exhaustion stage, but often lacked sufficient detail or failed to apply the stages clearly to Lydia’s symptoms. Weaker or non-creditworthy responses showed little or no knowledge of the GAS model, sometimes confusing it with measures of stress or giving general descriptions of stress without reference to the model.

(b) Responses to this question covered a limited range of marks, with many responses achieving only 1 mark. Better, 2-mark responses were able to explain a clear weakness of the GAS model and apply it directly to stress, for example by explaining that the model is deterministic or reductionist

© 2025

and illustrating how it fails to account for individual differences in stress responses or the complexity of multiple ongoing stressors. Some responses effectively linked this to the idea that not everyone progresses through the stages in the same way or experiences the same outcomes. However, many weaker responses simply stated a weakness such as 'it is reductionist' or 'it is deterministic' without any explanation, which was not creditworthy. Several responses did not know the GAS model and instead stated a weakness of the measure of stress they had incorrectly outlined in **part (a)** which was not creditworthy. As with **part (a)**, a notable proportion of candidates (approximately 17 per cent) did not attempt this question.

Question 12

(a) Responses to this question covered the full range of marks. Strong responses (Level 3) were able to describe acute, chronic and phantom limb pain clearly and accurately, often distinguishing between them in terms of duration, severity and treatability, and supporting these descriptions with appropriate examples. Phantom limb pain was generally well understood, with some responses accurately outlining mirror therapy as a way of reducing pain following limb loss. Stronger responses also demonstrated good knowledge of the gate control theory, using appropriate psychological terminology such as small and large nerve fibres, the spinal 'gate', and explaining how touch or emotional factors can reduce perceived pain. Many responses achieved Level 2 by giving generally accurate but brief descriptions, particularly of acute and chronic pain, often focusing only on duration. Weaker responses showed misunderstandings, such as assuming acute pain is mild, giving inappropriate examples of chronic pain, or providing vague accounts of gate control theory without reference to nerve fibres or the gating mechanism, which limited access to higher marks.

(b) Responses to this question covered the full range of marks, although a notable proportion of candidates (around 13 per cent) did not attempt it. Stronger responses were clearly evaluative and made effective use of the named issue of nature versus nurture, often explaining that acute and chronic pain are largely biologically based (nature), while phantom limb pain includes an environmental or psychological element, particularly through the use of mirror therapy. Better responses also evaluated gate control theory accurately, recognising its strong biological basis in nerve fibres and spinal gating, while also acknowledging that expectations, emotions and behaviours (such as rubbing an injury) introduce a nurture element. In addition to the named issue, some stronger responses discussed debates such as determinism versus free will or reductionism versus holism, usually by linking these explicitly to the theory or types of pain.

Many responses achieved Level 1 or Level 2. These responses often demonstrated limited understanding of the requirements of the question and failed to evaluate the material in **part (a)** beyond general statements. Weaker responses tended to be descriptive rather than evaluative, often outlining what the types of pain or gate control theory are without applying issues and debates. Some non-creditworthy responses evaluated pain itself rather than explanations or theories of pain, while others listed issues with little or no explanation.

Organisational Psychology

Question 13

Responses to this question covered the full range of marks available. Many responses were able to suggest at least one clear effect of moving from an open-plan office to individual offices, with common creditworthy effects including increased privacy, reduced noise levels and improved concentration. Stronger responses achieved full marks by clearly explaining how these changes would affect workers, for example by linking quieter conditions to improved productivity or job satisfaction, or by explaining how greater autonomy could increase motivation. Some responses also effectively identified potential negative effects, such as reduced communication or weaker teamwork, and explained how internal walls could limit informal interaction between colleagues. Weaker responses often stated an effect, such as 'better concentration' or 'more privacy', without providing any explanation or detail, which limited marks. A small number of responses repeated the same effect twice which was not credited for the second suggestion.

Question 14

(a) Responses to this question were mixed, with many achieving 1 mark but fewer gaining full marks. Most responses were able to outline holism reasonably clearly, typically referring to understanding behaviour by looking at the whole person or the wider context. In contrast, reductionism was less

© 2025

well understood. A common error was to state that reductionism only looks at one factor or ignores other factors, without explaining that it involves breaking behaviour down into simpler component parts. As a result, many responses provided only partial definitions of both terms and were limited to 1 mark. This question was one of the weaker performing issues questions on the paper, highlighting the need for clearer learning of precise definitions, particularly for reductionism.

(b) Many responses to this question were limited to 1 mark. Most responses were able to identify a basic strength of taking a holistic approach, such as considering the whole person rather than isolated factors, which was sufficient for the first mark. Stronger responses went on to explain this strength clearly in the context of shiftwork by recognising that health outcomes and accidents may be influenced by multiple interacting factors, including lifestyle, work demands, and non-work stressors. These responses often used relevant examples, such as the role of sleep, diet, or family responsibilities alongside shift patterns, and achieved full marks. Some responses just reworded the question which was not creditworthy.

Question 15

(a) Responses to this question covered a range of marks, with some strong answers achieving full marks. Better responses clearly identified two appropriate reasons for using the Job Descriptive Index (JDI) and applied these effectively to the context of a large school with a range of staff roles. Common creditworthy reasons included that the JDI is a self-report questionnaire that is easy to administer and complete, and that it is comprehensive, measuring several specific aspects of job satisfaction such as supervision, pay and co-workers. Stronger responses often supported these reasons with relevant examples from the JDI or explained why these features made it suitable for both teaching and non-teaching staff. Non-creditworthy responses did not know the JDI and sometimes just stated that it provides a 'job description' or made unsupported claims about high validity or reliability.

(b) Performance on this question was mixed. 2-mark responses were able to identify a clear weakness of the Job Descriptive Index and explain it with explicit reference to how the JDI operates. These responses mainly focused on the lack of qualitative data, explaining that the forced-choice format (yes / no / can't decide) limits the depth of information that can be gathered about workers' true feelings. Where responses illustrated this with an example, such as staff being unable to explain *why* they were unhappy with supervision or promotion prospects, full marks were awarded. It was also common for candidates to state that the JDI lacked validity or reliability without any supporting explanation, which was not creditworthy. Weaker responses tended to give generic weaknesses of questionnaires or self-report measures without linking these to the content of the JDI. Candidates who did not demonstrate clear knowledge of the JDI in **part (a)** often struggled to apply an appropriate weakness in this question.

Question 16

(a) Responses to this question covered a wide range of marks. Stronger, Level 3 responses demonstrated clear and accurate knowledge of the study by Claypoole and Szalma and were able to describe key elements of the research in sufficient detail. These responses typically included appropriate details of the sample, clearly identified the independent variable (electronic performance monitoring versus no monitoring) and a dependent variables (performance on the sustained attention task or correct detections, false alarms or response time). Better answers also outlined the procedure, including how participants were monitored, the nature and duration of the sustained attention task, and at least one result, often with accurate numerical data. Where these elements were combined, responses were able to access full marks. Many responses demonstrated a reasonable understanding of the study but lacked detail. These responses were often accurate but incomplete, for example stating that performance was better in the electronically monitored condition without describing how performance was measured or omitting any procedural details.

Weaker responses showed limited knowledge of the study. Some contained inaccuracies, such as incorrect sample size, describing the study as a repeated measures design, or suggesting it was conducted in a real workplace rather than a laboratory which was not creditworthy. Other responses focused heavily on the aim of the study at the expense of describing the sample, procedure and results, which restricted the marks available. Several responses did not include any results or gave very vague statements that performance 'improved' under electronic monitoring

© 2025

with no further detail. A significant proportion of candidates (approximately 21 per cent) did not attempt this question.

(b) Responses to this question covered a range of marks, though the majority were awarded Level 1 or Level 2. Stronger responses addressed the named issue of controls clearly and accurately, most commonly recognising the use of a control condition and explaining how this allowed for comparison with the electronically monitored group. Better answers also referred to specific procedural controls, such as the standardised sustained attention task and linked this to increased reliability of the findings. Application to everyday life was another commonly used evaluative issue. Stronger responses contextualised this by linking findings from the study to real-world uses of EPM, such as monitoring concentration in factory or office environments, often using a result from the study to support this point. Other issues used included determinism versus free will and individual versus situational explanations, though these were often only partially developed.

Weaker, Level 1 or 2 responses, tended to make very general evaluative statements, such as noting that the study was ‘controlled’ or ‘reliable,’ without explaining how or why this was the case in relation to Claypoole and Szalma’s methodology. Some candidates attempted to cover too many issues superficially, which limited depth and access to higher levels. Several responses were not clearly contextualised to the study which also limited the marks to Level 1 or 2. Approximately 27 per cent of candidates did not attempt this question, which significantly limited overall performance.

© 2025

PSYCHOLOGY

Paper 9990/33

Specialist Options: Approaches, Issues and Debates

Key messages

Questions 1, 3(a), 5, 7(a)(i), 7 (a)(ii), 9, 11(a), 13(a), 15(a) –

These questions in this exam asked candidates to apply an area of the syllabus (theory, treatment/explanation of a disorder, technique, self-report/measuring techniques etc.) to explain how it can be applied to a particular scenario or context. It is important that candidates are aware of the titles of the bullet points in the syllabus. It would be helpful for candidates to do revision notes with the title of the topic area and bullet point at the top so that they can identify which part of the syllabus these types of questions are referring to. Candidates should also refer directly to the scenario/context in the question in their response.

Questions 3(b), 7(b), 11(b) and 15(b) –

These questions in this exam asked candidates to evaluate with a strength or weakness the theory/treatment/measurement that was outlined in the candidate's response to **part (a)** of the question. In this exam, this type of question also asked the candidate to outline one problem psychologists have investigating a theory with participants. It would be helpful for candidates when doing revision to learn strengths and weaknesses of the theories, techniques, explanations, treatments, etc. they have learned and put these into their revision notes. In addition, candidates should learn problems psychologists might have when investigating what has been outlined in **part (a)**. They should also practice explaining the evaluation point in the context of the question.

Questions 2, 6, 10 and 14 –

Part (a) - These questions in this exam asked the candidate to outline one of the issues and debates, possibly with an example from the syllabus content. This part of the question could also ask the candidate to outline a theory, study, technique/treatment or self-report used by psychologists that is named in the syllabus. The revision technique outlined previously in this report will aid candidates to learn the syllabus material.

Part (b) – This part of the question may ask candidates to explain a strength or a weakness of the issue/debate or the syllabus content outlined in **part (a)**. The question could also ask candidates to explain how a bullet point in the syllabus links to or supports one of the issues or debates. It would also be useful for candidates to write revision notes where they define the issues/debates and prepare a strength and a weakness of each issue and debate to prepare for the **part (b)** of this type of questions. Candidates should also note how the topics covered in the syllabus fit with each of the issues/debates. These questions in this exam were worth 2 marks for each part of the response and therefore a short response is appropriate.

Questions 4(a), 8(a), 12(a) and 16(a)

These questions in this exam came from one or two of the bullet points in the syllabus. This exam asked the candidate to describe a study, theories, two treatments and theories (and/or studies) about making and presenting a diagnosis. For this exam, some of the responses used the incorrect topic area in the syllabus, the description was brief or lacked focus on the topic area (such as how the treatments treat depression rather than mental health problems in general).

It could be useful for candidates to create revision notes with the title of each topic area and these description in the bullet point as the header. Alternatively, candidates could create a mind map and put this information in the centre. Some of the responses were too long for the first option (e.g. **Question 4(a)**).

© 2025

These types of responses did usually receive a high mark. However, this often meant that the candidate did not have enough time to write a complete response for the second option (e.g. **Question 12(a)**). It would be useful for candidates to practice writing an appropriate length response to these types of questions. Teachers could set a word limit or do a timed response at home or in class. These should be done without referencing any notes or the textbook while doing the timed response.

Questions 4(b), 8(b), 12(b) and 16(b)

This question will always ask the candidate to evaluate the studies, theories, characteristics/explanations/treatments of disorders or techniques described in **part (a)** of the question. The response must include at least two evaluation issues, including the named issue, to be considered to have presented a range of issues to achieve the top band. However, most responses that evaluated using two issues in this exam, achieved in the lower bands due to the response being superficial and often with little analysis. Some responses that considered three issues tended to achieve higher marks as these responses were able to demonstrate comprehensive understanding with good supporting examples from the studies, theories, characteristics/explanations/treatments of disorders or techniques described in the **part (a)** of the answer.

The candidate must also provide some form of analysis to access level 3 and above. This could be done by discussing the strengths and weaknesses of the issue being considered, presenting a counterargument to the issue under discussion or comparing the issue between two studies and/or theories. The response needs to explain the comparison/strength/weakness or counterargument with examples from **part (a)** of the question. It was common for responses to state that two theories or treatments for example, were 'similar' or 'in contrast' for an issue without any explanation as to why they could be compared in this way. This is assessed as little analysis. It could be limited analysis if the response gives several contextualised examples to support the evaluation points. A conclusion at the end of each issue would be helpful to show excellent understanding of the issue under discussion. To achieve the requirements of the level 4 and 5 descriptors, it would be best to structure the response by issue rather than by study and/or theory. It would also be ideal for the response to start with the named issue to make sure the answer covers this requirement of the question. A small minority of candidates did not evaluate using the named issue. Quite a few of the answers were structured by study/theory/treatment rather than by the issue which often led the response to be quite superficial and repetitive. Some of the responses did do analysis. Candidates should be aware this question is worth 10 marks and need to include an appropriate amount of information.

General comments

The marks achieved by candidates for this session of the 9990 specification achieved across the full range of the mark band which was very pleasing to see. Some candidates were well prepared for the exam and showed good knowledge, understanding, application and evaluation throughout their responses. A significant number of candidates were not as well prepared and showed limited knowledge and understanding with brief, superficial and sometimes anecdotal responses. These candidates often had limited evaluation and application skills.

Time management for this paper was good for the majority of candidates and most attempted all questions that were required. A number of candidates did not respond to one or more of the questions asked in the option area. A very small number of the candidates attempted to respond to more than two topic areas but often did not attempt all the questions for each option chosen. These responses achieved at the lower end of the mark band.

The questions on clinical were the more popular choice of option, followed by health.

Comments on specific questions

Clinical Psychology

Question 1

The responses to this question covered the full range of marks. The vast majority were able to suggest at least one explanation of why Hugh has OCD and many suggested two. The most common explanation given was that Hugh inherited OCD from his mother. However most received 1 mark for this as responses did not give any research evidence and/or an outline of concordance rates where monozygotic twins have higher

© 2025

concordance than dizygotic twins. This was necessary to achieve 2 marks as this is the detail to explain genetic causes of OCD. In addition, higher mark responses also outlined either behavioural or psychodynamic explanations. Some who attempted behavioural simply suggested it was learned from the mother rather than using psychological terminology (such as through positive or negative reinforcement). This type of response as well as the many anecdotal responses, such as being traumatised as a child, were not creditworthy.

Question 2

(a) Most responses achieved at least 1 mark for this question with a significant number achieving full marks. Many responses gave a definition of culture accurately, referring to shared norms, beliefs and values, which was sufficient to gain 1 mark. To achieve the second mark, candidates needed to show understanding of the term *differences*. This was often done successfully by explaining how norms or beliefs vary between cultures or by giving a clear example of a cultural difference (for example, differences in attitudes towards eye contact or family roles). Weaker responses tended to be too vague, such as stating that cultures mental health, without further explanation or exemplification. Some responses used the wording in the question such as 'it is how cultures are different' which was not creditworthy.

(b) Responses to this question covered the full range of marks available. Many responses achieved 1 mark by giving a general reason why cultural differences may affect mental health, although these often made limited reference to anxiety disorders specifically. Stronger responses explained how anxiety disorders may be viewed differently across cultures, sometimes using examples from different cultural contexts. The strongest responses went on to explain how these differing views could affect diagnosis, for example from the doctor's perspective, or by considering how cultural expectations may influence what an individual reports during a clinical assessment. A small number of responses focused effectively on how anxiety disorders are perceived or accepted within a particular culture. Weaker or non-creditworthy responses tended to state simply that cultures view mental health differently, without further explanation or clear reference to diagnosis of an anxiety disorder.

Question 3

(a) Responses to this question covered the full range of marks available, with many achieving 3 or 4 marks. Strong responses clearly outlined the key stages of systematic desensitisation, including the teaching of relaxation techniques, the construction of a fear hierarchy, and gradual exposure to feared stimuli while practising relaxation. Many of these responses were well contextualised to flying insects, often using appropriate examples of low- and high-fear situations. Some responses made effective use of psychological terminology, such as reciprocal inhibition, and explained that anxiety must be reduced or absent before progressing to the next level of the hierarchy. Weaker responses tended to give a very generic outline of therapy, sometimes gaining only 1 mark for identifying a fear hierarchy or mentioning relaxation without further detail. A small number of responses incorrectly suggested that the phobic stimulus would be paired with some positive thoughts or a reward, which is not creditworthy for systematic desensitisation.

(b) Responses to this question covered the full range of marks. Many responses were able to gain 1 mark by identifying a general weakness of therapy, such as the treatment being stressful or requiring commitment over time. Stronger responses explained why this was a weakness specifically of systematic desensitisation, often by referring to the need for repeated exposure to the feared stimulus and the possibility that this may lead some individuals to discontinue treatment before it is completed. Some responses supported this with examples, including reference to Habiba's fear of flying insects. Other common weaknesses included practical issues such as time, cost, and access to a therapist; however, these often achieved only 1 mark where the explanation was brief or not clearly linked to systematic desensitisation. A large number of responses incorrectly stated that the treatment was unethical or that it would cause psychological harm or make the phobia worse; these were not creditworthy, as systematic desensitisation is a consent-based therapy and distress is managed through relaxation techniques.

© 2025

Question 4

(a) Responses to this question varied and covered the full range of marks available. Strong responses gave clear and accurate descriptions of both MAOI anti-depressants and Ellis's rational emotive behaviour therapy (REBT). It was not necessary for responses to include all of the content listed in the mark scheme; rather, higher-level responses selected relevant information and explained it accurately. For MAOIs, good responses typically explained that monoamine oxidase inhibitors prevent the breakdown of neurotransmitters such as serotonin, noradrenaline and dopamine, resulting in higher levels in the brain and an improvement in depressive symptoms. Some responses also included appropriate reference to effectiveness or side effects. For REBT, stronger responses clearly outlined the ABC (and sometimes ABCDE) model, explaining the role of irrational beliefs in depression and how disputing these beliefs can lead to more rational thinking and improved mood.

Weaker responses often lacked sufficient detail for one or both parts of the question. For MAOIs, some responses stated that they inhibit monoamine oxidase but did not accurately describe the effect on neurotransmitters or confused MAOIs with SSRIs. For REBT, weaker responses often described only the ABC model without explaining how disputing beliefs helps to treat depression or gave a generic outline of CBT or talk therapy, which was not creditworthy. Some responses confused REBT with Beck's cognitive therapy, limiting the level awarded.

(b) Marks for this part of the question covered the full range available, although the most common levels awarded were Level 1 and Level 2. A notable proportion of responses (approximately 13 per cent) did not attempt this question. Stronger responses (Level 3 and above) were clearly structured issue by issue and included evaluation of both MAOI anti-depressants and Ellis's REBT, often covering three or four issues and debates. In addition to the named issue of reductionism versus holism, common issues discussed included determinism versus free will, application to everyday life, and individual versus situational explanations. Better responses contextualised these issues effectively, for example by explaining how MAOIs may be seen as biologically deterministic but involve free will in choosing to take medication, or how REBT requires active engagement and choice by the patient. Some responses were also able to explain how both treatments could be viewed as reductionist while recognising attempts to address more than one aspect of depression.

Weaker responses (Level 1 or Level 2) often failed to apply issues and debates to the treatments, instead giving brief or generic definitions without clear reference to MAOIs or REBT. Many responses simply stated which side of a debate each treatment supported, such as claiming MAOIs are reductionist or REBT is holistic, without further explanation or contextualisation, which was not creditworthy. Some responses focused mainly on strengths and weaknesses of the treatments without linking these to the issues and debates, which limited the level awarded.

Consumer Psychology

Question 5

Responses to this question covered the full range of marks, although many were awarded 1–2 marks. Most responses were able to name or outline at least one appropriate decision-making strategy, most commonly satisficing or referring generally to non-compensatory strategies. Stronger responses achieved full marks by not only identifying a suitable strategy but also explaining why it was appropriate to the scenario, specifically linking their answer to the fact that the website design meant comparisons between chairs were not possible. These responses often correctly explained how Abdul might select a chair based on meeting basic requirements or prioritising a single key feature.

Weaker responses frequently failed to apply the strategy to the context of internet shopping where comparisons between chairs were not possible, limiting the marks awarded. Some responses described compensatory strategies, such as weighing up multiple features, despite the scenario indicating that comparisons could not be made, which was not creditworthy. Other non-creditworthy responses included references to the take-the-best heuristic, high or low self-monitors, anecdotal descriptions of shopping behaviour, or summaries of the scenario without identification of a decision-making strategy. Very few responses demonstrated a clear understanding that non-compensatory strategies are more likely to be used when website design restricts comparison.

© 2025

Question 6

(a) Responses to this question covered the full range of marks. Some responses achieved full marks by giving a clear definition of determinism and linking this directly to a choice heuristic, often explaining how behaviour is influenced by external factors rather than free choice. A small number of responses also referred to internal causes of behaviour leading to behaviour rather than a choice to behave in a particular way. Many responses achieved 1 mark by providing a basic definition of determinism but failing to link it to a choice heuristic, or by outlining what a choice heuristic is without explaining how it demonstrates determinism. Non-creditworthy responses often attempted to define determinism in terms of experiences determining behaviour without explaining the lack of cognitive control or did not make any relevant reference to choice heuristics being deterministic.

(b) Responses to this question were generally weak, with very few achieving any marks. A notable proportion of responses (approximately 10 per cent) did not attempt the question, and only a small number of candidates did the consumer option. A very small minority of responses achieved 1 mark by identifying a strength of determinism, such as the idea that behaviour is caused by external factors and briefly linking this to choice heuristics by implementing appropriate sales strategies. No responses achieved full marks, as responses rarely made a clear link between a strength of deterministic explanations and a specific choice heuristic. Many responses focused on stating that it helps with understanding consumer behaviour. This type of response did not identify or explain a strength of determinism. This highlights a need for greater awareness of the advantages of deterministic explanations, which are widely used in psychology.

Question 7

(a) (i) Responses to this question generally accessed the full range of marks. Many responses correctly identified Zelda as an explorer (or native) and Louis as a tourist (or raider), which gained full marks. The most common correct pairing was explorer for Zelda and tourist for Louis. However, a large number of responses demonstrated misunderstanding of the command word *identify*. Some responses provided extended descriptions or rewrote the scenario in different words rather than naming an appropriate shopper spatial behaviour pattern, which was not creditworthy. Common incorrect responses included identifying either shopper as a specialist or identifying Louis as a native. Some responses justified the behaviour in terms of shopping needs (for example, stating that Zelda was doing her main shop), rather than identifying a recognised spatial behaviour pattern from the syllabus.

(ii) Responses showed a wide range of quality. Many responses gained the first mark by correctly identifying a different spatial behaviour pattern for Frank, with tourist being the most commonly selected. A small number of responses achieved full marks by outlining one or more key features of a recognised shopper type and linking these clearly to Frank's behaviour in the supermarket. Some responses also gained credit by accurately describing a type of trip, such as a short or central trip.

Weaker responses tended to be very brief or vague, for example stating that Frank would 'zig zag' around the store, which did not clearly indicate a recognised spatial behaviour pattern. Common non-creditworthy responses included inventing behaviour patterns not in the syllabus, such as 'on a mission' or 'routine-based' or simply rewriting the scenario for Frank using different words. This was particularly common where a similar approach had been taken in **part (a)(i)**.

(b) Responses to this question covered a range of marks. Stronger responses were able to explain a clear problem faced by psychologists when investigating shoppers' spatial movement patterns and link this explicitly to shopper behaviour. Common creditworthy responses included explaining that shoppers may not behave consistently across different stores or shopping goals, making classification difficult, or that shoppers may change their behaviour if they are aware they are being observed. Some responses also appropriately referred to ethical issues, such as lack of consent when observing shoppers.

Weaker responses often identified a problem without any explanation or without clear reference to shoppers' spatial movement patterns, which limited marks to 1 or resulted in no credit. A large proportion of responses did not address the question in terms of a problem faced by the investigator. Non-creditworthy responses included incorrect issues such as ecological validity or cultural differences, or simply naming a problem (for example, 'demand characteristics' or 'unethical') with no further elaboration.

© 2025

Question 8

(a) Responses to this question covered the full range of marks available, although a notable proportion of responses (approximately 17 per cent) did not attempt the question. Strong responses demonstrated good knowledge of both the Yale model of communication and a study investigating types of advertising media. For the Yale model, Level 3 responses typically identified several of the five features and provided a brief outline of each, clearly linked to advertising. Weaker descriptions (Level 1 or Level 2) tended to identify fewer features, often focusing only on the channel or the audience, or simply listing features without explanation.

For the study on types of advertising media, the majority of responses that gained credit referred to Ciceri et al. Stronger responses outlined key features of the procedure, including the type of advertising media used and the task or measurement involved, and included at least one relevant result. Some responses included more detail than was required, such as sample characteristics or multiple results, although this was not necessary to achieve higher levels. Weaker responses often included inaccurate or incomplete procedural detail, omitted results, or failed to describe how advertising media were investigated. Non-creditworthy responses included outlining different types of advertising without reference to a study or describing studies that investigated only one type of advertising medium, such as Auty and Lewis or Fischer, which did not meet the requirements of the question.

(b) Marks for this part of the question covered the full range available, although the most common levels awarded were Level 1 and Level 2. A notable proportion of responses (approximately 27 per cent) did not attempt this question, which is likely to be linked to the weaker performance on **part (a)**. Stronger responses (Level 3 and above) were clearly structured issue by issue and included evaluation of both the Yale model of communication and a study on types of advertising media, most commonly Ciceri et al. In addition to the named issue of objective and subjective data, higher-level responses often discussed generalisability, ecological validity and application to everyday life. Some responses were able to explain effectively how Ciceri et al. included objective measures such as eye-tracking, EEG and recognition data, while also considering whether these fully captured subjective experiences such as frustration. Others applied the named issue to the Yale model by discussing the subjective nature of factors such as audience characteristics or message effectiveness.

Weaker responses (Level 1 or Level 2) often failed to apply issues and debates to the content in **part (a)**, beyond simply identifying the Yale model of communication or Ciceri et al. Some responses gave brief or generic definitions of issues such as objectivity or validity without contextualisation. Other non-creditworthy responses simply stated which side of a debate each approach supported, for example claiming that Ciceri et al. is reductionist or that the Yale model is holistic, without explanation or reference to advertising or the study, which did not receive credit.

Health Psychology

Question 9

Responses to this question covered the full range of marks available. Most responses were able to suggest two reasons for Chloe's unrealistic optimism, although it was common for responses to gain only 1 mark per reason where suggestions were outlined briefly. Popular reasons included Chloe's lack of personal or family experience of serious illness, and perceptions of personal control over health outcomes. Stronger responses achieved 2 marks per reason by extending these ideas with appropriate context or examples, such as explaining specific health behaviours Chloe believes will reduce her risk of illness or increase her life expectancy. Some responses also referred to stereotypes about who is most likely to become ill or to beliefs encouraged by family experiences.

Weaker or non-creditworthy responses often rewrote the scenario in different words, for example restating that Chloe believes other students are more likely to become ill than she is. Other responses explained why Chloe's beliefs are unrealistic, rather than giving reasons for her unrealistic optimism, which did not address the question.

© 2025

Question 10

(a) Responses to this question covered a range of marks, with many achieving full marks. Well-prepared responses clearly outlined both sides of the debate, defining individual explanations in terms of internal or dispositional factors (such as personality or innate characteristics) and situational explanations in terms of external or environmental influences. These responses gained both marks by addressing each explanation explicitly. Candidates were generally more successful in defining situational explanations than individual explanations.

Weaker responses often referred to the wording of the question without further explanation, for example stating that an individual explanation refers to the individual, which was not creditworthy. Some responses outlined only one side of the debate, limiting the marks awarded.

(b) Responses to this question were generally weak, with most achieving 0 or 1 mark. A small number achieved full marks by outlining a clear strategy for improving health, such as fear arousal or providing information, and then explaining why this strategy has a situational explanation. These stronger responses focused on the role of external factors, explaining how the strategy creates an environment that encourages healthy behaviour and does not depend on individual personality or disposition.

Many responses achieved only 1 mark by outlining a strategy for improving health but failing to explain why it is situational. A significant number of responses did not outline any relevant health strategy from the syllabus and therefore could not explain why it had a situational explanation. Very few responses referred to strategies for managing stress, despite these being included in the syllabus. Responses that identified a health strategy on its own, without explanation, did not receive credit.

Question 11

(a) Responses to this question covered a range of marks, with many achieving full marks. Strong responses were able to suggest two different measures of pain and provide clear detail about how each measure works. The most common measures described were the clinical interview, visual analogue scales, and the UAB pain behaviour scale. Good responses often named the measure explicitly, such as the McGill Pain Questionnaire, and included one or more relevant features explaining how pain is assessed.

Responses that achieved 1 mark typically did so by naming or briefly outlining a pain measure without providing sufficient detail about how it operates, or by describing the procedure without clearly identifying the measure used. Weaker or non-creditworthy responses included vague references to self-report or observation without linking these to a recognised pain measure from the syllabus or included inaccurate details about how questionnaires or observational scales are scored. It should be noted that biological measures such as blood or urine tests, or scans such as fMRI, do not measure pain and are not creditworthy.

(b) Responses to this question covered the full range of marks available. Stronger responses achieved full marks by clearly identifying a weakness of a pain measure and linking this explicitly to the specific measure suggested in **part (a)**. Common creditworthy weaknesses included lack of depth, differences in how individuals interpret scales or descriptors, the subjective nature of pain, and difficulty accurately observing pain behaviour when pain is not outwardly expressed.

Responses that achieved only 1 mark were often brief and failed to refer clearly to the pain measure suggested in **part (a)**, limiting development of the weakness. Non-creditworthy responses commonly identified only the type of data (for example stating that the measure used quantitative data) without explaining why this was a weakness. Other non-creditworthy responses included discussing weaknesses of pain treatments rather than pain measures or giving a weakness of a biological measure where such a measure had been incorrectly suggested in **part (a)**.

© 2025

Question 12

(a) Responses to this question covered the full range of marks available, although overall performance was weak and a notable proportion of responses (approximately 12 per cent) did not attempt the question. Where responses did gain marks, these were most commonly in Level 1 or Level 2. The vast majority of responses were able to define false positive and false negative diagnoses accurately, and many gave appropriate health-related examples of each. Some stronger responses also referred to difficulties with patient disclosure, including reference to research such as Robinson and West, explaining that patients may disclose more information in non-face-to-face formats, particularly when information is sensitive or embarrassing.

Stronger responses were relatively rare but typically included relevant research evidence linked directly to making or presenting a diagnosis. A small number of responses outlined Savage and Armstrong, which was credited, as this study relates to practitioner behaviour following diagnosis and sits within the same syllabus topic area. However, many responses lacked focus on diagnosis and instead included anecdotal descriptions of what a diagnosis is, reasons why doctors may make mistakes, or suggestions for improving communication. These responses often contained lengthy writing but did not address the specific syllabus content required and therefore received no credit.

(b) Marks for this part of the question covered the full range available, although overall performance was very weak and a substantial proportion of responses (approximately 22 per cent) did not attempt the question. A very small number of responses achieved Level 3 by structuring their evaluation issue by issue and applying it directly to the content in **part (a)**. These responses addressed the named issue of validity, most commonly discussing ecological validity and generalisability, for example by considering the extent to which findings based on specific patient groups (such as those diagnosed with skin cancer) could be applied more broadly.

Many responses achieved Level 1 or Level 2. These responses often demonstrated limited understanding of the requirements of the question and failed to evaluate the material in **part (a)** beyond general statements.

Common issues and debates selected included application to everyday life and generalisations from findings, but these were frequently described without sufficient contextualisation. The named issue of validity was often poorly addressed, with responses either defining validity in general terms or failing to apply it to research into making or presenting a diagnosis. As with **part (a)**, responses that had selected incorrect research were unable to access credit here, which further limited performance.

Organisational Psychology

Question 13

Responses to this question covered the full range of marks available, with a number of strong responses achieving full marks. Stronger responses demonstrated a clear understanding of the stages of group development and were able to apply these effectively to the context of the school council. These responses typically outlined three or more stages and linked most stages to relevant behaviours such as establishing rules, conflict between members, developing shared norms, and working cooperatively to make decisions about issues such as the school menu or recycling facilities.

Weaker responses often showed partial knowledge of group development but struggled to accurately describe the stages or match the correct stage name to the behaviours involved. Some responses listed stage names with little or no explanation, while others gave generic descriptions of group work without reference to specific stages or the school council context. These responses were typically awarded 1 or 2 marks. A small number of responses confused the stages entirely or gave anecdotal accounts of group decision-making that did not clearly reflect the group development model and were therefore not creditworthy.

© 2025

Question 14

(a) Most responses were able to achieve at least 1 mark by giving an accurate definition of the nomothetic approach, typically referring to the aim of identifying general laws of behaviour. Some responses were able to gain the second mark by applying this approach to a need theory of motivation. Stronger responses clearly linked the nomothetic approach to a named theory, most commonly Maslow's hierarchy of needs. Weaker responses often stopped after defining the nomothetic approach and did not include an example from a need theory, limiting them to 1 mark. Some responses focused on the quantitative nature of nomothetic research which made it difficult to link to a need theory. A small number of responses outlined Maslow's or McClelland's theory accurately but failed to link this to the idea of general laws of behaviour, which again restricted the mark awarded.

(b) Performance on this question was generally weak, with relatively few strong responses seen. A small number of responses achieved full marks by clearly outlining how an idiographic approach could be used to investigate motivation at work and linking this explicitly to a need theory. These responses often described methods such as interviews and explained how these would explore an individual employee's experiences of needs. Where this link was made, full marks were awarded. Some responses achieved only 1 mark by outlining an idiographic method, most commonly interviews, without linking this to a need theory of motivation. Other responses attempted to describe an idiographic approach but did not apply it to motivation at work or instead referred to general job satisfaction or workplace behaviour. In line with the mark scheme, questionnaires were only creditworthy where candidates made it clear that open-ended questions would be used. Responses that described need theories without reference to an idiographic method, or that confused idiographic and nomothetic approaches, were not creditworthy.

Question 15

(a) Responses to this question covered a range of marks, with some good answers achieving full marks. Stronger responses clearly identified two different extrinsic motivators and applied them effectively to the context of the carpet factory. The most common extrinsic motivators used were bonus payments and performance-related pay, with responses explaining how linking financial rewards to output could increase motivation and productivity. A small number of responses also referred to pay increases, which were creditworthy when clearly distinguished from other reward systems. Weaker responses often failed to apply their suggestions to the factory context, limiting the marks awarded. Some responses named an extrinsic motivator without explaining how it would operate in practice, resulting in 1 mark. A common error was to describe the same extrinsic motivator twice using slightly different wording, which restricted the total marks available. Other responses confused extrinsic and intrinsic motivation or gave general statements about motivating workers without reference to a reward system, which were not creditworthy.

(b) Performance on this question was mixed. Some responses were able to achieve full marks by identifying a clear strength of an extrinsic motivator other than increased production and explaining why this benefit would occur. Stronger responses referred to increased job satisfaction, a greater sense of fairness, or increased commitment to the factory, and explained how financial rewards could make workers feel valued or invested in the organisation. These responses demonstrated clear understanding and achieved 2 marks. Many responses achieved only 1 mark because the strength was identified but not sufficiently explained. These responses often stated that workers would be more satisfied or motivated without explaining why the suggested reward system would lead to this outcome. A common error was to focus on increased productivity, which was explicitly excluded by the question and therefore not creditworthy. Some responses also made unsupported claims, such as suggesting that bonus payments would automatically increase loyalty to the factory, without explanation. Overall, relatively few responses developed their points fully enough to access both marks.

© 2025

Question 16

(a) Responses to this question covered a range of marks, with some good descriptions achieving Level 3. Stronger responses demonstrated clear understanding of the difference between rapid rotation and slow rotation shiftwork, often referring to the frequency of shift changes and the impact on circadian rhythms. These responses accurately explained why rapid rotation is more disruptive and why slow rotation allows greater physiological adjustment. In addition, better responses included a clear outline of a relevant study, most commonly Gold et al., with brief procedural details and at least one appropriate finding relating to sleep disruption, accidents, or health outcomes.

Weaker responses often gave a very general outline of shiftwork without clearly distinguishing between rapid and slow rotation, or they defined one type without comparison to the other. Some responses failed to identify the types by name, which limited the level awarded. A number of responses either did not attempt the study element or gave a very brief or vague description without sufficient detail of the procedure or results, restricting them to Level 1 or Level 2. In some cases, responses appeared incomplete, suggesting that candidates may have run out of time.

(b) The marks for this question tended to be between level 1 and level 3. A small number of candidates produced strong evaluative responses to this question. These responses addressed the named issue of quantitative and qualitative data effectively and linked their discussion clearly to rapid and slow rotation shiftwork and to research such as Gold et al. The majority of research cited in this area, including Gold et al., relies heavily on quantitative data, for example questionnaire responses and accident rates. Strong responses recognised that quantitative data allows for easier comparison, the calculation of percentages and means, and the identification of patterns across large samples.

Other commonly used evaluative issues included application to everyday life and the use of questionnaires. In addition, weaker (Level 1 or Level 2) responses lacked evaluation points that provided any depth of explanation. Each point tended to be very brief, and reference was frequently made to the content in **part (a)** but this was merely identifying which side of the debate the theory was on or whether it did have applications or quantitative data but only occasionally was this explained. Sometimes analysis might be given but this was also just identified and not explained. Approximately 10 per cent of candidates did not attempt this question.

© 2025

PSYCHOLOGY

Paper 9990/41

Specialist Options: Application and Research Methods

Key messages

- What has been learned from the AS component of the syllabus should be transferred to the A2 component. For example, at AS, candidates learn about methodology, such as experiments, which also apply to A2.
- Questions should be read carefully, ensuring that the focus is on what the question asks.
- For ***Section A*** answers, candidates should relate their answer to the study in the question or include an example. Questions frequently end with 'in this study' and so the answer should be related to that specific topic area/study.
- All terminology should be explained. Writing 'it is valid and reliable' for example, is insufficient without explanation, application or example.
- The syllabus includes 'example studies' such as 'e.g., Oldham and Brass (1979)'. Example studies can be substituted for alternatives, but these alternatives must cover the same or very similar content to the example study. If the Oldham and Brass study is substituted, the alternative study must be about a move to open plan offices and the data that was gathered from that move. The alternative cannot be about something different.

General comments

Some candidates answered questions from one option only. Other candidates who correctly answered two options, sometimes performed considerably better in one option than the other.

A significant number of candidates answered two questions from ***Section B*** instead of one (only one of these ***Section B*** responses can receive credit). Candidates are advised to read the instructions on the front cover of the question paper and to read the heading instructions for each question section.

Candidates should double check that the terminology they use in their answers is correct. Often terms such as reliability and validity were used interchangeably, as were qualitative and quantitative, and independent and dependent variables. There was also confusion with the terms format and technique in relation to questionnaires and interviews.

Interview format is often misunderstood. A structured interview can consist of either closed questions, open questions or a combination of both. Structured means that the same questions are asked to all participants in the same order. An unstructured interview can also consist of closed questions, open questions or a combination of both.

Questionnaires are not objective. A person can be honest, but they can also be dishonest in their answers; the answers given are therefore subjective. Objective data is fact, such as the result of a blood test or the number of people attending a clinic.

Candidates should read all parts of a question and follow its instruction. If a question states: 'other than 'x', then 'x' should not be written about because 'x' is part of the question and will attract 0 marks.

Section A

There is no need for candidates to re-write the question before beginning their answer.

Some candidates apply the term 'demand characteristics' and 'social desirability' to every answer, when in most cases neither term applies.

© 2025

Question **part (c)** requires a general evaluative point that could relate to any study (such as a strength or weakness of a method) but it also requires the general point to be related to the specific sub-topic/study in the question. Answers often included strengths and weaknesses, but these were often not related to the question, and so restricted marks.

In question **part (c),** when answering the 'in this study' response, simply rewriting the words in the question is always insufficient and it will be awarded 0 marks. Doing this merely shows that candidates can copy a question. Something more needs to be added to show understanding of the study in question. See examples below, such as **Question 5(a)**.

Candidates should not use psychological terms without explanation. Frequently answers were limited to 'it is reductionist' or 'it is useful in everyday life' without further explanation. Stating 'it is reductionist' is *identification*, it is not *evaluation*. Explaining why something is or is not reductionist would be a strength or weakness.

Candidates should not use the terms' reliability and validity to answer every **part (c)** question for three reasons: they do not apply to most questions and so cannot be awarded marks, candidates using the terms often do not know how they apply to the specific question and candidates often confuse the terms.

Many candidates state that a weakness of a case study or longitudinal study is that the researcher has some kind of relationship with the participant. This is very, very rare and will never be an 'automatic' weakness of these methods. Any 'relationship' is often instant dismissal for a psychologist as well as being highly unethical.

Section B

Candidates should only answer one question from this section.

Many candidates appeared to assume that they must conduct an experiment whatever the question. An interview, questionnaire or observation are methods independent of an experiment and candidates should not try to make other methods 'fit' into an experimental format.

Some candidates evaluate their plan in **part (a)** by listing strengths and weaknesses. This should not be done because: the question does not ask for evaluation; there are no AO3 marks allocated to evaluation and evaluation is done in **Questions (c)(i), (c)(ii)** and **(c)(iii).**

Some candidates included a paragraph of results. This achieves no marks because the question asks for a plan only. Further, the proposed plan has not been carried out, so no actual results are gathered.

Every plan, whatever the question, should include a consideration of ethical guidelines, an essential element.

Candidates need to know the distinction between questionnaire format and technique, and interview format and technique. As stated on the syllabus: questionnaire technique: paper and pencil (i.e., done by a person with the researcher present), online or postal. Questionnaire format: open and/or closed questions. Interview technique: telephone or face-to-face. Interview format: structured, semi-structured, unstructured.

When using psychometric tests candidates should not use acronyms unless the full title of it is provided first. For example, 'Beck Depression Inventory (BDI)' is fine, with BDI used afterwards. Further, it is insufficient to simply state 'I would use a questionnaire similar to K-SAS' (such as when writing about pyromania, for example).

Answers to **part (a)** questions in this section should include an appropriate plan, applied a range (four or five) of specific methodological features, each of which should be explained fully, to show good understanding. Candidates should also include appropriate 'general' methodological features such as sample, sampling technique and location of the study. Many answers listed features such as 'I would have a random sample' and 'It would be an independent measures design' without explanation of why it would be a random sample, or how it would be obtained. Elaboration of these general sentences should be included.

In **part (b)(i)**, candidates should describe some relevant psychological knowledge that the whole question is based on. If the question, for example, asks about ways in which pain can be measured, then candidates should describe relevant measures.

© 2025

In **part (b)(ii),** candidates should explain what aspects of this psychological knowledge their **part (a)** plan is based on. These two question parts must be linked.

In **part (c)**, candidates must refer to what they did in their specific plan rather than give a generic answer that could apply to any study. Use of an example or quoting from their plan would be ideal.

Section B can be considered as follows: A teacher teaches a sub-topic from the syllabus and gives the candidate some psychological knowledge. The teacher then tells each candidate to plan a study using method 'x' to investigate some part of that sub-topic. The candidate plans the study using the psychological knowledge of the sub-topic and they use their methodological knowledge about method 'x'. In the examination, **part (a)** is the plan; **part (b)(i)** is the sub-topic knowledge and **part (b)(ii)** is how the knowledge was used to construct the plan. Exam question parts **(c)(i), (ii)** and **(iii)** then ask about some methodological decisions and evaluation about the plan.

Comments specific to questions

Section A

Question 1

(a) **(i)** Candidates either knew the two drugs used in the trials which were Nalmefene and Naltrexone and were awarded 2 marks, or they did not. Some candidates were awarded marks for misspellings that were close enough to the original. Candidates writing 'opioid antagonists' were awarded 0 marks.

(ii) Many candidates achieved full marks by stating '16 weeks' and '18 weeks'. It was not necessary to state which time period belonged to which trial or drug.

(b) Candidates frequently were awarded full marks for excellent answers showing good understanding of why placebos are used. For example, 'without a placebo group, participants would know that they are receiving an active drug and this alone might result in a change in behaviour' and when nalmefene or naltrexone are identified as the active drug, the answer was worth full marks. Some candidates were awarded 1 mark for 'it acts as a baseline'.

(c) Candidates were often confused about what location meant and frequently did not know that this key study was conducted in different locations. There were incorrect suggestions that it would cost more and participants would have to travel, without appreciating that different participants were used in each location. Also, there were comments about cultural differences. However, the participants were all gamblers, and the drug works physiologically on every human in the same way.

Question 2

(a) The syllabus requires candidates to know 'a case study of schizophrenia' and most chose to outline the first case study described in the study by Aneja (2008). Any two details achieved marks, such as aggressive father, living with mother and grandparents, showing aggressive outbursts and early onset schizophrenia, amongst others. Some candidates provided detail too brief for marks to be awarded such as 'a child with schizophrenia'. The Aneja study is an 'e.g.' study, so any appropriate alternative could have been substituted. Such instances could only receive credit if there was sufficient detail for an examiner to confirm the legitimacy of the answer. Providing a name and date would be ideal.

(b) Many candidates stated, 'the person might not tell the truth', which is correct, but this can apply to anyone and any study, not specifically someone with schizophrenia as the question requested. The question concerned reliability which is about consistency and a person who has schizophrenia, perhaps because of an impaired sense of reality, might not be consistent if data is gathered from them on two (or more) occasions.

© 2025

(c) Generic strengths were awarded marks, one for each strength such as 'studied in depth/detail'; 'can apply a range of different methods', etc. Candidates struggled to provide examples to support each strength, despite describing a case study in question part (a) where there were often many examples that could have been used.

Question 3

(a) (i) Some candidates wrote very impressive answers (awarded full marks) showing excellent knowledge of the procedure of the pre-test, including details such as the three pairs of product variants, the 12-item potency construct, and the 7-point rating scale. Others could only write that 'it is about yogurt packaging' (for 1 mark) and one candidate wrote 'what pre-test' scoring 0 marks.

(b) There were some excellent answers provided by candidates who suggested that there could be a lack of validity. What made such answers excellent was that they explained *why* it might lack validity and gave an example, such as testing something that would be unrealistic, like a square yogurt pot. Candidates writing nothing more than 'one effect is lack of reliability and validity' were awarded 0 marks.

(c) This was similar to other **part (c)** questions where candidates often gave a strength and/or weakness without relating it to the question. The answer could have been: 'one strength is that the location is congruent with the study' (1 mark) 'because yogurt in packaging is sold in a supermarket' (+1 mark). Here there is a generic strength and an applied example of yogurt and supermarket.

Question 4

(a) Nearly all candidates were awarded 1 mark for their answers describing retail atmospherics as the controllable characteristics of retail space which entice customers to enter the store, shop and purchase a product. Most candidates were awarded 2 marks, usually for providing an example, such as landscaping, window displays, background music or pleasant odours.

(b) Candidates frequently used 'classical music' and referred to the study by North *et al*. where the music increased spending, or to the study by Chebat and Michon who investigated the effect of odour on shopper ambience. The best answers related a specific retail atmospheric to the PAD model (pleasure/arousal/dominance).

(c) Although the question asked about strengths for the retailer (seller), some candidates still wrote about the purchaser (customer). For the retailer, strengths might be that they can apply psychological knowledge: if they have an 'up-market' establishment playing classical music it could lead to pleasure in customers and an increase in sales. Similarly, a retailer could use the findings by Chebat and Michon and create arousal in customers by spraying a pleasant odour in their store.

Question 5

(a) **(i)** Outlining the procedure was straightforward for most candidates who were awarded full marks for including the physiological (relaxation and diaphragmatic breathing) and psychological components (imagining a peaceful scene of their own choice). Some candidates only included one of these components so could only be awarded 1 partial answer mark.

(ii) The experimental design used was independent groups, and correct identification was awarded 1 mark. Continuing to identify what the groups were (control, relaxation only, and relaxation and imagery) resulted in the second mark being awarded.

(b) Many candidates suggested using a repeated measures design and the best answers went on to explain that this would not work because of a knock-on effect, where if relaxation and imagery were taught first it would make the control or relaxation only conditions redundant. Some candidates suggested that it would cause fatigue or perhaps lead to attrition.

© 2025

(c) Candidates sometimes struggled to suggest a strength of studying a sample of female participants, whilst others made good points such as 'stress reduction techniques (like relaxation and imagery) can be tailored specifically for females rather than a generic 'one size fits all'' or that 'techniques target specific female illnesses'. For the weakness, nearly all candidates suggested that the techniques may not apply to men.

Question 6

(a) This question was answered incorrectly by many candidates who outlined the *aim* of the study or the *findings*. The question asked for an outline of the *methodology* that was used. Two methods were used: a questionnaire-based survey and interviews (one of which was conducted in the home and a telephone interview to check on medication).

(b) Candidates could suggest an ethical guideline but often not relate it to this study. Some candidates identified a guideline, such as informed consent, but merely stating 'informed consent' was insufficient. There needed to be a comment about what informed consent is, or who is giving the consent, for 1 mark. Relating this guideline to the study (for 2 marks) could have been done by stating that parents had to give informed consent because the children were aged from 1.5 years to 6 years for example, which shows knowledge of the study.

(c) Answers to this question followed the typical pattern for **part (c)** answers, with generic weaknesses being provided such as 'a parent may give a socially desirable answer' with no elaboration or attempt to relate this to medication adherence in children. Candidates should appreciate that all question parts, i.e. **(a)**, **(b)** and **(c)** are part of the same question about a study or sub-topic and that for **part (c)** answers what has been written in **part (a)** can often be applied to strengths and/or weaknesses in **part (c)**.

Question 7

(a) **(i)** Some candidates could not provide any correct feature, often guessing, whilst others had little problem stating two features from: 136 participants, all psychology students, all from National Open University of Spain and that 53 per cent were women.

(ii) After reading the narrative, participants were asked to evaluate a supervisor (awarded 1 mark) and who was occupying that position for a trial period (1 additional mark). Marks were also awarded for stating 'asked to collaborate in the evaluation process by completing an anonymous questionnaire'. There were some perfect answers, but quite a few candidates could not provide a correct answer despite this being a key study.

(b) Some answers were awarded 1 partial mark for stating 'so all the conditions of the IV could be tested', whilst those who went on to state what the IV's were, and so relate the generic comment to this study, were awarded the full 2 marks. The IV conditions were: male stereotypical versus female stereotypical leadership style, and male leader versus female leader.

(c) Like many **part (c)** answers, candidates could provide two generic strengths (in this instance of using experiments) but could not relate these strengths to the study itself. For example, a candidate might write about the IV being manipulated (but not mention that the IV was male/female leader and male/female leadership style). Additionally, they might write about the standardisation of the procedure but not mention the evaluation of the leader always being the same, or the use of 14 adjectives, for example. Generic answers will never be awarded full marks; answers that are related/applied could be.

© 2025

Question 8

(a) Very few candidates could provide two questions from Walton's QWL. The syllabus is '*Walton's quality of working life (QWL) including eight conditions and QWL evaluation scale*', and so to know two questions from this scale should not be a problem for any candidate. The question asked for any two from the first four appearing on the QWL. Candidates are advised that the whole syllabus will be examined and so the whole syllabus should be revised thoroughly.

(b) Questions about reliability are the same whatever option it is related to. In this instance, there needed to be a comment about the same person taking the same test (i.e. the QWL) on a different occasion to test consistency. In other words, test-retest reliability. Candidates suggesting inter-rater reliability or validity were awarded 0 marks.

(c) In response to this question, candidates often gave very brief, single sentence answers that were very general, meaning that they could apply to any question from any option. Candidates should always apply the strength or weakness (or two weaknesses as in this instance) to the study or sub-topic the whole question (**parts (a)**, **(b)** and **(c)**) is about. Questions asking about weaknesses of questionnaires require answers about the *questionnaire* (such as a restricted scale or poor question wording) and not about a *participant* (who may lie or give a socially desirable answer).

Section B

Question 9

(a) There were several errors which restricted marks.

(i) Many candidates did not know what the term 'imaginal desensitisation' involved and in their answers never explained the treatment process, writing nothing more than the term itself, or they confused it with covert sensitisation (a very different therapy).

(ii) Some candidates decided to conduct an *experiment* comparing a desensitisation group with a non-desensitisation group with little or no observation. Similarly, some candidates decided to conduct an *interview*, the only reason for this appearing to be to gather qualitative data. If the question asks for an observation, then the plan should be about an observation *in detail*. Any additional method will gain no additional marks.

(b) (i) Following from **part (a),** many candidates could not answer this question part at all, showing no knowledge of imaginal desensitisation. Other candidates wrote incorrectly about covert sensitisation. Imaginal desensitisation involves teaching progressive muscle relaxation and then the person visualises themselves being exposed to the situation. Some candidates wrote legitimately about types of gambling behaviour.

(ii) If there was no answer in **(b)(i)** then no marks could be awarded in this question part as it links **(b)(i)** with **(a)**. Candidates explaining the link related details of their structured observation in **part (a)**, such as their response categories. Others had categories for behaviour before treatment and after treatment.

(c) **(i)** Candidates often struggled to decide whether a gambling casino (where a participant could be observed after being treated) was a natural environment or whether it was a controlled environment. Both were awarded credit. Some candidates chose to set up an artificial laboratory scenario which was correctly labelled as a controlled observation.

(ii) Most candidates gave an appropriate reason for their choice of structured observation using response categories but often there were no examples of what the categories would be. Others used a simple tally chart with a yes/no for a category of 'gambles (stops to play) when passing a fruit/slot machine'.

© 2025

(iii) Quantitative data was the most logical type to gather from response categories with statistical analysis and comparison as a strength. However, many candidates stated that they applied both quantitative and qualitative data and struggled to give a reason why they had used both. Candidates are advised to gather only one type of data to make answering questions much easier, and in this instance quantitative data was apposite.

Question 10

(a) All candidates planned an experiment, as the question required, but many gathered data through questionnaire or interview. This can be done, but only after full details of the named method have been included. Many candidates appeared to use a questionnaire or interview simply to say, 'I gathered qualitative data', when this was irrelevant to the effect of music on *spending*. Candidates are advised not to include qualitative data in questions where it is not required. For this question, many candidates incorrectly had the IV as different *types of music* (i.e. they replicated the North et al. study) and were awarded level 1 or 2 marks. The question asked for classical music in different *types of restaurants*.

(b) **(i)** For psychological knowledge, the most appropriate research would be that by North et al. (2003). This research played different types of music in the same restaurant and found that classical music resulted in increased spending when compared to pop and no music.

(ii) This question part, like all other **part (b)(ii)** questions, required an explanation to show how what was described in **10(b)(i)** informed the plan in **part (a)**. For example, candidates who outlined that North et al. used a classical music cd of 76 minutes, also used a classical music cd of 76 minutes. Similarly, North et al. studied an 'up-market' restaurant so many candidates also included an 'up-market' restaurant as one of their types. In other words, candidates based their plan on things taken from the North et al. study.

(c) **(i)** Many candidates listed the controls they had applied; the same person gathering data, the type and length of music and keeping menu items the same. Often candidates did not explain the reason why they had applied these controls when this is what the question asked. Some candidates wrote about generic controls here for the first time even though they had not included any controls in their **part (a)** plan.

(ii) Some candidates wrote about one *strength* of the controls they had applied and were awarded no marks as this did not answer the question. Others struggled to find any weaknesses. Some candidates wrote about a weakness of the controls that they had not applied in their plan. The best answers focused on having too many controls reducing ecological validity, or that too many controls could make participants suspicious and perhaps cause demand characteristics.

(iii) Top marks were achieved by candidates who knew the findings of North *et al*. and the requirements of this question. This meant that some candidates opted for a directional hypothesis because they expected classical music to increase spending in 'up-market' restaurants and not in other restaurants they had included in their plan.

Question 11

(a) There were two problems which reduced marks for this question.

(i) The inclusion of irrelevant questionnaires and interviews, as explained above, see **Question 10(a)**.

(ii) Candidates did not know what biofeedback was. To investigate the effectiveness, candidates need participants to have biofeedback and so they should outline the procedure of biofeedback. It is insufficient to write nothing more than the word 'biofeedback' throughout the answer. Top answers had an IV of biofeedback and no biofeedback with a DV of decrease in level of tone on a scale of 1–10 (if using GSR). Designs were usually independent measures with a range of controls applied.

© 2025

(b) **(i)** For psychological knowledge, the study by Budzynski and Stoyva (1969) was most appropriate, although as this is an 'e.g.' study any suitable alternative could be substituted. The Budzynski and Stoyva study used biofeedback to reduce forehead muscle tension. Some candidates described this study in detail and were awarded 4 marks, others provided a brief outline, and others could not identify any psychological knowledge at all.

(ii) Candidates awarded full marks wrote about how they used their knowledge of biofeedback to inform their plan. For example, because Budzynski and Stoyva focused on forehead muscle tension, so did many plans. Similarly, as their participants received feedback through an auditory 'tone' this also featured in many plans.

(c) **(i)** Most candidates chose to apply an independent measures design where different participants did different conditions, usually of biofeedback or no biofeedback. Some candidates chose a repeated measures design even though they had feedback or no feedback conditions.

(ii) Many candidates stated simply that because they had conducted a laboratory experiment in a controlled environment, a weakness was that it lacked ecological validity. Answers ending here were awarded 1 mark, but those going on to relate it to the study could be awarded an additional mark when stating, for example, 'stress in university students is when they are sitting at a desk looking at past examination papers'.

(iii) The choice of dependent variable is crucial in determining the effectiveness (or not) of the biofeedback. Following on from AS knowledge, the DV should be fully operationalised in order for full marks to be awarded, such as the 'DV is the decrease in level of tone on a scale of 1–10 when measuring GSR (Galvanic Skin Response)'. Answers such as 'the DV is level of stress' was too vague for marks to be awarded.

Question 12

(a) Answers to this question showed the same flaws as answers to other options, namely not using the correct method, not explaining terms (such as job enlargement) and showing poor methodological knowledge. In reverse order, many candidates confused question format and questionnaire technique despite the syllabus specifying the differences. Technique relates to whether the questionnaire is given to participants face-to-face (i.e. 'paper and pencil'), by post or online. Often candidates wrote 'gave the questionnaire to participants'. Many answers referred to 'job enlargement' without asking anything specific and so showing no knowledge of the term at all. Others focused on job satisfaction and often only one question was asked, such as 'do you feel more satisfied after your job enlargement'. Finally, many candidates decided to plan an experiment with IV and DV and no marks could be awarded for using the wrong method.

(b) **(i)** Relevant psychological knowledge here could include two things: job enlargement which transforms the jobs to include more and/or different tasks and has two types: horizontal and vertical enlargement, or job satisfaction rating scales such as the JDI could have been used, as could Walton's quality of working life evaluation scale.

(ii) Relevant psychological knowledge mentioned above in **(b)(i)** was lacking in most answers which made answering this question part very difficult. Candidates are advised that if they have no knowledge of a particular sub-topic then they should consider answering the question from their second studied option.

(c) **(i)** Many candidates answered this question part incorrectly. The question stated questionnaire *technique* and NOT questionnaire *format*. Technique relates to whether the questionnaire is given to participants face-to-face (i.e. 'paper and pencil'), postal or online. Candidates who wrote about an open or closed question *format* scored 0 marks.

(ii) Candidates writing about question format in **12(c)(i)** also provided incorrect answers to this question part when giving a weakness of their question format. Candidates answering the question correctly often gave a good weakness (online could raise doubts about the veracity of the data;

© 2025

postal may not get many returns; face-to-face may create demand characteristics) but often failed to relate their answer to what they had written in their plan.

(iii) Question format is about whether questions are open and/or closed. Some candidates were confused because they had already incorrectly written about question format in parts **(c)i** and **(c)ii**. Candidates could choose open or closed or a combination of the two. Most candidates opted to use both types 'so I could gather quantitative data'. It is advised that one question type should be used to avoid confusion.

© 2025

PSYCHOLOGY

Paper 9990/42

Specialist Options: Application and Research Methods

Key messages

- What has been learned from the AS component of the syllabus should be transferred to the A2 component. For example, at AS candidates learn about methodology, such as experiments, which also apply to A2.
- Questions should be read carefully ensuring that the focus is on what the question asks.
- For **Section A** answers, candidates should relate their answer to the study in question or include an example. Questions frequently end with 'in this study' and so the answer should be related to that specific topic area/study.
- All terminology should be explained. Writing 'it is valid and reliable' for example, is insufficient without explanation, application or example.
- The syllabus includes 'example studies' such as 'e.g., Oldham and Brass (1979)'. Example studies can be substituted for alternatives, but these alternatives must cover the same or very similar content to the example study. If the Oldham and Brass study is substituted, the alternative study must be about a move to open plan offices and the data that was gathered from that move. The alternative cannot be about something different.

General comments

Some candidates answered questions from one option only. Other candidates, who correctly answered two options, sometimes performed considerably better in one option than the other.

A significant number of candidates answered two questions from ***Section B*** instead of one (only one of these ***Section B*** responses can receive credit). Candidates are advised to read the instructions on the front cover of the question paper and to read the heading instructions for each question section.

Candidates should double check that the terminology they use in their answers is correct. Often terms such as reliability and validity were used interchangeably, as were qualitative and quantitative, and independent and dependent variables. There was also confusion with the terms format and technique in relation to questionnaires and interviews.

Interview format is often misunderstood. A structured interview can consist of either closed questions, open questions or a combination of both. Structured means that the same questions are asked to all participants in the same order. An unstructured interview can also consist of closed questions, open questions or a combination of both.

Questionnaires are not objective. A person can be honest, but they can also be dishonest in their answers; the answers given are therefore subjective. Objective data is fact, such as the result of a blood test or the number of people attending a clinic.

Candidates should read all parts of a question and follow its instruction. If a question states: 'other than 'x', then 'x' should not be written about because 'x' is part of the question and will attract 0 marks.

Section A

There is no need for candidates to re-write the question before beginning their answer.

Some candidates apply the term 'demand characteristics' and 'social desirability' to every answer when in most cases neither term applies.

© 2025

Question **part (c)** requires a general evaluative point that could relate to any study (such as a strength or weakness of a method) but it also requires for the general point to be related to the specific sub-topic/study in the question. Answers often included strengths and weaknesses, but these were often not related to the question, and so restricted marks.

In question **part (c),** when answering the 'in this study', simply rewriting the words in the question is always insufficient and will be awarded 0 marks. Doing this merely shows that candidates can copy a question. Something more needs to be added to show understanding of the study in question. See examples below, such as **Question 5(a)**.

Candidates should not use psychological terms without explanation. Frequently answers were limited to 'it is reductionist' or 'it is useful in everyday life' without further explanation. Stating 'it is reductionist' is identification, it is not evaluation. Explaining why something is or is not reductionist would be a strength or weakness.

Candidates should not use the terms reliability and validity to answer every **part (c)** question for three reasons: **(i)** they do not apply to most questions and so cannot be awarded marks, **(ii)** candidates using the terms often do not know how they apply to the specific question and **(iii)** candidates often confuse the terms.

Many candidates state that a weakness of a case study or longitudinal study is that the researcher has some kind of relationship with the participant. This is very, very rare and will never be an 'automatic' weakness of these methods. Any 'relationship' is often instant dismissal for a psychologist as well as being highly unethical

Section B

Candidates should only answer one question from this section.

Many candidates appeared to assume that they must conduct an experiment whatever the question. An interview, questionnaire or observation are methods independent of an experiment and candidates should not try to make other methods 'fit' into an experimental format.

Some candidates evaluate their plan in **part (a)** by listing strengths and weaknesses. This should not be done because the question does not ask for evaluation; there are no AO3 marks allocated to evaluation; evaluation is done in **Questions (c)(i), (c)(ii)** and **(c)(iii).**

Some candidates included a paragraph of results. This achieves no marks because the question asks for a plan only. Further, the proposed plan has not been carried out, so no actual results are gathered.

Every plan, whatever the question, should include a consideration of ethical guidelines, an essential element.

Candidates need to know the distinction between questionnaire format and technique, and interview format and technique, as stated on the syllabus: Questionnaire technique: paper and pencil (i.e., done by a person with the researcher present), online or postal. Questionnaire format: open and/or closed questions. Interview technique: telephone or face-to-face. Interview format: structured, semi-structured, unstructured.

When using psychometric tests candidates should not use acronyms unless the full title of it is provided first. For example, 'Beck Depression Inventory (BDI)' is fine, with BDI used afterwards. Further, it is insufficient to simply state 'I would use a questionnaire similar to K-SAS' (such as when writing about pyromania, for example).

Answers to **part (a)** questions in this section should include an appropriate plan, having applied a range (four or five) of specific (to the named method) methodological features, each of which should be explained fully, to show good understanding. Candidates should also include appropriate 'general' methodological features such as sample, sampling technique and location of the study. Many answers listed features such as 'I would have a random sample' and 'It would be an independent measures design' without explanation of why it would be a random sample, or how it would be obtained. Elaboration of these general sentences should be included.

In **part (b)(i)**, candidates should describe some relevant psychological knowledge that the whole question is based on. If the question, for example, asks about ways in which pain can be measured, then candidates should describe relevant measures.

© 2025

In **part (b)(ii)**, candidates should explain what aspects of this psychological knowledge their **part (a)** plan is based on. These two question parts must be linked.

In **part (c)**, candidates must refer to what they did in their specific plan rather than give a generic answer that could apply to any study. Use of an example or quoting from their plan would be ideal.

Section B can be considered as follows: A teacher teaches a sub-topic from the syllabus and gives the candidate some psychological knowledge. The teacher then tells each candidate to plan a study using method 'x' to investigate some part of that sub-topic. The candidate plans the study using the psychological knowledge of the sub-topic and they use their methodological knowledge about method 'x'. In the examination, **part (a)** is the plan; **part (b)(i)** is the sub-topic knowledge and **part (b)(ii)** is how the knowledge was used to construct the plan. Exam question parts **(c)(i), (ii)** and **(iii)** then ask about some methodological decisions and evaluation about the plan.

Comments specific to questions

Section A

Question 1

(a) There were two mistakes made by candidates when answering this question; the question asked for an outline and so writing nothing more than 'Y-BOCS' scored 0 marks because it was not an outline; Lovell et al. did not use the MOCI, (Mawdesley Obsessive-Compulsory Inventory) as many candidates thought, they used the BDI (Beck Depression inventory). If candidates identified both Y-BOCS and BDI positive marking allowed the allocation of 1 mark. Answers scoring full marks outlined both the Y-BOCS and BDI.

(b) Validity is not reliability and so many candidates writing about reliability could not be awarded marks. There were definitions of validity, which could be awarded 1 partial mark, but many answers failed to relate this to either the Y-BOCS or BDI. Top answers referred to concurrent validity, comparing data on Y-BOCS with data from another measure such as MOCI.

(c) A common failure in **part (c)** questions across all options is not relating the strength or weakness to the study or sub-topic the question is about. In this question there were many answers that provided two generic strengths of rating scales but failed to support these with any example, particularly the examples of Y-BOCS that had been provided in answering question **part (a)**.

Question 2

(a) Only a few candidates could not answer this question with most being awarded full marks for outlining Miller's theory, which argues that disorders are created when intense positive feelings (intense desire') become linked with specific behaviours (a 'triggering event'), and this creates a state-dependent memory or a 'feeling-state'.

(b) Some candidates suggested that the theory could be supported with evidence but often did not suggest what that evidence could be. Others suggested that experiments could be conducted but again failed to say what those experiments would be. Top answers were much more specific when outlining an IV and DV.

(c) Marks could not be awarded to some candidates who explained that generalisations could not be made because of individual differences. According to Miller, every person has the same mental process (his model) which does not have individual or cultural differences. However, credit was given to those arguing that what differs is the type of impulse control disorder and the specific triggering event for it. Nearly all candidates were awarded 1 mark for stating that Miller used a case study, restricting generalisation, but only a few went on to use the miller's example of 'John'.

© 2025

Question 3

(a) Marks for answers to this question covered the whole mark range. Top answers identified the image message (making hair look good rather than cleaning it) and the quality message (making hair clean rather than looking good). Some candidates wrote about 'soft and hard' sell terms not used by Snyder and DeBono and were often confused about what these terms meant. Like **Question 1(a)** 1 mark was awarded for *identifying* both terms.

(b) Very few candidates answered this question correctly. When interviewing anyone by telephone three important features are; questions should be as simple as possible (as nothing is written/seen); the time taken to complete the questionnaire should be short (e.g. 20 mins for total interview) and answers to questions should be uncomplicated, otherwise people forget what the first answer option is.

(c) As usual a strength (potentially a diverse sample) and a weakness (they are all students with limited experience) were given but very few answers applied these to the Snyder and DeBono study.

Question 4

(a) Nearly all candidates were awarded 1 mark for outlining 'scarcity' ('buy now before there are none left'; 'last day today'; 'limited offer'; 'only two more available') with significant numbers of candidates being awarded full marks for providing an appropriate example.

(b) This question also saw many candidates being awarded full marks because they provided elaboration or an example going beyond their initial reason. Answers could be: 'people like a bargain, want something for less than the usual cost' and 'if people delay, they may miss out, so buy not or be disappointed'.

(c) Few candidates continued to answer in relation to the consumer/ shopper rather than focus on the strengths for the seller, but many other candidates could be awarded full marks. Popular answers tended to be 'sellers can sell more of their product' and 'sellers can sell items at the end of a line, before they are out-of-date'.

Question 5

(a) The question asked for 'guidelines that were followed in this study', so only those addressed by Savage and Armstrong (informed consent, psychological harm and right to withdraw) could receive credit because details of how these guidelines were met were provided in the study. It was not a 'suggest' question where anything appropriate received marks. To match **Questions 1(a)** and **3(a)** candidates *identifying* (rather than *outlining* as the question requested) two guidelines received 1 mark.

(b) There was a wide range of different suggestions. It should be noted that at A2 level the word 'observation' or 'experiment' is insufficient. The type of experiment should be included or one (or more) features of an observation (such as controlled observation or participant observation. It should also be noted that repeating words in the question (for example 'patient satisfaction') without elaboration cannot be awarded marks. More knowledge should be shown such as 'patient satisfaction in relation to sharing or directing style'.

(c) For a strength, most students explained that patients have time to think about their answer rather than being rushed. For a weakness, patients forgetting was common. Again, answers needed to be related to the study to allow full marks to be accessed.

© 2025

Question 6

(a) A few candidates incorrectly wrote about the Tapper *et al*. study who did not use fear-arousal, and no marks could be awarded. Most candidates wrote about the Janis and Feshback study on dental hygiene. Answers receiving full marks referred to the different levels of fear presented to each of the four groups.

(b) This 'suggest' question saw many candidates successfully link fear arousal and providing information, by suggesting, for example, 'presenting an image/video of a fire to scare people and then information provided on how to put out the fire, how to escape from the fire and actions that could be taken to prevent a fire'.

(c) A range of good weaknesses were often clearly explained. Some explained that creating fear would be unethical, whilst others outlined that using fear would lead to 'reverse psychology'. Others proposed that using fear would be a waste of time because people (who smoke for example) simply ignore any warning.

Question 7

(a) **(i)** Many candidates opted for generic answers on controls such as standardising instructions and controlling participant variables because they did not know anything specific to this key study. To control for individual differences, Landry *et al*. asked participants to rate the perceived value of a reward on a 3-item scale and complete a positive and negative scale. Candidates outlining both these measures were awarded full marks.

(b) Many candidates suggested that using a laboratory would decrease ecological validity, and others suggested that the tasks done in a laboratory would not apply in an organisation. What was lacking was application of the suggestion to an example, whether it be from the Landry *et al*. study or any organisation.

(c) Candidates could give two strengths of using a laboratory, such as controlling situational variables and standardising instructions and were awarded partial marks. Only candidates using examples from the Landry *et al*. study to support the strengths were awarded full marks.

Question 8

(a) Many candidates did not understand Scouller's definition of a private level of leadership, using instead a commonsense definition of 'private' and as a result could not be awarded any marks. Scouller defines a private level of leadership as 'handling individuals on a one-to-one basis, where each person should be treated with respect because they have different knowledge, experience, motivation and confidence. This includes appraising, setting goals and recognising talent'.

(b) To be awarded full marks, candidates needed to include a question with answer options and the question asked needed to be relevant to private leadership. Many candidates included all three and were awarded full marks; others had one or more components missing and were awarded either partial or 0 marks.

(c) Questions about 'asking workers' always produces the same pre-prepared response: 'they might not be honest' and 'they might fear losing their job'. The same lack of relating the answer to the sub-topic was evident.

Section B

Question 9

(a) In order to be awarded marks, candidates must answer the question set. If they do not, then only level 1 marks (1 or 2) can be awarded. This meant that candidates writing about *therapies* to treat

© 2025

pyromania were not answering the question. The question asked about investigating *thoughts* about pyromania, so logically an *open* questionnaire should have been used. Instead, many candidates used closed questions which asked about diagnostic features such as 'when was the last time you set a fire', which is not addressing the question about their *thoughts*. Some candidates included an IV and a DV which were inappropriate as the method was questionnaire. Some candidates did not mention ethics at all, and this also reduced the marks that could be awarded. On the positive side there were some excellent answers which received very high marks.

(b) **(i)** For psychological knowledge, candidates should have focused on the diagnostic features of impulse control disorders, specifically pyromania from ICD-11. Many candidates did exactly this and linked the features with questions in their **part (a)** plan.

(ii) Candidates achieving full marks for this question explained how they had *used* the knowledge about pyromania described in **(b)(i)** in their plan. For example, asking a question about a sense of relief or satisfaction in **part (a)** because one feature of pyromania (included in **(b)(i)**) is 'must feel a sense of relief or satisfaction from setting the fire and witnessing it'.

(c) **(i)** Question format is about whether questions are open and/or closed. Candidates could choose open, closed or a combination of the two and in this question part the reason for that choice is explained. Most candidates opted to use a combination of both closed and open questions 'to gather both quantitative and qualitative data' which was illogical as the question asked about 'thoughts' about pyromania and open questions would be apposite.

(ii) As many candidates had applied a combination of both closed and open questions, they struggled to give a weakness of why they had used both. If open or closed only had been used, then a weakness would be obvious (no 'in-depth data' and 'no statistical data' for comparison). Candidates are advised to use only one question format to make answering questions like this much easier.

(iii) Questionnaire *technique* and questionnaire *format* are often confused and answers to this question were no different. Technique relates to whether the questionnaire is given to participants face-to-face (i.e. 'paper and pencil'), postal or online. Candidates who wrote about an open or closed question *format* scored 0 marks.

Question 10

(a) Answers in the higher levels included a range of relevant methodological features, for example type of experiment, IV, DV, controls and design, each outlined in good detail. Such answers also included a range of general features such as sample and sampling technique. For this question an aim is not required, and hypothesis stating was generally poor. Some candidates wrote complex plans with too many variables, for example males with male brands which females did not see, which could not allow any comparison.

(b) **(i)** Psychological knowledge was brand recognition which is, to clarify, 'the extent to which a consumer can correctly identify a product just by viewing the product's logo, tag line, packaging or advertising campaign'. Also relevant was brand recognition in children such as the study by Fischer *et al*. (1991).

(ii) Some candidates based their plans on what was done with children (following the study by Fischer) and adopted it to apply to adults, an appropriate strategy. Many others focused on just one form of brand recognition, typically logo, which was also creditworthy.

(c) **(i)** Some candidates opted for a directional hypothesis but could not provide a reason for their choice. If there is no evidence to suggest a directional hypothesis, candidates should choose non-directional. Many chose a non-directional with the logic that they did not know whether males or females would be better at brand recognition. Candidates should always state their appropriately tailed hypothesis if a question about hypotheses is included.

© 2025

(ii) Many candidates stated simply that they had conducted a laboratory experiment because it was a controlled environment. To achieve full marks, candidates need to relate this to their study, such as giving an example of a control that they had applied in their plan.

(iii) Many candidates could identify an appropriate descriptive statistic such the mean. Often, what they could not do was to give a reason why they chose that particular statistic. In this instance, as this was an experiment with IV and DV, the mean could be 'the mean number of correct brand images recognised by males and by females to give a quick look summary/comparison of any difference'.

Question 11

(a) Some candidates chose this question despite knowing nothing about unrealistic optimism. This was apparent by referring to nothing more than the words 'unrealistic optimism' in their interviews and being unable to write anything for **part (b)(i)** and **(b)(ii)**. Candidates should consider answering the question from their second option if needed. A further weakness for this question was that very little methodological knowledge about interviews was shown. Candidates should always mention interview format (whether it is structured, unstructured, semi-structured) and interview technique (whether the interview is telephone or face-to-face). A very common error is candidates assuming that a structured interview is 'closed' questions' when it means that interviewees are asked the same questions in the same order.

(b) **(i)** Many candidates did not know the term unrealistic optimism and so struggled with this and question **(b)(ii)**. For those who did know the term, their psychological knowledge was that of the work by Weinstein (1980). This research used positive and negative events, finding that participants rated their own chances to be above average for positive events and below average for negative events.

(ii) Many answers were left blank, and many answers focused incorrectly on methodology and sampling. Some candidates did answer the question correctly, linking their **part (a)** plan with the psychological knowledge they had outlined in **(b)(i)**.

(c) **(i)** In response to this question on unrealistic optimism, candidates could choose open or closed questions or a combination of both types. Most candidates opted to use a combination of both closed and open questions 'to gather both quantitative and qualitative data', without giving a reason as to why this was used in their plan.

(ii) As many candidates had applied a combination of both closed and open questions, they struggled to give a weakness of why they had used both. Choosing only one question format would make answering a strength or a weakness question like this much easier to answer.

(iii) Many candidates do not know about interview format. Candidates often wrote 'I would have a semi structured interview in order to obtain both quantitative and qualitative data'. This is incorrect. A semi-structured interview is when there are some fixed questions but with the option to ask additional questions. These questions can be entirely closed, or they could be entirely open. Similarly structured interviews can include open, closed or both types of question.

Question 12

(a) Some excellent plans used appropriate methodology to investigate praise and recognition. However, there were many plans which could only be awarded Level 1 or 2. Some candidates decided to conduct an experiment with IV conditions of praise or recognition using an independent measures design. This would never determine whether a worker preferred praise or recognition, especially when the question required an interview to be conducted. Many interviews were lacking in detail such as where the interviews would be conducted, who the interviewer would be and there was confusion about the interview format. Candidates frequently assumed incorrectly that a semi-structured interview consisted of both open and closed questions.

© 2025

(b) **(i)** Relevant psychological knowledge here was of intrinsic motivation, specifically praise such as 'well done' or a few words of encouragement from a manager to a worker or recognition for working hard, showing responsibility, leadership or other skills which could be in the form of an award or a certificate. A few candidates incorrectly focused on the work of Maslow.

(ii) Answers in the top Levels explained how they designed questions that could be asked about praise and recognition and whether they effectively motivated workers based on what they had written in **(b)i**. Answers that focused on some aspect of methodology, such as sampling and continued to write about features of Maslow's hierarchy, showing no knowledge of any intrinsic motivator, received no marks.

(c) **(i)** Many candidates chose an opportunity sampling technique and the stated reason for this choice was that workers in an organisation could be invited to participate. However, frequently there was no detail on how this would be achieved. Some answers merely stated, "I asked people at random' which achieved 0 marks.

(ii) A commonly stated weakness was that 'the sample is restricted' and often this had no elaboration to explain why the sample was restricted. Additionally, marks were restricted because there was little or no reference to the organisation the sample was selected from.

(iii) Many candidates do not know about interview format. Candidates often wrote 'I would have a semi structured interview in order to obtain both quantitative and qualitative data' but this is incorrect. A semi-structured interview is when there are some fixed questions but with the option to ask additional questions. These questions can be entirely closed, or they could be entirely open. Similarly structured interviews can include open, closed or both types of question.

© 2025

PSYCHOLOGY

Paper 9990/43

Specialist Options: Application and Research Methods

Key messages

- What has been learned from the AS component of the syllabus should be transferred to the A2 component. For example, at AS, candidates learn about methodology, such as experiments, which also apply to A2.
- Questions should be read carefully, ensuring that the focus is on what the question asks.
- For ***Section A*** answers, candidates should relate their answer to the study in the question or include an example. Questions frequently end with 'in this study' and so the answer should be related to that specific topic area/study.
- All terminology should be explained. Writing 'it is valid and reliable' for example, is insufficient without explanation, application or example.
- The syllabus includes 'example studies' such as 'e.g., Oldham and Brass (1979)'. Example studies can be substituted for alternatives, but these alternatives must cover the same or very similar content to the example study. If the Oldham and Brass study is substituted, the alternative study must be about a move to open plan offices and the data that was gathered from that move. The alternative cannot be about something different.

General comments

Some candidates answered questions from one option only. Other candidates who correctly answered two options, sometimes performed considerably better in one option than the other.

A significant number of candidates answered two questions from ***Section B*** instead of one (only one of these ***Section B*** responses can receive credit). Candidates are advised to read the instructions on the front cover of the question paper and to read the heading instructions for each question section.

Candidates should double check that the terminology they use in their answers is correct. Often terms such as reliability and validity were used interchangeably, as were qualitative and quantitative, and independent and dependent variables. There was also confusion with the terms format and technique in relation to questionnaires and interviews.

Interview format is often misunderstood. A structured interview can consist of either closed questions, open questions or a combination of both. Structured means that the same questions are asked to all participants in the same order. An unstructured interview can also consist of closed questions, open questions or a combination of both.

Questionnaires are not objective. A person can be honest, but they can also be dishonest in their answers; the answers given are therefore subjective. Objective data is fact, such as the result of a blood test or the number of people attending a clinic.

Candidates should read all parts of a question and follow its instruction. If a question states: 'other than 'x', then 'x' should not be written about because 'x' is part of the question and will attract 0 marks.

Section A

There is no need for candidates to re-write the question before beginning their answer.

Some candidates apply the term 'demand characteristics' and 'social desirability' to every answer, when in most cases neither term applies.

© 2025

Question **part (c)** requires a general evaluative point that could relate to any study (such as a strength or weakness of a method) but it also requires the general point to be related to the specific sub-topic/study in the question. Answers often included strengths and weaknesses, but these were often not related to the question, and so restricted marks.

In question **part (c),** when answering the 'in this study' response, simply rewriting the words in the question is always insufficient and it will be awarded 0 marks. Doing this merely shows that candidates can copy a question. Something more needs to be added to show understanding of the study in question. See examples below, such as **Question 5(a)**.

Candidates should not use psychological terms without explanation. Frequently answers were limited to 'it is reductionist' or 'it is useful in everyday life' without further explanation. Stating 'it is reductionist' is *identification*, it is not *evaluation*. Explaining why something is or is not reductionist would be a strength or weakness.

Candidates should not use the terms' reliability and validity to answer every **part (c)** question for three reasons: they do not apply to most questions and so cannot be awarded marks, candidates using the terms often do not know how they apply to the specific question and candidates often confuse the terms.

Many candidates state that a weakness of a case study or longitudinal study is that the researcher has some kind of relationship with the participant. This is very, very rare and will never be an 'automatic' weakness of these methods. Any 'relationship' is often instant dismissal for a psychologist as well as being highly unethical.

Section B

Candidates should only answer one question from this section.

Many candidates appeared to assume that they must conduct an experiment whatever the question. An interview, questionnaire or observation are methods independent of an experiment and candidates should not try to make other methods 'fit' into an experimental format.

Some candidates evaluate their plan in **part (a)** by listing strengths and weaknesses. This should not be done because: the question does not ask for evaluation; there are no AO3 marks allocated to evaluation and evaluation is done in **Questions (c)(i), (c)(ii)** and **(c)(iii).**

Some candidates included a paragraph of results. This achieves no marks because the question asks for a plan only. Further, the proposed plan has not been carried out, so no actual results are gathered.

Every plan, whatever the question, should include a consideration of ethical guidelines, an essential element.

Candidates need to know the distinction between questionnaire format and technique, and interview format and technique. As stated on the syllabus: questionnaire technique: paper and pencil (i.e., done by a person with the researcher present), online or postal. Questionnaire format: open and/or closed questions. Interview technique: telephone or face-to-face. Interview format: structured, semi-structured, unstructured.

When using psychometric tests candidates should not use acronyms unless the full title of it is provided first. For example, 'Beck Depression Inventory (BDI)' is fine, with BDI used afterwards. Further, it is insufficient to simply state 'I would use a questionnaire similar to K-SAS' (such as when writing about pyromania, for example).

Answers to **part (a)** questions in this section should include an appropriate plan, applied a range (four or five) of specific methodological features, each of which should be explained fully, to show good understanding. Candidates should also include appropriate 'general' methodological features such as sample, sampling technique and location of the study. Many answers listed features such as 'I would have a random sample' and 'It would be an independent measures design' without explanation of why it would be a random sample, or how it would be obtained. Elaboration of these general sentences should be included.

In **part (b)(i)**, candidates should describe some relevant psychological knowledge that the whole question is based on. If the question, for example, asks about ways in which pain can be measured, then candidates should describe relevant measures.

© 2025

In **part (b)(ii),** candidates should explain what aspects of this psychological knowledge their **part (a)** plan is based on. These two question parts must be linked.

In **part (c)**, candidates must refer to what they did in their specific plan rather than give a generic answer that could apply to any study. Use of an example or quoting from their plan would be ideal.

Section B can be considered as follows: A teacher teaches a sub-topic from the syllabus and gives the candidate some psychological knowledge. The teacher then tells each candidate to plan a study using method 'x' to investigate some part of that sub-topic. The candidate plans the study using the psychological knowledge of the sub-topic and they use their methodological knowledge about method 'x'. In the examination, **part (a)** is the plan; **part (b)(i)** is the sub-topic knowledge and **part (b)(ii)** is how the knowledge was used to construct the plan. Exam question parts **(c)(i), (ii)** and **(iii)** then ask about some methodological decisions and evaluation about the plan.

Comments specific to questions

Section A

Question 1

(a) **(i)** Candidates either knew the two drugs used in the trials which were Nalmefene and Naltrexone and were awarded 2 marks, or they did not. Some candidates were awarded marks for misspellings that were close enough to the original. Candidates writing 'opioid antagonists' were awarded 0 marks.

(ii) Many candidates achieved full marks by stating '16 weeks' and '18 weeks'. It was not necessary to state which time period belonged to which trial or drug.

(b) Candidates frequently were awarded full marks for excellent answers showing good understanding of why placebos are used. For example, 'without a placebo group, participants would know that they are receiving an active drug and this alone might result in a change in behaviour' and when nalmefene or naltrexone are identified as the active drug, the answer was worth full marks. Some candidates were awarded 1 mark for 'it acts as a baseline'.

(c) Candidates were often confused about what location meant and frequently did not know that this key study was conducted in different locations. There were incorrect suggestions that it would cost more and participants would have to travel, without appreciating that different participants were used in each location. Also, there were comments about cultural differences. However, the participants were all gamblers, and the drug works physiologically on every human in the same way.

Question 2

(a) The syllabus requires candidates to know 'a case study of schizophrenia' and most chose to outline the first case study described in the study by Aneja (2008). Any two details achieved marks, such as aggressive father, living with mother and grandparents, showing aggressive outbursts and early onset schizophrenia, amongst others. Some candidates provided detail too brief for marks to be awarded such as 'a child with schizophrenia'. The Aneja study is an 'e.g.' study, so any appropriate alternative could have been substituted. Such instances could only receive credit if there was sufficient detail for an examiner to confirm the legitimacy of the answer. Providing a name and date would be ideal.

(b) Many candidates stated, 'the person might not tell the truth', which is correct, but this can apply to anyone and any study, not specifically someone with schizophrenia as the question requested. The question concerned reliability which is about consistency and a person who has schizophrenia, perhaps because of an impaired sense of reality, might not be consistent if data is gathered from them on two (or more) occasions.

© 2025

(c) Generic strengths were awarded marks, one for each strength such as 'studied in depth/detail'; 'can apply a range of different methods', etc. Candidates struggled to provide examples to support each strength, despite describing a case study in question part (a) where there were often many examples that could have been used.

Question 3

(a) (i) Some candidates wrote very impressive answers (awarded full marks) showing excellent knowledge of the procedure of the pre-test, including details such as the three pairs of product variants, the 12-item potency construct, and the 7-point rating scale. Others could only write that 'it is about yogurt packaging' (for 1 mark) and one candidate wrote 'what pre-test' scoring 0 marks.

(b) There were some excellent answers provided by candidates who suggested that there could be a lack of validity. What made such answers excellent was that they explained *why* it might lack validity and gave an example, such as testing something that would be unrealistic, like a square yogurt pot. Candidates writing nothing more than 'one effect is lack of reliability and validity' were awarded 0 marks.

(c) This was similar to other **part (c)** questions where candidates often gave a strength and/or weakness without relating it to the question. The answer could have been: 'one strength is that the location is congruent with the study' (1 mark) 'because yogurt in packaging is sold in a supermarket' (+1 mark). Here there is a generic strength and an applied example of yogurt and supermarket.

Question 4

(a) Nearly all candidates were awarded 1 mark for their answers describing retail atmospherics as the controllable characteristics of retail space which entice customers to enter the store, shop and purchase a product. Most candidates were awarded 2 marks, usually for providing an example, such as landscaping, window displays, background music or pleasant odours.

(b) Candidates frequently used 'classical music' and referred to the study by North *et al*. where the music increased spending, or to the study by Chebat and Michon who investigated the effect of odour on shopper ambience. The best answers related a specific retail atmospheric to the PAD model (pleasure/arousal/dominance).

(c) Although the question asked about strengths for the retailer (seller), some candidates still wrote about the purchaser (customer). For the retailer, strengths might be that they can apply psychological knowledge: if they have an 'up-market' establishment playing classical music it could lead to pleasure in customers and an increase in sales. Similarly, a retailer could use the findings by Chebat and Michon and create arousal in customers by spraying a pleasant odour in their store.

Question 5

(a) **(i)** Outlining the procedure was straightforward for most candidates who were awarded full marks for including the physiological (relaxation and diaphragmatic breathing) and psychological components (imagining a peaceful scene of their own choice). Some candidates only included one of these components so could only be awarded 1 partial answer mark.

(ii) The experimental design used was independent groups, and correct identification was awarded 1 mark. Continuing to identify what the groups were (control, relaxation only, and relaxation and imagery) resulted in the second mark being awarded.

(b) Many candidates suggested using a repeated measures design and the best answers went on to explain that this would not work because of a knock-on effect, where if relaxation and imagery were taught first it would make the control or relaxation only conditions redundant. Some candidates suggested that it would cause fatigue or perhaps lead to attrition.

© 2025

(c) Candidates sometimes struggled to suggest a strength of studying a sample of female participants, whilst others made good points such as 'stress reduction techniques (like relaxation and imagery) can be tailored specifically for females rather than a generic 'one size fits all'' or that 'techniques target specific female illnesses'. For the weakness, nearly all candidates suggested that the techniques may not apply to men.

Question 6

(a) This question was answered incorrectly by many candidates who outlined the *aim* of the study or the *findings*. The question asked for an outline of the *methodology* that was used. Two methods were used: a questionnaire-based survey and interviews (one of which was conducted in the home and a telephone interview to check on medication).

(b) Candidates could suggest an ethical guideline but often not relate it to this study. Some candidates identified a guideline, such as informed consent, but merely stating 'informed consent' was insufficient. There needed to be a comment about what informed consent is, or who is giving the consent, for 1 mark. Relating this guideline to the study (for 2 marks) could have been done by stating that parents had to give informed consent because the children were aged from 1.5 years to 6 years for example, which shows knowledge of the study.

(c) Answers to this question followed the typical pattern for **part (c)** answers, with generic weaknesses being provided such as 'a parent may give a socially desirable answer' with no elaboration or attempt to relate this to medication adherence in children. Candidates should appreciate that all question parts, i.e. **(a)**, **(b)** and **(c)** are part of the same question about a study or sub-topic and that for **part (c)** answers what has been written in **part (a)** can often be applied to strengths and/or weaknesses in **part (c)**.

Question 7

(a) **(i)** Some candidates could not provide any correct feature, often guessing, whilst others had little problem stating two features from: 136 participants, all psychology students, all from National Open University of Spain and that 53 per cent were women.

(ii) After reading the narrative, participants were asked to evaluate a supervisor (awarded 1 mark) and who was occupying that position for a trial period (1 additional mark). Marks were also awarded for stating 'asked to collaborate in the evaluation process by completing an anonymous questionnaire'. There were some perfect answers, but quite a few candidates could not provide a correct answer despite this being a key study.

(b) Some answers were awarded 1 partial mark for stating 'so all the conditions of the IV could be tested', whilst those who went on to state what the IV's were, and so relate the generic comment to this study, were awarded the full 2 marks. The IV conditions were: male stereotypical versus female stereotypical leadership style, and male leader versus female leader.

(c) Like many **part (c)** answers, candidates could provide two generic strengths (in this instance of using experiments) but could not relate these strengths to the study itself. For example, a candidate might write about the IV being manipulated (but not mention that the IV was male/female leader and male/female leadership style). Additionally, they might write about the standardisation of the procedure but not mention the evaluation of the leader always being the same, or the use of 14 adjectives, for example. Generic answers will never be awarded full marks; answers that are related/applied could be.

© 2025

Question 8

(a) Very few candidates could provide two questions from Walton's QWL. The syllabus is '*Walton's quality of working life (QWL) including eight conditions and QWL evaluation scale*', and so to know two questions from this scale should not be a problem for any candidate. The question asked for any two from the first four appearing on the QWL. Candidates are advised that the whole syllabus will be examined and so the whole syllabus should be revised thoroughly.

(b) Questions about reliability are the same whatever option it is related to. In this instance, there needed to be a comment about the same person taking the same test (i.e. the QWL) on a different occasion to test consistency. In other words, test-retest reliability. Candidates suggesting inter-rater reliability or validity were awarded 0 marks.

(c) In response to this question, candidates often gave very brief, single sentence answers that were very general, meaning that they could apply to any question from any option. Candidates should always apply the strength or weakness (or two weaknesses as in this instance) to the study or sub-topic the whole question (**parts (a)**, **(b)** and **(c)**) is about. Questions asking about weaknesses of questionnaires require answers about the *questionnaire* (such as a restricted scale or poor question wording) and not about a *participant* (who may lie or give a socially desirable answer).

Section B

Question 9

(a) There were several errors which restricted marks.

(i) Many candidates did not know what the term 'imaginal desensitisation' involved and in their answers never explained the treatment process, writing nothing more than the term itself, or they confused it with covert sensitisation (a very different therapy).

(ii) Some candidates decided to conduct an *experiment* comparing a desensitisation group with a non-desensitisation group with little or no observation. Similarly, some candidates decided to conduct an *interview*, the only reason for this appearing to be to gather qualitative data. If the question asks for an observation, then the plan should be about an observation *in detail*. Any additional method will gain no additional marks.

(b) (i) Following from **part (a),** many candidates could not answer this question part at all, showing no knowledge of imaginal desensitisation. Other candidates wrote incorrectly about covert sensitisation. Imaginal desensitisation involves teaching progressive muscle relaxation and then the person visualises themselves being exposed to the situation. Some candidates wrote legitimately about types of gambling behaviour.

(ii) If there was no answer in **(b)(i)** then no marks could be awarded in this question part as it links **(b)(i)** with **(a)**. Candidates explaining the link related details of their structured observation in **part (a)**, such as their response categories. Others had categories for behaviour before treatment and after treatment.

(c) **(i)** Candidates often struggled to decide whether a gambling casino (where a participant could be observed after being treated) was a natural environment or whether it was a controlled environment. Both were awarded credit. Some candidates chose to set up an artificial laboratory scenario which was correctly labelled as a controlled observation.

(ii) Most candidates gave an appropriate reason for their choice of structured observation using response categories but often there were no examples of what the categories would be. Others used a simple tally chart with a yes/no for a category of 'gambles (stops to play) when passing a fruit/slot machine'.

(iii) Quantitative data was the most logical type to gather from response categories with statistical analysis and comparison as a strength. However, many candidates stated that they applied both quantitative and qualitative data and struggled to give a reason why they had used both.

© 2025

Candidates are advised to gather only one type of data to make answering questions much easier, and in this instance quantitative data was apposite.

Question 10

(a) All candidates planned an experiment, as the question required, but many gathered data through questionnaire or interview. This can be done, but only after full details of the named method have been included. Many candidates appeared to use a questionnaire or interview simply to say, 'I gathered qualitative data', when this was irrelevant to the effect of music on *spending*. Candidates are advised not to include qualitative data in questions where it is not required. For this question, many candidates incorrectly had the IV as different *types of music* (i.e. they replicated the North et al. study) and were awarded level 1 or 2 marks. The question asked for classical music in different *types of restaurants*.

(b) **(i)** For psychological knowledge, the most appropriate research would be that by North et al. (2003). This research played different types of music in the same restaurant and found that classical music resulted in increased spending when compared to pop and no music.

(ii) This question part, like all other **part (b)(ii)** questions, required an explanation to show how what was described in **10(b)(i)** informed the plan in **part (a)**. For example, candidates who outlined that North et al. used a classical music cd of 76 minutes, also used a classical music cd of 76 minutes. Similarly, North et al. studied an 'up-market' restaurant so many candidates also included an 'up-market' restaurant as one of their types. In other words, candidates based their plan on things taken from the North et al. study.

(c) **(i)** Many candidates listed the controls they had applied; the same person gathering data, the type and length of music and keeping menu items the same. Often candidates did not explain the reason why they had applied these controls when this is what the question asked. Some candidates wrote about generic controls here for the first time even though they had not included any controls in their **part (a)** plan.

(ii) Some candidates wrote about one *strength* of the controls they had applied and were awarded no marks as this did not answer the question. Others struggled to find any weaknesses. Some candidates wrote about a weakness of the controls that they had not applied in their plan. The best answers focused on having too many controls reducing ecological validity, or that too many controls could make participants suspicious and perhaps cause demand characteristics.

(iii) Top marks were achieved by candidates who knew the findings of North *et al*. and the requirements of this question. This meant that some candidates opted for a directional hypothesis because they expected classical music to increase spending in 'up-market' restaurants and not in other restaurants they had included in their plan.

Question 11

(a) There were two problems which reduced marks for this question.

(i) The inclusion of irrelevant questionnaires and interviews, as explained above, see **Question 10(a)**.

(ii) Candidates did not know what biofeedback was. To investigate the effectiveness, candidates need participants to have biofeedback and so they should outline the procedure of biofeedback. It is insufficient to write nothing more than the word 'biofeedback' throughout the answer. Top answers had an IV of biofeedback and no biofeedback with a DV of decrease in level of tone on a scale of 1–10 (if using GSR). Designs were usually independent measures with a range of controls applied.

(b) **(i)** For psychological knowledge, the study by Budzynski and Stoyva (1969) was most appropriate, although as this is an 'e.g.' study any suitable alternative could be substituted. The Budzynski and Stoyva study used biofeedback to reduce forehead muscle tension. Some candidates described this

© 2025

study in detail and were awarded 4 marks, others provided a brief outline, and others could not identify any psychological knowledge at all.

(ii) Candidates awarded full marks wrote about how they used their knowledge of biofeedback to inform their plan. For example, because Budzynski and Stoyva focused on forehead muscle tension, so did many plans. Similarly, as their participants received feedback through an auditory 'tone' this also featured in many plans.

(c) **(i)** Most candidates chose to apply an independent measures design where different participants did different conditions, usually of biofeedback or no biofeedback. Some candidates chose a repeated measures design even though they had feedback or no feedback conditions.

(ii) Many candidates stated simply that because they had conducted a laboratory experiment in a controlled environment, a weakness was that it lacked ecological validity. Answers ending here were awarded 1 mark, but those going on to relate it to the study could be awarded an additional mark when stating, for example, 'stress in university students is when they are sitting at a desk looking at past examination papers'.

(iii) The choice of dependent variable is crucial in determining the effectiveness (or not) of the biofeedback. Following on from AS knowledge, the DV should be fully operationalised in order for full marks to be awarded, such as the 'DV is the decrease in level of tone on a scale of 1–10 when measuring GSR (Galvanic Skin Response)'. Answers such as 'the DV is level of stress' was too vague for marks to be awarded.

Question 12

(a) Answers to this question showed the same flaws as answers to other options, namely not using the correct method, not explaining terms (such as job enlargement) and showing poor methodological knowledge. In reverse order, many candidates confused question format and questionnaire technique despite the syllabus specifying the differences. Technique relates to whether the questionnaire is given to participants face-to-face (i.e. 'paper and pencil'), by post or online. Often candidates wrote 'gave the questionnaire to participants'. Many answers referred to 'job enlargement' without asking anything specific and so showing no knowledge of the term at all. Others focused on job satisfaction and often only one question was asked, such as 'do you feel more satisfied after your job enlargement'. Finally, many candidates decided to plan an experiment with IV and DV and no marks could be awarded for using the wrong method.

(b) **(i)** Relevant psychological knowledge here could include two things: job enlargement which transforms the jobs to include more and/or different tasks and has two types: horizontal and vertical enlargement, or job satisfaction rating scales such as the JDI could have been used, as could Walton's quality of working life evaluation scale.

(ii) Relevant psychological knowledge mentioned above in **(b)(i)** was lacking in most answers which made answering this question part very difficult. Candidates are advised that if they have no knowledge of a particular sub-topic then they should consider answering the question from their second studied option.

(c) **(i)** Many candidates answered this question part incorrectly. The question stated questionnaire *technique* and NOT questionnaire *format*. Technique relates to whether the questionnaire is given to participants face-to-face (i.e. 'paper and pencil'), postal or online. Candidates who wrote about an open or closed question *format* scored 0 marks.

(ii) Candidates writing about question format in **12(c)(i)** also provided incorrect answers to this question part when giving a weakness of their question format. Candidates answering the question correctly often gave a good weakness (online could raise doubts about the veracity of the data; postal may not get many returns; face-to-face may create demand characteristics) but often failed to relate their answer to what they had written in their plan.

© 2025

(iii) Question format is about whether questions are open and/or closed. Some candidates were confused because they had already incorrectly written about question format in parts **(c)i** and **(c)ii**. Candidates could choose open or closed or a combination of the two. Most candidates opted to use both types ‘so I could gather quantitative data’. It is advised that one question type should be used to avoid confusion.

© 2025