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 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, otherwise it can mean a candidate may run out of time for other questions.

Candidates should be able to give full definitions of terms 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.

General comments

The marks achieved by the candidates sitting this examination covered a wide spread of possible marks. Three-quarters of the candidates scored 25 marks or below (and one-quarter scored 10 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 (all 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 3, 5, 6, 9** and **10** had the strongest correlation with a candidate's overall score.

Comments on specific questions

Question 1

(a) The majority of candidates could state that four elephants could successfully complete the trunk-wash test. Common errors included stating the total sample size or choosing a sample size from a different core study.

(b) (i) (ii) Performance on **Question 1bi** was better than **Question 1bii**. The majority of candidates could identify the primary reinforcer but only a minority of candidates could identify the secondary reinforcer. For **Question 1bi**, a common error was to answer with food or treat and for **Question 1bii** was to answer banana, clicker, or punishment.

(c) Only a small minority of responses could clearly outline the behaviour. Common choices included focusing on the trunk being still but that the elephant could move other body parts to fulfil 'steady'. There were many examples of responses being tautological. That is, candidates would often state that 'steady' was holding the trunk steady. Responses like this cannot be given credit as there is no way of knowing if the candidate is simply copying the question to try and be given credit. For definitions, candidates need to be able to define the term without using the term.

Question 2

(a) Only a minority of responses could identify the two possible ways the test ended for a participant. These were that the figures collided or that the participant pressed a key/space bar to stop the figure from approaching any further. Common examples of incorrect responses included that the computer had technological issues or that they had completed all 96 trials. Candidates are expected to know the full procedure of all core studies. This question had the joint second highest rate of blank responses.

(b) Stronger responses could clearly identify a correct ethical issue related to the study by Perry *et al*. and then provide an example from the study to show why there was an issue. Popular examples included the potential for harm as mentally invading personal space and deception in relation to the private conversation causing anticipated harm. However, there were some significant knowledge-based errors based around the ethical guideline of deception. The first example was simply stating that them not actually interacting with someone at a later date 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 here, candidates were claiming that the participant not knowing if the were given placebo or oxytocin meant the guideline of deception was broken. This is not breaking the guideline. Candidates are expected to know details about all ethical guidelines for human participants.

Question 3

(a) The majority of responses could provide a full aim of the study. Popular choices included investigating the role of classical conditioning and investigating whether imagery exposure could help the boy with his phobia. There were examples when candidates wrote about disgust in their chosen aim but this was mentioned in the stem for the question so these could only be awarded partial credit. It is important that candidates read questions carefully and do not repeat what has already been given to them in the stem and/or question.

(b) Stronger responses could identify two other features of the sample. Popular choices were the boys age, his ethnic background and his experience at school that triggered his phobia. Some candidates repeated the two features already given in the question and could therefore not be

given credit. Common errors included providing results from the study or outlining what the boy was expected to do during therapy sessions.

(c) This type of question requires two parts to be awarded two marks. Firstly, we need to know what the application is and secondly, we need to be told how this would be achieved. Stronger responses could clearly cover both parts with popular choices being utilising what happened during the sessions to use it with other (different) phobias and educating therapists on which techniques cold be more successful with clients. There were some responses that simply outlined how someone at home could use a hierarchy to help them with their own phobia, but this is not answering the question. There were virtually no unethical responses.

Question 4

Candidates are required to know the results of studies including key quantitative results. Stronger responses could clearly compare the dream recall from REM and NREM sleep, proving one correct data point. The most common correct response was comparing the number of dreams recalled under both stages of sleep. It is important for candidates to provide a full comparison when presenting results so stating that there was more dream recall in REM without telling the Examiner what this was compared to can only ever be awarded partial credit. There were many examples of candidates mis-representing the data they provided in their responses. For example, a significant minority of responses claimed that 80 per cent of participants recalled a dream in REM. This is incorrect as every participant recalled in REM. Common errors included providing examples of dreams and length of time in dream. Neither of these were answering the question.

Question 5

(a) Stronger responses could clearly outline one main assumption of the biological approach. The most common choice was about how specific biological mechanisms (e.g., brain, hormones) affect our behaviour or cognitions. Some candidates provided assumptions from different approaches or gave specific findings from a biological approach study.

(b) Many responses could provide a key finding that linked to a biological concept. However, only a small minority of responses could then explain why it links back to the assumption they presented in **Question 5a**. Candidates tended to simply re-write the assumption they had already given in **Question 5a** with no attempt at explaining how it links back. Responses need to be explicit and clearly use examples from the study. A common correct response was explicitly stating that the differences in human behaviour could be understood as brain density increased in the mindfulness group (internal factor) as a result of participating in a mindfulness-based stress reduction program (external factor). This example would be awarded two marks. This question had the highest blank response rate on this paper.

Question 6

(a) This type of question requires candidates to present generic psychological information linked to the named study. Stronger responses could clearly describe eyewitness testimony, what social demands are and/or what a false positive response is. However, the majority of responses focused on what Pozzulo et al. did in their study, which can only ever be awarded the one available example mark. To improve, candidates need to give generic descriptions of psychological concepts that Pozzulo et al. based their study on.

(b) Candidates were equally likely to choose Eric or Quinn. For Eric, popular choices included internal validity and control of extraneous variables. Stronger responses could clearly provide evidence from the study to explain why Eric was correct. For Quinn, the most popular choice was ecological validity. As with Eric, stronger responses could clearly provide evidence from the study. There was a significant minority of responses that confused validity with reliability, with a focus on standardisation and replicability which could not be given any credit as it is not answering the question.

Question 7

The average score on this question was around two marks. Stronger responses that were awarded five could clearly describe what one trial consisted of with specific examples and a logical progression to the response. However, a significant minority of responses focused on what happened post-trial and could not be awarded credit (e.g., how the data were analysed). It is very important for candidates to read each question carefully

and ensure that their response is covering the demands of the question set rather than simply writing all they know about a particular part of the core study.

Question 8

Stronger responses focused on the situation Imelda was in and gave appropriate advice to help improve her child's behaviour. Popular technique included using a model to show peaceful behaviour, using a male role model, and modelling specific toy-related behaviour whilst the child observed. However, a significant minority of responses focused on trying to explain why the child was behaving as they did in the scenario which could not be awarded credit as it was not answering the question. This was another example where candidates needed to read the question carefully and give the advice that Imelda had asked them for. Some common errors focused on using positive reinforcement (not social learning) and/or punishment (unethical).

Question 9

(a) The majority of responses could describe two features of the sample used in the study by Andrade. Popular choices included the sampling technique, that they had just completed a different study and that the majority were female. Some candidates described the procedure rather than the sample of participants or gave incorrect information about the sample (e.g., that the sample was all male participants). It is essential for candidates to know the features of the samples used in all 12 core studies.

(b) Stronger responses could clearly explain two differences. Popular choices to compare the studies on included experimental designs, sampling techniques and validity. 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 to 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 the sample, yet a number of candidates did use the sample 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 cognitive.

In relation to this specific series, candidates need to ensure that their chosen 'differences' are based on psychological principles and/or are factually correct. There were several of these errors that were seen repeatedly:

1. 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.
2. Comparing Andrade and Baron-Cohen on the number of groups they used. This is not psychological so could not be given any credit
3. Explaining that Pozzulo was not standardised as the children and adults had different procedures. All children completed the same procedure and all adults completed the same procedure so the study was standardised and could be replicated.

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 Milgram in depth and in terms of two strengths and two weaknesses with at least one of these points covering the named issue of the sampling technique used. 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 Milgram. 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 Milgram 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 give Level 5.

Candidates need to know that any description of the study does not gain credit in these type of questions as it is testing their evaluation skills *only*. There were some common errors evident in responses, for example, that operationalising a variable means it automatically has validity, that it was a laboratory experiment, and that Mr. Wallace was the stooge, and he vocalised and screamed. 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 <u>why</u> 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 <u>each</u> evaluative point made. These are the requirements for a Level 5 response. The average response was Level 2 (three marks) for this cohort.

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 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, otherwise it can mean a candidate may run out of time for other questions.

Candidates should be able to give full definitions of terms 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.

General comments

The marks achieved by the candidates sitting this examination covered a wide spread of possible marks. Three-quarters of the candidates scored 27 marks or above (and one-quarter scored 45 marks or above). Many 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 limited blank responses in this series. However, as positive marking is used, candidates should attempt all questions even if they are unsure of the response they are providing.

Performance on **Questions 2, 6, 9** and **10** had the strongest correlation with a candidate's overall score.

Comments on specific questions

Question 1

(a) A substantial majority of candidates could name one of the objects. Common errors included naming an object from the CIDS in Experiment 1.

(b) The majority of candidates could partially outline how participants were deceived. The most common choice was them being told they would discuss personal matters at a later date. Some responses focused on the application of placebo or oxytocin which was not answering the question.

(c) Stronger responses could focus on the empathy group and the role of oxytocin within Experiment 2. The most popular choice was to compare within the high empathy group when given placebo or oxytocin. It is important when presenting quantitative results that a comparison is provided by the candidate or only partial credit can be awarded. There were some errors presented by candidates when presenting comparisons between high and low empathy groups in the oxytocin conditions by getting the comparison the wrong way round. Some responses did not focus on the empathy grouping so subsequently presented an incorrect result.

Question 2

(a) Only a minority of responses could identify the two features of the target. Popular choices included her age and that she was a candidate. Most responses identified one feature. Common errors included focusing on Dora the Explorer or presenting features not mentioned in the Pozzulo et al. study. It is important for candidates to know how all 12 core studies were set-up methodologically.

(b) Responses tended to present what the strength could be or an example from the study. Stronger responses did both, focusing in the evaluative part on reliability (that standardisation improves replicability) and then presenting a clear example from the study by Pozzulo et al. Standardisation or replicability presented by itself is not evaluative as it is simply describing part of the study. It is important for candidates to read the question carefully to ensure they are evaluating and not just describing standardisation. Some candidates provided a strength of the study but not in relation to reliability, for example, internal validity.

Question 3

(a) The majority of responses could provide two features of the sample. Popular choices included the sample size and that they were all males. Common errors included describing what the sample did in the study or providing incorrect demographics, sometimes even the sample information from a different core study.

(b) Stronger responses could present a full conclusion. Popular choices included that people are more obedient than expected and that people will follow the orders of an authority figure even if it means harming another person. Some candidates presented a result and could not be given credit. A result is based on factual data collected in the study whereas a conclusion is the general 'meaning' of the data. It is important for candidates to be able to differentiate between a result and a conclusion.

(c) This type of question requires two parts to be awarded two marks. Firstly, we need to know what the application is and secondly, we need to be told how this would be achieved. Stronger responses covered both parts, with popular choices being using formal dress in the classroom for teachers to be more likely to be obeyed and using the findings of the study to create whistleblowing policies in organisations. There were some responses that attempted to explain everyday behaviours which is not the focus of this style of question. Responses similar to 'it helps us to understand obedience' are not answering the question.

Question 4

Candidates are required to know the results of studies, including key quantitative results. Stronger responses could clearly compare the two types of toys played with by the male monkeys in terms of either frequency of play or duration of play. Some candidates could then present a correct data point to highlight the difference. The most popular choice was comparing the wheeled and plush toys on duration of play. Some candidates did provide 'correct data' but for the wrong comparison – for example, presenting duration data when the comparison was for frequency. Some candidates presented sex difference results, but this was not answering the question. It is important for candidates to read the question carefully to ensure they are presenting the correct results.

Question 5

(a) Stronger responses could clearly outline one main assumption of the learning approach. The most common choice was learning through types of conditionings. Some candidates provided assumptions from different approaches or gave specific findings from a learning approach study. Some responses stated that 'the learning approach is about how we learn' but this could not be given any credit. Assumptions need to be generic, rather than focused on specific mechanisms like positive reinforcement.

(b) Many responses could provide a key finding that linked to a learning concept. However, only minority of responses could then explain why it links back to the assumption they presented in **Question 5(a)**. Candidates tended to simply re-write the assumption they had already given in **Question 5(a)** with no attempt at explaining how it links back. Responses need to be explicit and clearly use examples from the study. A common correct response was explicitly stating that the boy was helped by his mother in therapy sessions via rewards and how this then helped to decrease his fear ratings of the buttons. This example would be awarded two marks.

Question 6

(a) Stronger responses could clearly outline positive reinforcement using an example from the study by Fagen *et al*. A majority of responses could identify that this is a type of reward and then give the example of how chopped bananas were used for this purpose in the study. However, the idea of increasing the probability of repeating behaviour was the component that was missing from most responses. Some candidates appeared to confuse operant and classical conditioning, whilst others tended to focus on primary and secondary reinforcers which was not answering the question.

(b) Candidates were much more likely to choose Roksana over Tien. For Roksana, popular choices included performance in the various tasks with explanations. Stronger responses could clearly provide evidence from the study to explain why Roksana was correct. For Tien, the most popular choice was the use of training methods to help the elephants pass the trunk wash test. As with Roksana, stronger responses could clearly provide evidence from the study. There was a significant minority of responses that confused individual explanations with situational explanations (or presented both in the response). It is important for candidates to know how each study fits into all issues and debates where appropriate.

Question 7

The average score on this question was around three marks. Stronger responses that were awarded five marks could clearly describe the Aggression Arousal procedure with specific examples and a logical progression to the response. However, a significant minority of responses focused on the final experimental room. It is very important for candidates to read each question carefully and ensure that their response is covering the demands of the question set rather than simply writing all they know about the procedure of a core study.

Question 8

Stronger responses focused on the situation Rosalia was in and gave appropriate advice to help reduce her 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 Rosalia to go and have a series of MRI scans to check her progress – this is both impractical, and the question was asking about reducing her stress levels and an MRI cannot measure that. Candidates need to ensure that they are giving meaningful advice.

Question 9

(a) The majority of responses could describe both victims clearly and concisely. Some candidates could outline one of the victims very well and then only identify the second. A small minority of responses focused on either the sample used in the study, or the models used.

(b) Stronger responses could clearly explain two differences. Popular choices to compare the studies on included the mandatory sample, type of experiment, validity and experimental design. 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 they **must** refer to the sample, yet some gave two other differences. In this case, only the best of these would be the mark awarded. 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 Baron-Cohen *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 quantitative data. 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 Baron-Cohen *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 Baron-Cohen *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 give Level 5.

Candidates need to know that any description of the study does not gain credit in these type of questions as it is testing their evaluation skills *only*. There were some common errors evident in responses, for example, focusing on the generalisability of the control groups which has nothing to do with the aim of the study so cannot be awarded any credit. 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 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 3 (five marks) for this cohort.

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, otherwise it can mean a candidate may run out of time for other questions.

Candidates should be able to give full definitions of terms 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.

General comments

The marks achieved by the candidates sitting this examination covered a wide spread of possible marks. Three-quarters of the candidates scored 18 marks or less. 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 a lot blank responses in this series (where every question 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 5, 6, 7, 9** and **10** had the strongest correlation with a candidate's overall score.

Comments on specific questions

Question 1

(a) A substantial majority of candidates could state the number of elephants. Common errors included stating the number of elephants that passed or providing a 'random' number.

(b) Only a small minority of candidates could identify **one** criterion used to select elephants. The most popular choices were lack of pregnancy and being docile. Common errors included describing what the elephants were being trained to do or other criteria never mentioned in the original paper.

(c) A small minority of responses focused on the question set. The most popular correct answer was that the end of the trunk needed to be placed on the palm of the trainer. However, many responses simply stated 'the elephant was asked to put its trunk here' which cannot gain any credit as it is copying the question. It is important for candidates to know the behavioural checklist aspects of studies that use observations as a research tool.

Question 2

(a) The majority of candidates could outline one place where the EEG electrodes were placed on participants. The most popular choice was on the scalp/head. Common errors included on the arms, near the heart or on the eyes. Some candidate focused on the electrodes being gathered to make a ponytail, but this was not answering the question.

(b) Responses tended to present what the weakness could be or an example from the study. Stronger responses could do both, focusing on the evaluative part of validity (usually about ecological) and then presenting a clear example from the study by Dement and Kleitman. It is important for candidates to read the question carefully to ensure they are evaluating and not just describing validity. Some candidates provided a weakness of the study but not in relation to validity, for example, generalisability of the sample.

Question 3

(a) The majority of responses could provide a partial aim. Popular choices included testing imitation of aggression and sex differences in imitation. Candidates need to be able to write full aims of studies to show they fully understand the purposes of each core study. Some responses focused on the findings of the study whilst others provided the aim already given in the question. It is important for candidates to read the question carefully.

(b) It is important for candidates to know the sample of every core study. Responses to this question rarely noted one correct feature even though the question required two. Popular correct choices included the sample size and sex distribution. Many responses focused on aspects of the study that were not about the sample, for example, watching the model or beating up the Bobo Doll.

(c) This type of question requires two parts to be awarded two marks. Firstly, we need to know what the application is and secondly, we need to be told how this would be achieved. Stronger responses covered both parts with popular choices being about censorship on aggressive behaviours in TV shows and video games and promoting prosocial behaviour in schools. There were some responses that attempted to explain everyday behaviours which is not the focus of this style of question. Responses similar to 'it helps us to understand aggression' are not answering the question.

Question 4

Candidates are required to know the results of studies including key quantitative results. A very small minority of responses could clearly compare two approaching figures in Experiment 1. The most popular choice was comparing the friend and stranger. However, a significant majority could not present a result from Experiment 1. Candidates tended to focus on high compared to low empathy or gave generic descriptions of the procedure. It is important for candidates to read the question carefully to ensure they are presenting the correct results. This had the joint highest rate of blanks on this paper.

Question 5

(a) Stronger responses could clearly outline one main assumption of the social approach. The most common choice was being influenced by actual or imagined people. Some candidates provided assumptions from different approaches or gave specific findings from a social approach study. Some responses stated that 'the social approach is about how we learn from social things', but this could not be given any credit. Assumptions need to be generic, rather than focused on specific mechanisms like obedience.

(b) Some responses could provide a key finding that linked to a social concept. However, only a minority of responses could then explain why it links back to the assumption they presented in **Question 5a**. Candidates tended to simply re-write the assumption they had already given in **Question 5a** with no attempt at explaining how it links back. Responses need to be explicit and clearly use examples from the study. A common correct response was explicitly stating that passengers were more likely to help the ill victim as it was an actual person requiring help as they did not know they were in a study. This example would be awarded two marks.

Question 6

(a) This type of question requires candidates to present generic psychological information linked to the named study. Stronger responses could clearly describe sex differences in relation to biology and what socialisation means. However, the majority of responses focused on what Hassett et al. did in their study which can only ever be awarded the one available example mark. To improve, candidates need to give generic descriptions of psychological concepts that Hassett et al. based their study on.

(b) Candidates were equally likely to choose Madison or Zach. For Madison, popular choices included internal validity and control of extraneous variables. Stronger responses could clearly provide evidence from the study to explain why Madison was correct. For Zach, the most popular choice was ecological validity. As with Madison, stronger responses could clearly provide evidence from the study. There was a significant minority of responses that confused validity with reliability, with a focus on standardisation and replicability which could not be given any credit as it is not answering the question.

Question 7

The average score on this question was around one mark. Stronger responses that were awarded five marks could clearly describe the shock generator. However, a significant majority of responses focused on how it was used as part of the procedure which was not answering the question. It is very important for candidates to read each question carefully and ensure that their response is covering the demands of the question set rather than simply writing all they know about the procedure of a core study.

Question 8

Stronger responses focused on the situation Alfonso was in and gave appropriate advice to help him improve his concentration at school. Popular advice included drawing then shading shapes on paper and not worrying about how neat it as. For these types of questions, the advice must be manageable and/or practical. There were several responses where the advice to Alfonso was to doodle. This is not practical as it does not tell Alfonso 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.

Question 9

(a) The majority of responses could not describe any of the questionnaire. For those responses that were awarded credit, it was mainly for the rating scale and its anchor points. Candidates were more likely to briefly describe how it was used in the study or create items that could measure mindfulness rather than answering the question set. It is important for candidates to read the question carefully. This had the joint highest rate of blanks on this paper.

(b) Stronger responses could clearly explain one similarity and one difference. Popular choices to compare the studies on included the sample, type of brain scans, validity and type of data collected. 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 or different species does not allow the response to be detailed so will 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 they **must not** refer to questionnaires, yet some focused on the use of questionnaires.

Question 10

The strongest responses evaluated the study by Saavedra and Silverman in depth and in terms of two strengths and two weaknesses with at least one of these points covering the named issue of subjectivity. Common choices included ethics, generalisability, longitudinal studies, 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 Saavedra and Silverman 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 Saavedra and Silverman 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 give Level 5. Candidates need to know that any description of the study does not gain credit in these type of questions as it is testing their evaluation skills *only*. There were some common errors evident in responses, for example, stating it was a laboratory experiment which cannot be awarded any credit. 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 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 1 (two marks) for this cohort.

PSYCHOLOGY

Paper 9990/21
Research Methods

Key messages

- Candidates need to ensure that they understand what is meant by key research methods terms and how to apply them to data given. This was especially the case with the standard deviation.
- Where candidates did know the meaning of the terms such as standard deviation, often they would just give a definition of that term rather than applying it to the data given. This highlights another key message which is that candidates need to ensure that their responses are focused on the questions within the exam paper. There was more than one question where candidates appeared to have 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 detail 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 link their answers to the information given in the stem if asked to do so. Often candidates showed excellent understanding of named issues/studies but then lost marks for giving generic responses.
- It is worth noting that candidate responses for the 6-mark question showed good knowledge and understanding of sampling techniques. Candidates just 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 such as saying random sampling is where you pick people randomly.
- Candidates need to ensure that they know the differences between human and animal ethics. Some candidates lost marks for suggesting ethical issues such as 'protection from harm' for animal studies, which is not creditworthy.
- It is worth noting, that candidate performance on the 10-mark extended response question asking candidates to describe an experiment was much improved from previous series. There were some thoughtful responses which clearly had good understanding of the requirements of an experiment but also were very much aware of replicability. For this, candidates should be commended.

General comments

This was the second June series for the new Psychology specification and significant improvements in many areas were seen. Paper 21 scripts provided the full range of marks, showing a good level of knowledge and understanding across many areas of the specification. Where performance was limited, it was due to a lack of knowledge of key terms, or a misunderstanding of the demands of the question. This was clear when looking at questions on the mean **(4a)**, standard deviation **(4b)**, and correlations **(8b, 8ci and ii)**. Candidate responses showed good knowledge and understanding of key studies, such as when referring to the Bandura et al. study and the Milgram study, although responses did show some misunderstanding of how to apply that knowledge to the question. Candidates did show their ability to structure 10-mark responses **(10a)** 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 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.

Comments on specific questions

Section A

Question 1

(a) This two-mark question required candidates to describe what is meant by a questionnaire. To achieve two marks candidates needed to give two **distinct** descriptive points related to the questionnaire research method. A full range of marks was seen for this question, with many candidates' showing good knowledge and understanding.

Where candidate responses did not achieve the two marks it was due to a misunderstanding about the demand of the question. For example, writing about closed and open questions or structured and unstructured questionnaires would achieve one mark for the 'questions' part; Saying that it could be written or online could only achieve one mark for the way a questionnaire can be completed. However, most candidates got at least one mark, and many got the two marks for simply saying 'written questions'.

(b) This two-mark question asked candidates to explain one practical strength of using a questionnaire. To achieve the two marks candidate responses needed to:

- identify/explain one strength
- give some detail/justify/give a comparison to another method.

Candidates struggled slightly with this question. Many candidate responses either went down the line of ethical strengths or just suggesting it is quick and easy which was not creditworthy. Responses would often make a vague point about being able to achieve a wider sample but not actually say why this is the case. Where candidate responses did explain one strength they often did not go on to justify/elaborate on this strength. Comparisons to other methods were very rarely seen in responses.

Where candidates did achieve marks, the most common strength was the idea of lower risk of social desirability due to them being completed without the researcher present. Where candidate responses did achieve marks for the wider sample it was usually because they said that the questionnaire could be accessed online. Where candidate responses did this, they often achieved the second mark for the ideas of representativeness/generalisability.

Candidates often find these types of questions more of a challenge and therefore teachers could use past papers and unseen scenarios to allow candidates to practise this type of response

Question 2

(a) This one-mark question asked candidates to outline what is meant by consent. There were no major issues with this question with most candidates able to show accurate understanding of the term. Where candidates did not achieve the one mark available it was due to responses talking about consent in a non-psychological way. For example, if they stated that it was when people agreed to take part it was not credited as this is not linked to a study however IF the response mentioned the idea of being a participant or part of a study then it could gain credit. Many candidates went further than was needed for this question and outlined informed consent. This was credited but candidates should ensure they are not using up more time expanding their answers than is necessary.

(b) This two-mark question asked candidates to suggest why obtaining consent would have been important in this study (Bandura et al.). To achieve two marks for this question candidate responses needed to:

- provide a linked suggestion why consent was needed
- EITHER provide elaboration on the above suggestion OR give an entirely separate second one.

Candidates struggled with this question at times, mainly because many responses just talked about generic issues around using children in a study and did not relate it back to the issues of using children in a study about **aggression**.

Where at least one mark was achieved it was usually due to candidates suggesting that there could be possible harm due to them seeing aggression (this is fine as a linked ethical issue). Stronger candidate responses would then go on to suggest that this was due to them being scared or frightened by seeing aggression or that they could actually be hurt by the aggression being shown. Often, however, candidate responses achieved zero or one mark

Question 3

(a) This two-mark question asked candidates to suggest one reason why using participants with different occupations could be important to Milgram's test of obedience, using an example in their answer. To achieve both marks on this question candidate responses needed to:

- make a suggestion linked to obedience
- provide an example to exemplify this point.

Some thoughtful candidate responses were seen for this question, however candidate performance on this question was mixed.

The most common responses linked to certain occupations needing to show more or less obedience within their profession. The examples then were often related to people in the military having to obey and that managers did not obey as much as the workers that they manage. Where performance was more limited it was due to candidate responses talking about generalisability, which was in the question so could not be credited, or just giving generic answers which did not link back at all to obedience. It is important therefore that candidates always ensure that their response is linked explicitly to the question being asked.

(b) This two-mark question asked candidates to suggest one reason why using participants with different ages could be important to Milgram's test of obedience. To achieve both marks on this question candidate responses needed to:

- make a suggestion linked to obedience
- provide an example to exemplify this point.

How well candidates performed on this question was often linked to their performance on **3(a)**, as this was similar in style and demand to that question, only linked to age rather than occupation. In the main, however, this question performed slightly better than **3(a)**. More candidate responses were able to give the suggestion that some age groups may be more obedient than others and then go on to either suggest older people would be more obedient as they respect authority more or conversely that younger people will be more obedient as they will not have as much knowledge about the world and authority as older people. Both interpretations were creditworthy. Very few candidates suggested that children have to obey parents more so they will naturally be more obedient, however there were a few more that talked about children having to obey teachers! The most common error, as in **part 3(a)** above, was responses that focused on generalisability, or did not mention obedience at all, but these were far fewer than **part 3(a)** above.

Question 4

(a) This one-mark question asked candidates to state what the means show about the ages of the two groups (from a given table of data). To achieve the mark responses needed some reference to one group (1) being **older** than the other group.

Candidate performance on this question was below what was expected. The main issue came with candidate responses which just defined what the mean is or just repeated the results from the table. Candidates were given the means, so just copying out the table was not creditworthy.

The other common error on responses came with candidate responses trying to be more specific but slightly misunderstanding the data by saying group 3 are mainly children which was not the case.

This question shows the importance of candidates reading the question properly and making sure they understand what the question is asking for.

(b) This one-mark question asked candidates to state what the standard deviations show about the ages of the two groups (from a given table of data). To achieve the mark candidate responses needed to show understanding that group 3 is more 'spread out' in terms of ages than group 1. A greater range of ages is also fine.

Candidate performance on this question was below what was expected, although this question performed slightly better than **4(a)** above. The main issue came with candidate responses which just defined what a standard deviation is or just repeated the results from the table.

Question 5

This two-mark question asked candidates to explain **one** ethical weakness of the case study research method. Candidates were told not to use the idea of consent in their response. To achieve the two marks candidate responses needed to:

- identify an ethical weakness
- elaborate/justify this reason.

Candidate performance on this question was pleasing with most candidates able to get at least one mark, with many achieving both marks. Acceptable ethical issues for this question were:

- Confidentiality
- Distress/harm
- Right to withdraw
- Privacy

Deception was not creditworthy in this case. If candidates explained the issue above without specifically naming it, they would also gain the identification mark.

Where performance was more limited it was due to not elaborating on their identified weakness enough to achieve the second mark. For example, right to withdraw is mainly linked to the relationship built up with the researcher but many suggested they could not withdraw as they were the only participant which of course is not really true. The most common response was linked to confidentiality (only person so more likely to be identifiable) or possible harm (due to the intensity of the research).

Some candidates would link their ethical weakness to a specific case study; especially in relation to distress i.e. Saavedra and Silverman, and this was absolutely fine as long as their elaboration was linked to the ethical weakness that they had stated.

Question 6

This six-mark question required candidates to describe opportunity sampling and random sampling, using any example(s). Awarded marks could come from accurate description of each term such as:

- opportunity is whoever is available at the time
- random is where everyone has an equal chance of being selected.

and appropriate examples such as:

- Piliavin et al. using opportunity sampling as they used whoever was on the subway.
- Baron-Cohen et al. used random sampling to get participants in group 4 who were age and IQ matched to group 1.
- Their own study examples which clearly showed knowledge of the different methods.

It is also worth noting that for random sampling candidates could get up to two marks for the PROCESS of random sampling e.g. getting a numbered list of the target population (1) putting these into a hat and drawing out the number needed (1) – many candidates did get one or two marks for this.

Candidates found this question fairly challenging and, although they were able to achieve marks for opportunity sampling in particular, there were some misunderstandings in relation to random sampling.

At the lower end of the mark range, candidates usually achieved one or two marks for opportunity sampling. They may be able to give an appropriate example, such as the use of subway passengers for Piliavin et al. At this mark range misunderstandings about random sampling were present, for example, random sampling is where you pick random people off the street; or opportunity sampling being when you put posters up.

At the high end of the mark range candidates produced some thoughtful responses and were able to give a detailed description of the process of random sampling, as well as knowing it presents an equal chance for everyone in the target population to be selected. Opportunity sampling was still slightly stronger at this level with relevant examples and detail given. Candidate responses at this level also gained a mark for defining what is meant by a 'sampling technique'.

It is also worth noting that responses needed an explicit example in order to access all six marks. There were more instances than in previous series where responses showed fantastic knowledge about the sampling methods but did not give any examples and therefore could only achieve four marks. In terms of candidates own examples, candidates need to be aware that these examples need to link to a research question so just saying 'where you get people playing football at the park you are at' is not enough as this is not linked and is just a generic statement.

Question 7

(a) (i) This two- mark question asked candidates to suggest one way that Dr Smith could produce a quantitative measure of how easily the animals learn. To achieve two marks candidate responses needed to:

- identify one way
- give some detail about that way.

Candidates were often able to get at least one of the marks for this question with many using the idea of the time taken, or the number of times a behaviour happened, but were not able to achieve the second mark. This second mark could be achieved in one of two ways:

1. They could say where the animal would learn their behaviour e.g. in a Skinner box.
2. Stating what this measure could say about learning e.g. the quicker the time that they learn to press the lever the quicker they have learnt.

The second point was more common than the first, but it was not seen very often, and therefore often candidate responses were limited to one mark.

(ii) This two-mark question asked candidates to explain whether the way identified in **part 7(a)(i)** would be valid for both species (a smaller and larger one). To achieve two marks on this question candidate responses needed to:

- give a reason why it may or may not be valid
- provide some elaboration on that reason.

Candidate performance on this question was directly related to their success on **part 7(a)(i)**. If candidates had achieved no marks on part 7(**a)(i)** they inevitably achieved lower marks on **part (a)(ii)** (although this was not always the case). A full range of marks was seen for this question.

By far the most common response which achieved marks on this question was the idea that it was valid because the task would be equally as easy for both species. This focus on the cognitive element of the task often meant that candidates achieved marks.

The idea that it is not valid produced some very creative and thoughtful responses such as the bigger animal would need more food so would press the lever more, that they would have too big paws and may not be able to press it properly, that the bigger animal could cover the ground more easily. This was really pleasing to see and showed a level of understanding that was well worth the marks.

(b) This three-mark question asked candidates to draw the axes for a graph that could be used to show the results of Dr Smith's study using the measure of learning suggested in **part (a)(i)**. This question caused very few issues for candidates with the vast majority getting most if not all of the three marks.

Where candidates did not achieve full marks it was due to both axes not being labelled correctly. For example, some candidate responses did not use the measurement stated in **part (a)(ii)** or just gave a vague label such as 'time'. In addition, some candidates did label the axes but then did not provide any units for either label (either species 1 and 2 or minutes for time, for example). It is worth reminding candidates to ensure that they give units for axes whenever drawing a graph.

(c) This two-mark question asked candidates to suggest how Dr Smith could follow **one** ethical guideline in relation to animals, other than 'species. To achieve the two marks candidates needed to:

- identify an animal guideline
- provide a link to the study.

Candidates performed well on this question with the majority showing they understood the difference between an animal and human ethical issue. Issues which could be credited were:

- Distress
- Housing
- Deprivation

The most common response for this question was related to housing. Candidate responses would then elaborate successfully suggesting that the housing needed to be with other social animals/enough space.

Where candidate responses performed less well it was due to the lack of elaboration from their identified ethical issue, or their elaboration not relating to the ethical issue stated. For example, if candidate responses had mentioned housing, some would then go on to talk about ensuring the animal is not hungry which is a different issue. The other error seen was the use of a human ethical issue such as protection from harm, or that animals cannot give consent. It is worth reminding candidates that human and animal ethical issues are not the same, and to be aware of the differences.

Question 8

(a) (i) This two-mark question asked candidates to explain one problem about validity or reliability that Josie could have with one of her measures of tidiness. To achieve the two marks on this question candidate responses needed to:

- identify a problem linked to the scenario
- state whether this is a problem with validity or reliability.

Candidates performed fairly well on this question. Most were able to state a problem with either one of the measures with creditworthy responses such as:

- Josie being subjective in what she believes is tiredness.
- Josie not understanding the system the participant is using.
- Josie's rating of tiredness being different to what other researchers would give.
- The idea of social desirability with the candidates rating themselves.
- The idea of candidates interpreting the numbers on the scale differently.

The vast majority of candidate responses who suggested one of the above were then able to correctly link it to validity/reliability. Often candidate responses would specify a certain type of validity or reliability, but this was not necessary for the mark.

Where performance was limited, it was due to candidate responses choosing validity/reliability incorrectly e.g. suggesting that subjectivity would be reliability, or the different interpretation of numbers was validity.

(ii) This two-mark question asked candidates to explain one ethical problem that Josie could have with one of her measures of tidiness. To achieve the two marks candidate response needed to:

- identify a problem linked to the scenario
- link this to the correct ethical guideline.

Candidate performance on this question was pleasing, with the majority of candidates achieving the two marks available. The most common response was the idea of being told you are untidy being an embarrassment and then linking this to protection from harm/distress. If candidate responses did talk about distress/harm it was likely they would achieve the two marks.

The other most common response was the idea of privacy, because she was looking at their work which may contain private details. Candidate responses were not quite as strong for this issue with many just stating 'because they would not want her to look at their work' and not linking it to a guideline. A few candidate responses did not link back to the scenario so were unable to get two marks, but this was rarely seen.

(b) This two-mark question asked candidates to explain which one of the measures of tidiness in **part (a)** is not suitable for Josie to use in her correlation. To achieve the two marks on this question candidates needed to:

- provide an explanation **why** one of the measures is not suitable
- link this explanation to the scenario given.

It is worth re-iterating that for an EXPLAIN question candidates will not achieve marks for just choosing a measure, they needed to explain WHY It has been chosen.

The majority of candidates did not achieve any marks on this question. Often candidate responses would discuss the idea of subjectivity/bias meaning you cannot use it – this was not creditworthy.

Where candidate responses did achieve marks, it was due to the understanding that correlations would need continuous date such as a rating scale and that by giving participants only a yes/no choice it was categorical/nominal. However, where candidates did achieve this mark, they at times forgot to link it back to the scenario. For example, they would say 'because it was only categorical data and you need a continuous scale' – which is absolutely right but tell us nothing about the scenario. Candidate responses only needed to add the yes/no or the word tidiness to get the second mark here.

It is worth reminding candidates that whenever a question talks about **this** study etc. there will **always** need to be a link to the study within their response.

(c) (i) This two-mark question asked candidates to explain what Josie will conclude about her correlation. To achieve the two marks candidates needed to give a linked explanation about what Josie could conclude. In essence they need to identify it was a positive correlation due to both tidiness and enjoyment increasing together.

Most candidate responses were able to achieve at least one mark for identifying that it was a positive correlation (this part did not have to be linked) but fewer were able to achieve the second mark. This was often due to candidate responses such as:

- 'Positive correlation with tidiness and wellbeing' – this was not enough for the second mark as candidate responses just repeated the two variables.

'Positive correlation as tidiness and wellbeing both increased/went up together' was enough to achieve the two marks.

Where performance was limited, it was either due to a lack of a link to the scenario, or the identification of the wrong type of correlation. The other error seen in some responses was candidates suggesting a causal relationship such as 'Tidy work **causes** people to enjoy their lessons' which is incorrect.

(ii) This two-mark question asked candidates to explain whether Josie can conclude that having tidier classwork influences candidates' enjoyment of lessons. To achieve the two marks candidates needed to:

- give an explanation as to whether Josie can conclude that having tidier classwork influences candidates' enjoyment of lessons
- provide some elaboration on this explanation.

Candidate performance on this question was below expectations, with zero being the most common mark. Some candidates seemed to get confused and believed that you can conclude that having tidier classwork influences candidate enjoyment. This suggests an idea of causation which, being a correlation, you cannot do.

Some candidates did achieve marks by saying that you cannot show cause and effect with a correlation, and if they did so candidates often went on to suggest that there could be a third variable affecting results, or that it could be either variable that is the cause.

Question 9

(a) (i) This one-mark question asked candidates to explain which of these intervals (yearly or weekly) is the most suitable for Dr Gao's study. To achieve the mark available for this question candidate responses needed to do more than just state yearly or weekly which many candidates did. Others tried to explain it should be weekly, but this would not be the most suitable in these circumstances due to the length of the study. Where candidate responses did provide an explanation for yearly, they often got the mark through suggesting that reading habits do not change weekly. This is all that they needed to get the mark, but only the minority of candidates achieved the mark on this question.

(ii) This one-mark question asked candidates to suggest how Dr Gao could contact her participants during the study. This was a very straightforward question for all candidates who only had to state any appropriate form of communication such as phone/test/email/through WhatsApp etc. This did not present any real problems for candidates and inevitably they achieved the mark available.

(iii) This two-mark question asked candidates to suggest two problems that Dr Gao may have when trying to contact her participants at regular intervals. To achieve the two marks for this question candidate responses needed to give two distinct issues with trying to communicate using the sort of means they have mentioned i.e. phone/email etc.

Candidate performance on this question was as expected, with most candidates able to give two problems with the most common ones being:

- Change of number/email.
- Being away/at work when the call comes.
- Not wanting to take part any more so does not answer.
- Has become too ill to participate.

The most common error was misunderstanding the demand of the question and talking about issues such as social desirability/exaggeration and order effects which were not creditworthy.

(b) This two-mark question asked candidates to suggest one advantage of Dr Robert's idea (for a snapshot study) compared to Dr Gao's plan for a longitudinal study (without using the idea of contacting participants). To achieve the two marks for this question candidates needed to:

- suggest an appropriate advantage
- link this advantage back to the scenario.

Candidate performance on this question was below expectation with few candidate responses able to achieve the two marks available. Many candidate responses talked about snapshot studies being shorter in terms of time and cost which on its own is not always true as snapshot studies may get results sooner but may be a lot more time intensive and costly for the period of time of the study. The only way time could be credited as an advantage was linking it to being able to publish the study results sooner but candidate responses did not do this.

Where candidate responses did get marks, it was usually due to the idea of subject attrition being less likely in a snapshot than longitudinal study. However, candidates often did not link this response to the study and therefore were limited to one of the two marks.

Question 10

(a) This ten-mark essay question asked candidates to describe how Dr Miller could conduct an experiment using an independent measures design to investigate whether creative activities help people to sleep well. 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 level (5) but there were a number of level 3/4 responses.

Many candidates had a sound understanding of experiments and were able to make relevant decisions about the dependent variables and controls. Where candidates struggled slightly was with explicitly showing what independent measures design meant for their study. Often the design was implicit within the response rather than explained, which often meant that candidates were unable to achieve the higher levels. In addition, often candidate responses would not operationalise their IVs just stating the type of creative activity which was given in the scenario itself.

Candidate responses within the lower mark range were often able to give an appropriate way to measure the dependent variable, e.g. with an EEG or a questionnaire, and were able to suggest some form of controls. However, at this level, candidate responses would have significant gaps within the procedure which would mean that it would be difficult to replicate. Candidate responses may also use other research methods which could not be linked to the experimental method or would suggest that all participants would do all levels of the IV which is repeated rather than independent measures. In this case, the response could still achieve marks for their DVs and controls, 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.

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 how to operationalise the creative activities that they were using e.g. playing the piano for 30 minutes or listening to opera for 30 minutes. They would provide a detailed way of measuring the DV showing knowledge of what an EEG would show or giving examples of questions which could be asked to participants about their quality of sleep. At this range candidates would go into detail about types of controls needed e.g. same length of time doing the activity; no caffeine or alcohol, same room to sleep in etc. This element of the responses was probably the strongest. What differentiated candidate responses at this level was whether they explicitly talked about independent measures as a research design choice. Some candidates would talk about random allocation, and about the fact that participants would only be part of one level of the IV. If candidates did so, they would be given a detail mark. 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) (ii) These two, two-mark questions asked candidates to explain one advantage/disadvantage of using an independent measures design in this study compared to using a repeated measures design. To achieve the two marks candidates needed for each part of the question to:

- give an advantage/disadvantage of independent measures
- link this to the scenario/their study.

Most candidates could give an appropriate advantage of independent measures, such as a lack of order effects, and an appropriate disadvantage, such as the idea of participant variables. The most common error in both questions was not linking this back to their study/the scenario, meaning they were only able to access one of the two marks.

PSYCHOLOGY

Paper 9990/22
Research Methods

Key messages

- Candidates need to ensure that responses are focused on the specific questions on the exam paper. There were instances where it was clear candidates misunderstood questions which impacted their marks.
- 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/outline key terms within the specification such as opportunity sampling, participant and situational variables, and order effects.
- 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.
- It is worth noting that candidate responses for the higher (6 and 10 mark) tariff questions showed good knowledge and understanding of order effects and correlations. There were a number of thoughtful responses for the extended essay question, and candidates should be commended for their performance on these questions.

General comments

This was the second summer 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 participant and situational variables (**7(a)(i)** and **7**(**b)(i)**), social desirability (**7**(**c)(i)**) and when asked to use specific measures of central tendency (**8(a)(i)**/**8(a)(ii)**). Candidate responses also showed some gaps in knowledge when referring to the Milgram study (obedience), with responses showing some misunderstanding in relation to the findings of the study i.e. voltages, participant comments.

Candidates showed good ability when structuring six- and 10-mark responses (**6** and **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 exam in general and 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

This two-mark question required responses to explain how the independent variable (IV) was operationalised in the study by Andrade (doodling). To achieve both marks candidate responses needed to:

- Identify the independent variable (whether the participants doodled or not).
- Provide some detail about how this was operationalised (given paper with shapes on to doodle/just given a sheet of paper).

A full range of marks was seen for this question, with many candidates' showing good knowledge and understanding of the Andrade study. Where candidate responses did not achieve the two marks it was often due to a lack of operationalisation of the IV, or suggesting the amount of information remembered was the independent variable (it was the dependent variable).

Question 2

There was an error in question 2, the word 'monkeys' should not have been included at the end of the stem, as the graph the question referred to in the Hassett et al. study was about male and female humans. This has been corrected in the published version of the paper.

Due to the issue with question 2, full marks have been awarded to all candidates for this question to make sure no candidates were disadvantaged.

Question 3

(a) (i) This one-mark question required candidates to (in relation to the voltage level in the Milgram study) identify the type of data. The definitive answer to this question was quantitative. No explanation of 'why' was needed. Most candidates achieved the one mark available for this question.

(ii) This one-mark question asked candidates to (in relation to the voltage level in the Milgram study) state the highest voltage at which all participants obeyed. The definitive answer for this question was 300v.

The vast majority of candidate responses achieved the one mark available for this question. Where candidate responses did not achieve the mark, it was usually due to suggesting the maximum was 450v. This was not creditworthy as although it was the maximum voltage on the 'shock' machine not all participants went on to that level.

(b) (i) This one-mark question asked candidates to (in relation to the comments by participants) identify this type of data in Milgram's study. The definitive answer for this question was qualitative and most candidates achieved the one mark available. Where candidate responses did not achieve the marks, it was mainly due to getting the type of data the wrong way round for both **part (a)(i)** and **(b)(i)**.

(ii) This one-mark question asked candidates to outline one of the comments made by a participant in the Milgram study. There were a number of different comments which were creditworthy for this question and most candidates were able to quote one of these comments.

It is worth noting that the responses did not have to be exactly the same as the ones in the journal article as long as the same meaning could be inferred. In addition, comments from during the study as well as the interviews after the study were creditworthy.

(c) This four-mark question asked candidates to suggest one strength for each type of data (qualitative and quantitative) in the Milgram study. To achieve the two marks for each type of data candidate responses needed to:

- Give one strength of qualitative/quantitative data.
- Link this stated strength to the Milgram study.

Candidate performance on this question was mixed. The majority of candidates were able to identify a strength for each type of data but often did not link this strength to the Milgram study. The most common responses for each type of data were:

- Quantitative: easy to compare/analyse.
- Qualitative: provides detail/feelings behind an action.

Where candidates did achieve full marks, they would link their strengths, such as the ability to compare voltage levels between participants; or showing the strain the participants were going

through when they 'shocked' the learners. This was not the majority of candidate responses and therefore the most common mark for this question was two marks.

Question 4

This one-mark question required candidates to suggest how 'learning' could be operationalised in an animal study about dropping small objects in a hole. To achieve the mark candidate responses needed to provide an operational definition for the learning above. Candidate performance on this question was as expected with many candidates able to provide an appropriate definition such as:

- how many go into the hole
- how fast they can get a certain number of objects in the hole.

It is worth noting that a definition of learning was creditworthy on this occasion. Where candidate performance was limited, it was either due to responses which were too vague, not measurable, or responses that would explain **how** they get taught i.e. through the use of positive reinforcement, which was not creditworthy.

Question 5

This four- mark question asked candidates to suggest **two** weaknesses of using scientific equipment to measure variables. To achieve the four marks on this question candidates were required to:

- Provide a weakness of scientific equipment (x2).
- Provide an elaboration/some detail about the specific weakness (x2).

A variety of responses was seen, by far the most common weaknesses were:

- The idea of low 'ecological validity' due to people sleeping with electrodes on their heads/performing tasks in a scanning machine.
- The possibility of harm due to the use of scanners with enclosed spaces or that some people can be harmed by certain types of scanner. Often with this response candidates did not give enough detail to achieve the second mark.

Where candidate performance was more limited it was due to responses giving generic answers about cost. Cost could be creditworthy if clarified in terms of the budget of the study being high/that they would therefore only be able to afford a smaller sample size, however there were very few responses that gave that sort of detail.

It is also worth noting that if candidate responses linked their weakness to a study i.e. the idea of low ecological validity linked to Dement and Kleitman. This was fine as long as it was very clear what the study itself would show in terms of the weakness identified.

Question 6

This six-mark question asked candidates to describe order effects, using any example(s). Awarded marks could come from specific/accurate features of order effects such as:

- Repeating a task more than once.
- Used in a repeated measures design.
- The order a specific task is done (in a lab experiment for example).
- Can be minimised through counterbalancing.

Specific examples for any of the above such as:

- Pozzulo counterbalanced/randomised the order to counteract order effects.

Specific examples such as practice effects (getting better due to repeating a task) or fatigue effects (getting worse due to becoming bored/tired) were also creditworthy. Other study examples which were creditworthy included:

- Dement and Kleitman randomized the waking up in NREM/NREM so less chance of order effects.

- Andrade used counterbalancing of names/places sets.
- Perry et al. was repeated measures in both Experiment 1 and 2.
- Holzel et al. possible order effects on FFMQ (done before and after).
- Pozzulo et al. randomised position of target in photos/video clips so less chance of order effects.

Candidate responses to this question were pleasing. At the lower end of the mark range, candidates usually achieved one or two marks for suggesting that order effects are usual with a repeated measures design and giving an example, however responses did not give the detail required to achieve the higher marks or describe the specific types of order effects such as practice or fatigue.

At the higher end of the mark range candidates produced some thoughtful responses which looked at a variety of features of order effects, and relevant studies linked to such techniques as randomisation/counterbalancing. It is worth noting that examples do not have to be from named studies, and there were some creative responses which used a variety of examples, which are indeed creditworthy.

Question 7

(a) (i) This one-mark question asked candidates to explain which **one** of two uncontrolled variables is a participant variable. To achieve the mark for this question candidate responses needed to do more than identify the participant variable; they also had to explain **why** it is a participant variable. As an explain question rather than a state or name this extra bit of detail is needed.

Candidate performance on this question was mixed. Some candidate responses produced nice explanations about it being something to do with the individual / a specific characteristic of the person. Where performance was more limited, it was either due to:

- candidate responses which only named the specific variable
- candidate responses that just repeated the word participant in their explanation
- candidate responses that used a reverse argument i.e. not to do with the environment.

For future series it is important that candidates know the different requirements for command words such as name/state/explain.

(ii) This two-mark question asked candidates to suggest how Inma could limit the effect of the participant variable. To achieve the two marks for this question candidates needed to:

- suggest a way the issue (the headache) can be solved or reduced
- provide some detail for this suggestion OR provide a second distinct way of reducing the effect of the variable.

Candidate performance on this question was pleasing with many creative responses being seen. Creditworthy responses included:

- Ask all participants if they are feeling well/have a headache (suggestion) and if they do remove them from the sample (detail).
- Tell the participant to take a tablet which will help stop their head from hurting, and then observe them once they feel better (detail).
- Observe them over a variety of different periods of time (suggestion) which would mean the results could be averaged so the effects of the headache would not affect results (detail).
- Ask the participants after the study finishes whether they feel well (suggestion) if they do not then remove their results from the study (detail).

All of the above responses were seen from many candidate responses showing the thoughtful and creative nature of their responses. Where performance was limited, it was usually due to candidate responses which misinterpreted the demands of the question and talked about how the variable would affect the results rather than how it could be reduced. It is also worth noting that if candidate responses had chosen the wrong variable in **part (a)(i)** then it was inevitable that they would achieve very few marks on this question.

(b) (i) This one-mark question asked candidates to explain which **one** of two uncontrolled variables is a situational variable. To achieve the mark for this question candidate responses needed to do more

than identify the situational variable; but also explain **why** it is a situational variable. As an explain question rather than a state or name this extra bit of detail is needed.

Candidate performance on this question was mixed. Some candidate responses produced really nice explanations about it being something to do with the environment/something external. Similar to **part (a)(i)** where performance was more limited, it was either due to:

- candidate responses which only named the specific variable
- candidate responses that just repeated the word situation in their explanation
- candidate responses that used a reverse argument i.e. not to do with the person.

For future series it is important that candidates know the different requirements for command words such as name/state/explain

(ii) This two –mark question asked candidates to suggest how Inma could limit the effect of the situational variable. To achieve the two marks for this question candidate responses needed to:

- suggest a way the issue can be solved or reduced
- provide some detail for this suggestion OR provide a second distinct way of reducing the effect of the variable.

Candidate performance on this question was mixed, with far fewer candidate responses achieving the two marks available. The important part of this question is the fact that very few people are there in the environment (so less likely you will talk to anyone) and that you need more people there to allow more interactions. Unfortunately, many candidates interpreted the question differently, thinking that it was the variation in numbers that was the issue and suggested standardising numbers which was not creditworthy (as this could mean that numbers were still too small). Creditworthy responses included:

- Ensure you observe when there are lots of people around in the playground (suggestion), for example when it's close to school start time when everyone is there.
- Go and observe where you know there will be a lot of people (suggestion) for example, you could observe in the canteen at lunch time/corridor before lessons (detail).

Candidate responses which used the above possibilities were in the minority.

(c) (i) This two-mark question asked candidates to suggest one way that social desirability could affect Inma's results. To achieve the two marks for this question candidate responses needed to:

- identify a possible reason why social desirability may occur
- provide some detail about how that may affect the results of the study.

Candidate responses showed that they know social desirability well and often they could achieve the identification mark by talking about the participants wanting to look good in front of the researcher and therefore changing their behaviour. Where performance was limited, it was due to candidate responses not going on to suggest how that would affect the results of Inma's study, such as only suggesting it would 'change their behaviour in the playground' which is not creditworthy as this does not specify in terms of the results of the study (talking).

There were some really thoughtful candidate responses that did achieve two marks who talked about how a person may not want to be seen as shy/too talkative and therefore would change the amount that they talk to people to make themselves look better. These were excellent responses which fully deserved the two marks available for this question.

For future exam series it is worth reminding candidates that they need to answer all parts of the question and not focus just on the key words such as social desirability.

(ii) This two-mark question asked candidates to suggest how Inma might reduce the effect of social desirability on her results. To achieve the two marks for this question candidate responses needed to:

- identify a way to reduce social desirability

- provide detail about this way linked to the scenario OR provide a second way to reduce social desirability distinct from the first one above.

Candidate performance was directly linked to their understanding of the term social desirability in **part 7(c)(i)**. Candidate performance on this question was mixed with 1 mark out of 2 being the most common mark. Candidate responses were able to provide a way to reduce social desirability such as performing a covert observation or deceiving participants about the aim which is creditworthy and achieved the 1 mark. What fewer candidate responses did is then go on to give some **linked** detail. This is especially true when they are talking about covert observation. Many responses would say:

"Do a covert observation, so people do not know they're observing"

This response would only get 1 for 2 reasons. Firstly, it is not linked, it is not detailed/another brief way of reducing social desirability. To achieve the second mark for this example candidate responses could go on and suggest that the researcher should pretend to be a teacher in the playground, or they could put up cameras which overlook the playground which she could then watch.

In terms of deceiving about the aim, the detail mark was often not acquired. Candidate responses needed to provide some detail about what they would tell the participants the study was about such as how weather affects play.

Question 8

(a) (i) This one-mark question asked candidates to state which measure of central tendency Kong should use to analyse his data from timing the duration of eye contact. The definitive answer for this question is the mean. Most candidates achieved the one mark available for this question.

(ii) This one-mark question asked candidates to state which measure of central tendency Kong should use to analyse his data from asking participants 'Are you excited or scared at parties'. The definitive answer for this question was the mode. Most candidates achieved the one mark available for this question.

(b) (i) This two-mark question asked candidates to explain one reason why timing the duration of eye contact could break one ethical guideline. To achieve the two marks candidate responses needed to:

- Identify an appropriate ethical guideline the timing may have broken.
- Provide some detail about that guideline in relation to the study.

Candidate performance on this question was pleasing with many achieving the two marks available. The ethical issues which were creditworthy included:

- Consent
- Deception
- Protection from harm
- Right to withdraw

The most common response discussed the potential for harm due to Kong's prolonged eye contact. Although the question did not say that Kong is deliberately trying to maintain eye contact, it is a perfectly reasonable point given what candidates know about researcher bias, and therefore this was a very thoughtful response based upon knowledge and therefore was seen as creditworthy.

It is worth noting that although creditworthy, most responses that used right to withdraw only achieved one mark as responses found it difficult to give the detail needed for the second mark. For the linked detail they really need to say about the fact they would be happy with a study on free time but not friendliness (as more personal), but they have not been told so cannot withdraw.

(ii) This two-mark question asked candidates to explain one reason why asking participants 'Are you excited or scared at parties?' could break one ethical guideline (a different one than used for **8(b)(i)**). To achieve the two marks for this question candidate responses needed to:

- Identify an appropriate ethical guideline **different** to **part (b)(i)**.
- Provide some detail about the guideline in relation to the study.

Candidate performance on this question was pleasing with many candidate responses achieving the two marks available for this question. Where performance was more limited it was due to either a lack of detail for the second mark OR candidates using the same ethical issue as they used in **part (b)(i)**.

By far the most common response here was linked to the ethical guideline of privacy with many responses providing detail such as it being a personal question about their partying habits, although many responses also talked about the possibility of harm due to the participant being embarrassed at the thought of saying they were scared at parties. This was less common mainly since a lot of candidates had used possible harm in their response to **part (b)(i).**

For future series it is worth reminding candidates to read the whole of the question prior to starting to answer it, as some candidates seemed to miss the part about using a different ethical issue than the previous question.

(c) This two-mark question asked candidates to outline what is meant by 'opportunity sampling', using Kong's study as an example. To achieve both marks for this question candidate responses needed to:

- give an outline of opportunity sampling
- give an example of this sampling method using Kong's study.

Most candidates were able to outline opportunity sampling as when you get participants who are available at the time with many then going on to suggest that Kong could go to a local park/playground and ask whoever is there. Where candidate performance was limited, it was either due to:

- Completely generic response such as 'opportunity sampling is where you can get people who are available in the local park' as there is no reference to the study i.e. to friendliness/Kong as the researcher.
- Candidate responses talking about the wrong sampling method, which was mainly volunteer and suggesting you could put posters up.

This was the minority of candidate responses, and many got the full marks available.

Question 9

(a) This two-mark question asked candidates to identify the two techniques for presenting a questionnaire to participants. The definitive answers here were paper and pen AND online. It is worth noting that physical copy/written were creditworthy for paper and pen, and digital/through email creditworthy for online. Most candidate responses achieved the two marks available for this question. Where performance was limited, it was due to responses suggesting TYPES of questions such as open/closed, or questionnaires such as structured/unstructured.

(b) This two-mark question asked candidates to suggest one reason why one of the techniques identified in **part (a)** would be appropriate for Dr Shaw to use. To achieve both marks for this question candidates needed to:

- suggest one reason
- give linked detail about the suggestion.

Candidate performance on this question was mixed. Candidate responses which did not get **part 9(a)** right usually did not achieve the marks on this question. Where candidate responses did achieve marks, it was usually one out of two. Many candidate responses chose online as it is easier to get a larger and more diverse sample but then forgot to link back to the study and so did not get any further credit.

Where candidate responses did get the two marks it was usually by choosing paper and pencil and suggesting that it may be more private especially given that childhood memories could be a sensitive topic area. Of course, candidate responses could also suggest this for online as well because it could be anonymised and if done well this could also be credited.

Question 10

(a) This ten-mark essay question asked candidates to describe how Dr Brent could conduct a study to investigate whether there is a correlation between how noisy wards are and the quality of patients' sleep. Candidate responses showed the full range of marks, with some really thoughtful responses 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 correlations and were able to make relevant decisions about how to measure the two co-variables. The part which candidates found more challenging was the operationalisation of the two variables. Many would just repeat noise and sleep; and although it could get minimal credit for a) it could not get a mark for detail. To get the detail mark here candidate responses needed to first show that they knew that these were co-variables and not an independent and dependent variable, and then show some level of operationalisation such as **level** of noise, and **quality/length** of people's sleep. This is something that many candidate responses did not do.

Candidate responses within the lower mark ranges were often able to show knowledge of what the variables were but did not operationalise them. They were able to show how they would measure these two variables, but often the details were fairly brief and did not give enough detail for it to be repeated. They were able to say that the results could be plotted onto a scattergraph but did not go any further than that with their results.

Generally, at this level, candidate responses would have significant gaps within the procedure which would mean that it would not be replicable. Candidate responses may also mistakenly suggest that there was an IV and a DV in the study and try and measure using qualitative methods which would not be appropriate for a correlation.

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 operationalised co-variables and then describe in detail how they would be measured. For example, they would suggest that the researcher should use an EEG to measure brain wave activity and note down how long the participant was in deep sleep, or suggest noting down the length of sleep, or how many times they woke up. For noise they would suggest the use of a decibel machine and then would go on to state when it should be used, and how a mean should be taken of the results. Candidates at this level would discuss what a positive and negative correlation would mean for the results, with some drawing their expected conclusions on a scattergraph.

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) This four-mark question asked candidates to explain one practical strength and one practical limitation in relation to one variable identified in their study. To achieve the four marks on this question candidate responses needed to:

- identify one strength and one weakness
- provide some detail for their stated strength and weakness.

Candidate performance on this question was pleasing with many candidate responses showing good understanding of research methodology. Where performance was limited, it was due to either:

- Candidates using different variables for their strength and weakness.
- Giving vague and brief answers about objectivity/reliability and not giving any detail so were limited to one mark maximum.

Candidate performance on this question was often linked to the level of detail given in **part (a)**, with those using scientific measurements often able to give both a strength and weakness and therefore gained the marks available for this question.

PSYCHOLOGY

Paper 9990/23
Research Methods

Key messages

- Candidates need to ensure that their responses are focused on the questions within the exam paper. There was more than one instance 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 detail 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 are able to define/outline key terms within the syllabus such as 'confidentiality' and 'subjectivity' and know the difference between terms such as semi-structured and structured interviews. 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 excellent understanding of named issues/studies but then lost marks for giving generic responses.
- Candidates seem to struggle with questions which ask them to make a judgement about the validity of something. Often candidates would just talk about it 'not measuring what it supposed to measure' which is not enough and is a basic definition of validity. Candidates need to look in detail at the information given and make an appropriate judgement about validity **based on** that information not validity in general.
- Candidates seem to struggle with questions which ask them about hypotheses in general, and non-directional and directional specifically. This is an area that continues to challenge candidates and could benefit with some extra targeted preparation with candidates for future series.
- It is worth noting that candidate responses for the 10-mark extended response question showed good knowledge and understanding of an experiment. There were a number of thoughtful responses for this extended response question, and candidates should be commended for their performance.

General comments

This was the second summer 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 a misunderstanding of the demands of the question. This was clear when looking at questions on subjectivity (**Question 2(a)(b)**), interviews (**Question 6**), and ethical issues when using animals (**Question 8(a)**).

Candidate responses showed good knowledge and understanding of key studies, such as when referring to Bandura et al. and Dement and Kleitman, which meant that many candidate responses we able to achieve most of the marks. Questions which required candidates to comment on the validity of a method or 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.

Comments on specific questions

Section A

Question 1

(a) This one-mark question required candidates to outline the ethical guideline of confidentiality. To achieve the mark available candidates needed to give a brief but accurate outline of the term, such as 'keeping the identify of a participant secret'. This was indeed the most common response seen. Candidate performance on this question was pleasing, most candidates were able to achieve the one mark available. A creditworthy response for this question would be:

- Where the researcher does not reveal the names/identity of the participants.

When candidates did not achieve the mark, it was due candidates giving very vague responses such as 'participants privacy' as this does not show understanding of the term; this was a small minority of candidates however.

(b) This two-mark question asked candidates to explain one way that the ethical guideline of confidentiality was followed in the study by Dement and Kleitman. To achieve the two marks candidate responses needed to provide a linked explanation for this question.

The vast majority of candidate responses were able to achieve at least one of the marks, usually for a generic response e.g. the participants names were not published. For the second mark candidate responses needed to give something specific about the study and how confidentiality was kept. For example, that Dement and Kleitman only used the initials of the participants in their study. The other creditworthy response for this question was reference to only sharing the occupation and gender of their participants.

Question 2

(a) This one-mark question asked candidate to outline what is meant by subjectivity. To achieve the mark candidate responses needed to give a brief but accurate definition of the term. A creditworthy response for this question would be:

- That the way you interpret things/behaviours may be skewed/changed by your own personal bias/viewpoint/opinion.

Candidate performance on this question was mixed with fewer responses achieving the one mark available. This was usually due to vague responses such as:

- Unable to be objective **or**
- Linked to qualitative data.

Both are true in a way but they do not really show knowledge of the term so did not gain any credit.

(b) This two-mark question asked candidates to suggest how subjectivity could be a problem when studying people's recall of past events. To achieve the two marks available candidates needed to:

- Give a reason that subjectivity is a problem e.g.,
  - Post-event details changing memory
  - Older memories are displaced by newer memories, so they forget
  - Personal experience may distort how an event is remembered.
- Detail the consequences of the problems above e.g. that two people therefore may remember the same event differently.

Candidate performance on this question was mixed and marks achieved were very much linked to the knowledge of the term shown in part **(a)**. Invariably if the candidate response had not given the right definition of the term in part **(a)** then they would not be able to achieve marks on part **(b)**. Where candidate responses did achieve a mark, it was usually due to their knowledge of post event information distorting memories with some discussing about the researcher's experiences affecting how they interpret the participants response which was also creditworthy.

Question 3

(a) This three-mark question asked candidates to sketch a graph to show the results of this study. Including labelling the axes. To achieve the three marks candidate response needed to label the two axes with sleep duration and physical health, provide some units (hours) for the sleep duration **or** physical health. The candidates also needed to draw points/line which would suggest a positive correlation. Candidates could not achieve full marks without the plotting of the points/line. Candidate performance on this question was pleasing with the vast majority of candidate responses achieving the three marks available. Where candidate responses did not achieve the three marks it was usually due to the lack of a sketch that showed a positive correlation, or drawing a bar chart or histogram.

(b) This four-mark question asked candidates to suggest two problems with collecting data about the number of hours of sleep in this study. To achieve the 2 marks for each point candidates needed to:

- suggest a problem and
- link this problem to the scenario.

Creditworthy responses for this question include:

- that participants may lie and say they've **slept** longer
- the **sleep** may be disturbed by external factors such as cats
- the environment they **sleep in** can't be controlled such as noise.

The important part to remember for a question like this is it needs to be linked to sleep. Candidate performance to this question was pleasing. Most candidates were able to achieve at least one mark for each point for the idea of extraneous variables affecting their sleep (as at home not in a lab). Where candidate performance was limited to two marks it was due to the candidate response not mentioning sleep in their response and therefore meaning it was a generic response and only able to achieve the one mark. For future series candidates need to remember that if a question is related to an unseen study, then their responses need to explicitly reference the study.

Question 4

This two-mark question asked candidates to state whether this is a directional or a non-directional hypothesis and to justify their answer. To achieve the two marks candidate responses needed to:

- justify why the hypothesis is non directional
- and to link this response to the scenario.

The important point to make for this question is that candidates did not achieve any mark for identifying whether the hypothesis was directional/non-directional they needed to say **why** it was one or the other and then linking this response to the scenario. Therefore, a generic justification would get 1, and a linked justification would get 2. For example:

- That it only talks about a difference rather than the way the difference will go – 1
- That it only says that levels of obedience will be different dependent on levels of intelligence = 2

Candidate performance on this question was mixed. Candidates seem to struggle with hypotheses questions especially when asked about the type of hypotheses seen. Consequently there was a significant proportion of candidate responses which did not achieve any marks, especially given that many talked about a relationship which was not creditworthy.

Question 5

This two-mark question asked candidates to suggest why Bandura et al. may **not** have used volunteer sampling. To achieve the two marks candidate responses needed to suggest a problem with using volunteer sampling and provide a link to the scenario within their response.

Candidate responses to this question were below expectations. A significant proportion of candidate responses talked about volunteer samples in general or suggested that the children are unlikely to see adverts. Candidates need to ensure that they read the whole question rather than focusing on one or two words and then just discussing those words.

Creditworthy response would include therefore:

- because you may get a biased sample = 1 (generic)
- because you may only get children whose parents think they're child is not **aggressive (linked)** – 2
- parents of aggressive children are unlikely to volunteer them for the study (linked) = 2.

The most common mark on this question was zero and one, with few two marks responses seen.

Question 6

This six-mark question required candidates to describe structured interviews and semi-structured interviews, using any examples. Awarded marks could come accurate description of each term, further detail about the term and appropriate examples such as showing knowledge of the interviews within studies on the specification such as Dement & Kleitman, and Milgram. Candidate performance on this question was mixed with many struggling to give enough detail to gain the 6 marks even with a decent amount of knowledge of each term. It is worth noting that knowledge of general interviews would gain some credit. Examples of possible creditworthy responses were:

For interviews in general:

- self-report
- can use open and closed questions
- can be face to face/online etc.

For Structured interviews:

- follows a standardised set of question/a script/predetermined set of questions
- the order of the questions is the same
- set tone of voice
- wears the same clothes
- **more likely** to be closed ended.

For Semi structured interviews:

- some set/predetermined questions
- have the ability to ask to follow up questions dependent on responses
- can be open or closed questions (only credited once)
- no fixed order in how the questions are asked.

At the lower end of the mark range, candidates usually achieved one or two marks for suggesting that interviews are self-reports, and that structured interviews have set questions. Candidate responses often struggled with semi-structured interviews at this range which led to weak and confused responses.

At the higher end of the mark range candidates produced some thoughtful responses and the vast majority were able to give detailed explanations of interviews in general and structured interviews but still struggled slightly with semi-structured interviews. The examples given were appropriate and explicitly linked to the question. These candidate responses often achieved around four to five marks.

Question 7

(a) This two-mark question asked candidates to suggest two questions that Dr Jones could use in a self-report (questionnaire or interview) that would investigate different features of Kevan's mindfulness. These questions could be open or closed. This question gave candidates a fair amount of scope for the questions and in the main candidate performance on this question reflected this. To achieve the marks candidates needed to have some reference to mindfulness (or a feature of mindfulness) in their question e.g. being in the moment. The vast majority of candidates response achieved some marks for this question, with many achieving both marks.

(b) (i) This two-mark question asked candidates to suggest two changes in Kevan that his friends could observe. These changes needed to be behavioural, emotional or cognitive. Similar to the previous question there was significant scope for candidates to answer this question.

The most common creditworthy responses were:

- Have they observed him smiling more (happier)/more or less angry
- Have they seen increased incidence of kindness
- Has he started to improve his performance at work – behavioural/cognitive
- Has he started to solve problems/resolve issues more easily
- Does he seem more motivated at work.

The vast majority of candidates were able to achieve at least one mark on this question. However, many candidate responses lost marks due to vague responses which did not really answer the question.

(ii) This two-mark question asked candidates to explain why one of the changes suggested in part **(b)(i)** would be difficult to investigate. Candidate performance on this question was directly related to their performance on the previous question.

To achieve the two marks candidate responses needed to say:

- why it may be a problem to observe that specific characteristic
- some elaboration/detail on this problem identified.

Creditworthy responses, therefore, could include:

- Friends do not often sleep in the same house/room together
- Friends would only see a specific aspect of their life
- Such traits as kindness are subjective and one person may see something as kind and another does not
- You cannot observe happiness as external features, such as smiling does not mean he is happy.

In essence, if candidates had explained a characteristic in **7(b)(i)** that was explicit and easy to discuss in terms of evaluation then the usually were able to achieve the marks available on this question. However, a lot of candidate responses on **7(b)(i)** were fairly complicated/vague which meant that they then struggled to identify a problem for this part of the question.

(c) (i) This one-mark question asked candidates to explain which problem, from a given scenario, relates to reliability. The most important part of this question is the command word 'explain'. Candidates do not get any marks for identifying that this is problem 1, they needed to explain why. In essence, candidates needed to discuss the lack of **consistency** in interpretation. Candidate performance on this question was mixed, with many just identifying the problem rather than explaining it.

(ii) This one-mark question asked candidates to explain which problem, from a given scenario, relates to validity. Like previously, the most important part of this question is the command word 'explain'. Candidates do not get any marks for identifying which problem this was (in this instance it could be 1 or 2), they needed to explain why. In essence, candidates needed to say that it is not testing what it is supposed to test (for Item 2) or because the inconsistencies mean Dr Jones cannot be sure that any of the interpretations are true/correct (for item 1).

Candidate performance on this question was mixed, with many just identifying the problem rather than explaining it. For futures series candidates need to ensure that they understand the demands of each command word.

Question 8

(a) This two-mark question asked candidates to explain why Suggestion 1 (animals with less complex brains) would make Dr Mohan's study more ethical. To achieve the two marks candidate responses needed to:

- Give an animal ethical guideline related to using animals with less complex brains
- Provide some detail about the guideline identified.

Candidate performance on this question was as expected with many candidate responses achieving at least one mark on this question. The two main creditworthy responses for this question were:

- A simpler animal will not become as **distressed** as more complex animals = two (as the guideline of distress was part of their explanation)
- **Choosing** the least complex animal will mean that they are less likely to suffer **harm** = two.

Where candidate responses did not achieve any marks, it was often due to the use of human ethical guidelines such as protection from harm, or withdrawal. For future series candidates need to ensure they know the differences between animal and human ethics.

(b) This two-mark question asked candidates to explain why Suggestion 2 (using videos of the animals in the natural habitat) would make Dr Mohan's study more ethical. Similar to **Question 8(a)**, to achieve the two marks candidate responses needed to:

- Give an animal ethical guideline related to using videos of animals
- Provide some detail about the guideline identified.

Candidate performance on this question was mixed, and often candidates struggled more to get the second mark on this question than part **8(a)**. One mark was often achieved for the basic suggestion that it meant that the animals would not need to be in a laboratory, but then the second mark was not gained as this explanation was not linked back to a specific ethical issue. The most common error on this question was that candidate responses went into answers that were not related to ethics such as talking about seeing more natural behaviour which is a **strength** of using animal videos but not an ethical strength.

Question 9

(a) This one-mark question asked candidates to suggest a suitable frequency for Kioni to use to test her participants throughout their lives. To achieve the mark candidates needed to give a frequency that was between one year and (at the limit) 20 years. 20 years was the highest due to the need to do multiple tests (at least 3) over the period of time.

Many candidates achieved the mark on this question. Where candidates did not achieve the mark, it was due to suggesting a frequency that was too close together (with every six months being popular).

(b) This two-mark question asked candidates to suggest one way for Kioni to re-contact her participants each time she needs to test them. To achieve the marks on this question candidate responses needed to:

- Give an appropriate way of recontacting
- Link their way of recontacting to the scenario/give some detail how that number was given i.e. linked to personal space **or** that they got the number to recontact from the participants at the beginning of the study.

Most candidates were able to achieve at least one mark for suggesting an appropriate way of recontacting e.g., phone/email/text. However, only a minority of candidate responses were able to

achieve the second mark as there was no actual link to the study itself/detail to where the number came from therefore one was the most common mark for this question

(c) (i) This two-mark question asked candidates to describe one way that Kioni could measure personal space. To achieve the two marks on this question candidate responses needed to describe the way chosen in detail. Basic or incomplete answers could achieve one mark. The most common creditworthy responses for this question included:

- A replication of the methodology used by Perry.
- A questionnaire with examples of specific questions needed for the two marks.
- An observation was creditworthy, though candidates who did suggest an observation often struggled to give detail of how they would do this e.g., observing someone move closer to a participant and observe for signs of discomfort.

Candidate performance on this question was directly related to the method chosen to measure personal space. Those candidates using a simple questionnaire often were able to achieve one or two marks however those that decided to replicate Perry often gave slightly confused answers which often would only achieve one mark maximum.

For future series candidates need to ensure that they write their responses as clearly as possible and not over-complicate their answers as this often leads to confusion and a reduction in marks awarded.

(ii) This two-mark question asked candidates, for the way described in part **(c)(i)**, to suggest one strength of measuring personal space in this way. As would be expected, candidate performance on this question directly related to the answer/methodology given in part **9(c)(i)**.

Candidates who described a simple questionnaire often were able to gain marks on this question through the use of closed questionnaire enabling the researcher to gain quantitative data. In this case, however, many did not give enough detail to enable the response to achieve the second mark. Candidates who gave a replication of Perry in part **(c)(i)** often struggled to achieve marks on this part, with many giving weaknesses e.g., lack of real-life interactions, rather than a strength.

(d) (i) This two-mark question asked candidates to suggest two variables that could affect an individual's personal space, other than age and gender. Candidate performance on this question was pleasing with many candidates getting the full two marks. There were a number of thoughtful responses for this question including:

- Culture
- Past experiences
- Health status
- Personality e.g., extrovert or introvert
- High/low empathy
- Mental health
- Smell (too much perfume)
- Dress (some dress more intimidating).

All of the above were creditworthy and showed a good understanding and appreciation of the topic area.

(ii) This two-mark question asked candidates, for one of the variables suggested in part **(d)(i)**: to suggest how this variable could affect an individual's personal space. Candidate performance on this question was directly linked to the answers given on part **(d)(i)**. Where candidates had written about such factors as culture/personality they often went on to achieve the 2 marks for suggesting that introverts may not want to be around a number of people, that some cultures prefer a larger area of personal space. Where performance on this question was more limited it was invariably due to either candidates not writing an appropriate answer in part **(d)(i)** or the vagueness of an answer where only one mark could be awarded.

Question 10

(a) This ten-mark essay question asked candidates to describe how Isla could conduct an experiment to test whether attention is better with a pen or an object that is not used for writing. 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 but there were a number of level 3/4 responses which was pleasing to see.

Many candidates had a sound understanding of experiments in general and were able to make relevant decisions about the independent and dependent variables. However, where candidates struggled slightly was with the operationalisation of the independent variable 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 controls needed to make the experiment standardised/reliable and candidates excelled on this part of their response, with many giving relevant and accurate examples of the controls they could use.

Candidate responses within the lower mark range were often able to give a basic independent variable e.g., use of a pencil and use of non-writing object, but would not go further to say what this non-writing object would be. However, 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 experimental design element was often missing or very basic i.e., a brief identification of repeated or independent measures. 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 a non-writing object that could be used opposite a pen, with some having more than one non-writing condition and comparing which was fine. The controls given by candidates at this level would be well thought out e.g. using the same room/lecture/information. What differentiated candidate responses at this level was whether they explicitly talked about the experimental research design used for the experiment, and how it could be implemented. 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.

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) (i) This two-mark question asked candidates to explain why one feature described is a strength of their study. To achieve the two marks candidates needed to:

- Identify a strength
- Give some explanation of the strength identified.

Creditworthy identifications could be:

- The use of independent measures/ repeated measures
- That there was a high level of control
- That attention was operationalised through test scores/quantitative measures.

The detail about **why** this is a strength for the second mark could include:

- Independent measures meant a lack of order effects
- High level of control means replicability to test for reliability
- Use of standardised/quantitative measures means results objective/can be compared/analysed.

Candidate performance on this question was mixed. The most common error in responses was a lack of detail meaning that the second mark was often not available. Unfortunately, when candidates had not given sufficient detail in their part **(a)**, they often struggled to achieve the marks on this part.

(ii) This two-mark question asked candidates to explain how that feature could be a weakness of the study. Like part **10(b)(i)**, candidates needed to:

- Identify a weakness
- Give some explanation of the weakness identified.

Creditworthy weaknesses could be similar/the same as the strengths such as:

- The use of independent measures/ repeated measures
- That there was a high level of control
- That attention was operationalised through test scores/quantitative measures.

The detail for the second mark could then be:

- That participant variables such as intelligence may be an issue for independent measures/order effects for repeated measures
- That the high control meant it was more artificial so susceptible to demand characteristics
- That the use of this type of measure does not give any insight why.

Candidate performance on this question was mixed. Like responses for part **(b)(i)**, the most common error was a lack of detail meaning that often the second mark was not available. Unfortunately, when candidates had not given sufficient detail in their part **(a)**, they often struggled to achieve the marks on this part.

PSYCHOLOGY

Paper 9990/31

Specialist Options: Approaches, Issues and Debates

Key messages

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

These questions in this exam asked candidates to apply an area of the syllabus (theory, diagnostic criteria, technique/treatment, self-report, etc.) to explain how it is relevant 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 the suggestion such as the theory/technique that was outlined in the candidate's response to part **(a)** of the question. In this exam, these types of questions asked the candidate to evaluate the theory/technique outlined in part **(a)** such as with a problem with the theory/technique and explain why this is a problem in the scenario given in the stem of the question. It would be helpful for candidates when doing revision to learn strengths and weaknesses/problems of the theories, techniques, using diagnostic criteria, treatments, etc. that they have learned and put these into their revision notes. They should also practice explaining the evaluation point in the context of the question.

Questions 2, **6**, **10 and 14**

Part (a)

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

Part (b)

This part of 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 might ask candidates to explain a strength or a weakness of the issue/debate or the syllabus content outlined in part **(a)**. It would 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 question. 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 two 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 two treatments, two theories (possibly with reference to research that investigated the theory), and a key study identified in the specification under the appropriate bullet point. For this exam, some of the answers used the incorrect topic area in the syllabus or the description was brief. 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. 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 study/studies, theories, 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, in order 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 study/studies, theories, 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, 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. Several of the responses did do analysis. Candidates should be aware this question is worth 10 marks and they 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 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. Responses that achieved 3-4 marks were able to explain the different components of CBT such as creating a therapeutic alliance, identifying faulty thinking, reality testing (often with an example of either a delusion or a hallucination), keeping a diary and an example of homework that Mahmoud could do outside of sessions. Weaker responses were sometimes brief and just identified that CBT can challenge thinking without any reference to schizophrenia and how CBT can assist with positive symptoms. A small number of responses outlined how CBT could help with a negative symptom which was not creditworthy. Some responses included a detailed outline of the symptoms of

schizophrenia which was not creditworthy. These responses often then outlined how CBT could be applied to a positive symptom, but it was often very brief.

Question 2

(a) There were some good responses that were able to clearly outline Miller's feeling-state theory of impulse control disorders. Full mark responses outlined the different components of the feeling state and that the memory of the positive feelings lead to the individual wanting to perform the impulsive behaviour again. Other responses outlined an example of the feeling-state for a specific impulse control disorder such as gambling or pyromania which was also creditworthy. Weaker, 1-mark responses outlined a link to positive feelings and specific behaviours (e.g., gambling). There were also many responses that focused on the negative rather than the positive experience that are felt when either engaging in or after the compulsive behaviour which was not creditworthy.

(b) Some responses were able to achieve one mark for this question by explaining one reason why Miller's feeling-state theory could be considered holistic. These often explained that it was due to the theory considering emotions, physiological arousal and cognition (memory) to explain impulse control disorder. A small number of responses achieved the second mark for providing a definition of holism with reference to it being about the bigger picture or how components work together to cause behaviour. Credit was given to genetic definitions or holism or in the context of Miller's feeling-state theory. Candidates found it difficult to define holism and frequently stated that it involved multiple factors which was not creditworthy. In addition, many were unable to explain one reason why Miller's feeling-state theory could be considered holistic. These types of responses often incorrectly identified the different parts of the feeling state or just identified one or two.

Question 3

(a) (i) Most responses were able to achieve at least one mark by suggesting one reason why Emi may be diagnosed with pyromania. This was achieved by citing an example from the stem such as feeling better after setting a fire, fascination with fire or feeling excited when thinking about starting a fire. Responses that scored full marks were able to relate the example to one of the diagnostic criteria in the ICD-11. It was not a requirement of the question to specifically identify the ICD-11 but the candidate needed to make it clear how the example given of Emi's behaviour/thoughts fit a symptom/diagnostic criterion of pyromania. A few responses did not receive credit, and this was often because they made a vague link to the stem such as 'she felt anxious about the situation'.

(ii) Those responses that received credit either referred to the fact that the factory fire had been done out of revenge or that Emi had only set one (or two, including the matches) fire. A few candidates linked this to ICD-11 criteria that there is no motive for setting the fire or there are multiple acts of fire-setting. Some responses suggested that Emi had not had these impulses/fire-setting behaviour for very long which was not creditworthy as the ICD-11 does not state a time period for the symptoms. Some responses suggested that she would not be diagnosed with pyromania because either Emi did not act on her thoughts to start a fire when she was in a new job or that it has not impacted her life significantly enough. Neither of these are in the diagnostic criteria so these types of responses were not creditworthy.

(b) There were some full mark responses to this question. Common problems included Emi being dishonest with the doctor about the fire setting due to legal reasons or comorbidity with other disorders such as anxiety. Full mark responses gave a clear link to pyromania whereas the weaker, 1-mark responses outlined a generic problem. Some responses just repeated the question in their answer e.g. Emi does not meet all of the diagnostic criteria, which was not creditworthy.

Question 4

(a) Responses varied for this question and covered the full range of the marks available. Level 3 responses were less common with most responses achieving either Level 1 or Level 2. Those that did achieve Level 3 kept their response focussed on SSRIs and exposure and response prevention (ERP) as a treatment/way to manage obsessive-compulsive disorder (OCD). Level 3 descriptions of SSRIs went beyond that it just blocks the re-uptake of serotonin. These responses were able to fully describe that it blocks the reuptake into the presynaptic neuron therefore leaving more serotonin available in the synaptic gap. They were also able to talk about increasing serotonin receptors and then finally linking it to specific symptoms of OCD. ERP was often described in more depth and with more accuracy of the steps involved in this therapy and how this would lead to a

reduction in obsession(s) and compulsion(s) of the client with OCD. Many responses included an outline of Lehmkuhl et al.'s study on Jason with a focus on ERP and Rapaport's study on Charles with the focus on the SSRI used and its effect on Charles.

Weaker responses showed a poor understanding of SSRIs and how they work to help with symptoms of OCD. Some included dopamine and oxytocin which SSRIs do have an indirect effect on but that is not the main neurotransmitter and neither does it link to OCD, so this was incorrect. There were a number of incorrect responses that attempted to describe what 'blocking the re-uptake of serotonin' means. Many stated it was blocked in the brain or in the post-synaptic neuron which was not creditworthy. Level 1 responses were often very brief for both SSRIs and ERP. Some responses failed to provide a clear outline of ERP and gave a generic description of talking therapy or mixed it up with other treatments such as systematic desensitisation (linked to phobias). Neither of these types of responses were creditworthy.

(b) The marks for this part of the question did cover the full range of marks available with the most frequent levels awarded being Level 1 and 2. The responses that did achieve Level 3 and above structured their response issue by issue and often started with the named issue of the use of children in research, along with clear and specific examples from either Lehmkuhl et al.'s study on Jason and/or Rapaport's study on Charles and analysis. Frequent points raised regarding the use of children included application to everyday life, lack of demand characteristics, communication issues and ethics as both a strength and a weakness. Good responses were able to discuss how the research on children is ethical as consent was obtained from parents and the child's OCD symptoms reduce. Weak responses stated that the studies were unethical as consent is not given by the child. This is not a creditworthy point as consent from the parent makes the study ethical and, depending on the age of the child, consent could be given by them as well. In addition, many attempted to argue that both ERP and SSRIs are harmful. Credit was given to points regarding the side effects of SSRIs as harmful to children. If the candidate stated that ERP is harmful as it causes anxiety; this was only credited if it was made clear in the response that the anxiety was short-term and the therapy helped reduce anxiety in the long term due to treating the child's OCD symptoms. Apart from the named issue, other popular issues covered were determinism versus free-will, situational versus individual, generalisability, idiographic and nomothetic, application to everyday life and reductionism versus holism.

Weaker responses achieving Level 1 or Level 2 did not contextualise their response. While some of the evaluation points were valid, the lack of context prevented responses from achieving a higher band. There were several generic definitions and a very weak attempt to include context by identifying either SSRIs or ERP as the example. Some responses took the approach to evaluate the two treatments one after the other. This usually resulted in very superficial points with a lack of context. For example, the candidate would state that SSRIs are unethical as they have side effects without giving any examples or explaining that side effects will lessen the longer the patient takes them for. It was also very common for responses to state that SSRIs reduce symptoms immediately which is incorrect as they will take a few weeks to start to reduce symptoms of OCD.

Consumer Psychology

Question 5

There were a few good responses to this question, and some achieved full marks. Full mark responses were able to suggest two ways that the shop could change their virtual store layout that could reduce customer complaints. Strong responses were often concise. Grid layout was seen as easiest and freeform as most entertaining. There was a clear link to reducing complaints, particularly when referring to the current racetrack layout. Weaker responses did not relate their suggestion to reducing complaints. Some responses mixed up the effect of a grid layout for a freeform layout or vice versa which was not creditworthy.

Question 6

(a) There were a number of full mark responses that gave a clear definition of free-will and linked this to how system 2 uses free-will. Most were able to outline a definition of free-will. Some candidates gave a vague definition of free-will such as 'being able to express yourself' which was not given credit. In addition, many responses could not give an example of how system 2 uses free-will although some of these did understand what system 2 thinking involves. There were some responses that compared system 2 thinking to system 1 which did not address the question and therefore was not creditworthy.

(b) Some responses were able to gain full marks by explaining one problem a psychologist may have when they investigate whether consumers use free-will when deciding to purchase products. A common response was that it is difficult to see/measure what the consumer is thinking and determine whether they are using free-will. In addition, there may be external factors such as advertising/promotions/packaging that could influence the consumer's choice without the consumer being aware of this. Therefore, the consumer would tell the psychologist that they used free-will when in fact the purchase was determined by external factors. Few responses were able to relate the problem to purchasing a product to achieve the second mark. Some responses explained that external factors could affect purchase without linking this to a problem that a psychologist may have when investigating free-will. These types of responses were not creditworthy.

Question 7

(a) (i) Some responses were able to achieve full marks by identifying the anchoring heuristic and then provided an example from the stem on how Sally used this heuristic when making the decision of which running shoes to purchase. These responses often stated Sally was influenced the most by the first piece of information she received. Weaker responses could not explain from the stem how Sally used the anchoring heuristic. Many candidates simply mentioned it was the first item she saw without explaining how the anchoring heuristic comes into effect. A significant number of candidates identified the incorrect heuristic or restated what had already been outlined in the stem e.g., Sally bought the shoes from the first shop she went in or she bought them as they were first. These types of responses were not creditworthy.

(ii) Compared to part **(a)(i)**, a higher proportion of responses correctly identified the recognition heuristic and referred to Dan's brand loyalty due to familiarity, knowing the brand well or trusting the brand due to past experience to achieve full marks. Weaker/non-creditworthy responses often provided an incorrect heuristic such as 'take-the-best' and/or were not able to explain from the stem how Dan used the recognition heuristic. Many candidates simply mentioned he knew the running shoe brand but did not elaborate about Dan trusting the brand/long history with the brand. These types of responses were not creditworthy.

(b) A few responses were able to achieve one or two marks for this question. Creditworthy problems outlined included that consumers may use multiple heuristics or it is difficult because researchers do not know what a consumer is thinking. Some explained that an interview or questionnaire would need to be used, and this might lead to invalid results as consumers may lie/produce socially desirable answers. Alternatively, consumers may not be aware of which heuristic they used as these are mental shortcuts and are often unconscious. The best responses were able to give an example in the context of deciding on a product to purchase. Some responses used an example from the STEM, although this was not required for full marks.

Weaker responses frequently gave a brief outline of a problem without any reference to choice heuristics used when purchasing a product so could only achieve one mark. Some responses attempted to explain that other variables might influence the choice of product such as pricing/advertising. These types of responses were not creditworthy as consumers still use heuristics with these factors. There was a poor understanding in many responses of what is meant by a choice heuristic which meant that they failed to achieve any marks. 15 per cent of candidates did not attempt to answer this question and often achieved 0 marks for their responses to part **(a)(i)** and **(a)(ii)**.

Question 8

(a) There were a few Level 3 responses to this question. Responses that obtained this mark band were able to describe example studies to support their response. The best described clearly the Porublev et al. study on gift-wrapping. Atalay et al. on attention and shelf position was sometimes well-described by candidates who described one of the three studies. The best responses had an outline of participants, procedure, and the results from the example studies. Some responses outlined either the Howard study or the Rixom et al. study on gift-wrapping. However, candidates who described one of these studies frequently had very limited knowledge of the study and either gave a result or a conclusion from the study which was awarded Level 1. Some responses attempted to outline the three studies done by Atalay et al. but frequently none of them were described in any depth and candidates often confused the procedure and result between the three studies.

There were a significant number of weaker responses that often gave anecdotal/generic descriptions of gift-wrapping including beliefs of giver and recipient. These included comments about buying and receiving presents and how we would like them wrapped or not wrapped. There was lots of inclusion of cultural differences for colour that had no real psychological evidence to back it up. These types of responses often achieved Level 1. Similarly, responses then attempted to outline what psychologists have discovered about attention and shelf position but did not know a relevant study so simply outlined where products should be placed on shelves in supermarkets and shops which was not creditworthy.

(b) There were a very small number of Level 3 and above responses to this question. Those candidates who had outlined at least one study in part **(a)** were able to discuss the named issue of generalisations from findings. Candidates often used cultural difference in their discussion about generalisations, mentioning that different cultures may wrap gifts differently thus traditional wrapping/non-traditional wrapping preferences may not be generalised across different cultures. Weaker responses for the named issue frequently lacked knowledge of the sample from the studies outlined in part **(a)** so just identified that the studies were done in the 'West' and therefore could not be applied to other cultures. These types of responses were awarded Level 1 for this issue as this is basic evaluation.

The best responses covered this question by structuring it by issues, along with supporting examples from the studies in part **(a)** and analysis. Apart from the named issue, other popular issues covered were deterministic versus free-will, application to everyday life, validity and idiographic versus nomothetic. Some responses attempted to use applications to real life but simply stated that the findings will 'improve sales' without elaborating how or providing evidence from the studies. Better responses were able to outline how stores should include gift wrapping services to improve customer service or put their best valued products at the middle of the shelf to improve profits.

Most responses for this question achieved in Level 1/Level 2. This was due to writing superficial evaluation points. The issue was named and either the theory or the study was either stated to support the issue or which side of the debate it supported. This question had several very brief responses often with just a few sentences within which numerous issues were identified with little else provided. It was also noted in a significant number of candidates who had chosen consumer as one of their options, did not attempt the question with 20 per cent leaving the answer space blank.

Health Psychology

Question 9

There were some strong responses to this question which achieved 3–4 marks. The best responses were able to outline what is meant by false positive and provide several reasons why a doctor might make a false positive diagnosis. Common suggestions included patients giving false information either deliberately or due to shyness/embarrassment, the patient might exaggerate their symptoms, or language barriers. Other creditworthy suggestions included problems of clinical tests such as blood tests, lack of time during consultation and lack of expertise by doctor. Some wrote about how the doctor's skill level caused them to misinterpret symptoms/medical readings which led to a false positive.

A few responses confused false positive diagnosis with a false negative, which was not creditworthy. However, responses that mentioned factors applicable to both types of diagnostic errors, such as patient misinformation or clinician inexperience, were awarded marks.

Weaker answers often lacked a definition or example of a false positive diagnosis or provided only one reason without elaboration. Some responses listed multiple similar points – such as 'patient lying,' 'patient exaggerating,' and 'patient providing false information' – which were considered a single idea and thus limited the marks awarded. Additionally, some candidates focused on why a patient might provide false information (e.g., due to conditions like Munchausen syndrome), which did not directly address the doctor's role in making a false positive diagnosis and therefore did not receive credit. Some responses suggested that that the doctor misdiagnosed the patient with a different disease, which is not a false positive diagnosis. This part of the response did not receive credit.

Question 10

(a) There were a number of full mark responses that accurately defined the nomothetic approach and then gave an example from a study on preferences for practitioner clothing. The vast majority used the McKinstry and Wang's study as their example, such as patients' preferences for male doctors in suits and ties and female doctors in white coats. Weaker responses often were able to define the nomothetic approach but unable to give an example from a study or gave an incorrect result.

Some candidates correctly stated that the nomothetic approach involves collecting quantitative data, earning partial credit. However, they often failed to link this to a concrete example from the study, limiting their marks. Merely stating that the study used closed questions without elaboration was too vague to be credited.

A number of responses lacked a clear definition of the nomothetic approach or confused it with other concepts like reductionism, which did not receive any marks. Additionally, some candidates described the study's aim or procedure without explaining how it exemplified a nomothetic approach. These examples from the study were not creditworthy.

(b) There were a few good responses to this question. The most common weakness given was that the nomothetic approach does not consider individual or cultural differences. Full mark responses frequently gave an example of a difference such as the older participants in the McKinstry and Wang study felt more strongly about the importance of dressing formally compared to younger participants. The candidate did not need to refer to a result in a study and some gave an example of a possible, specific cultural difference in preferences.

Some responses explained the weakness was due to a lack of in-depth/qualitative data. These often achieved 1 mark as the candidate was unable to elaborate or just briefly explained that we do not know the reason for the choice of practitioner clothing which was a basic outline of this weakness.

A common error was that the nomothetic approach uses qualitative data instead of quantitative and an outline of a weakness of qualitative data was given that did not receive credit. Some responses outlined a result from the McKinstry and Wang study but did not link this to the weakness identified. These type of responses were awarded one mark.

Question 11

(a) This question was well answered by candidates and many achieved full marks. Common suggestions included pill counting, TrackCap, and blood/urine tests. Many responses were able to achieve two marks for each suggestion by explaining how the measure works or how it will show that Dr Smith's patients are taking their medication correctly. Weaker, one-mark responses often just identified the method but were unable to suggest how it would measure patient adherence. Some responses suggested that a Trackcap can count the pills removed but this is incorrect as it tracks the time the bottle was opened. In addition, some candidates who suggested a blood or urine test did not explain that Dr Smith will need to check the results to see if their medication is present in the blood or urine so were unable to achieve full marks.

(b) There were a number of full mark responses to this question where the problem was explained and why this might lead Dr Smith to an inaccurate measure of his patients' adherence. Common responses for this question used pill counting and TrackCap. Most candidates explained that patients could possibly remove the pills on their own where neither method would be able to track this. Another common problem was that social desirability may take place when an interview is used. Candidates achieved full marks by explaining that this will mislead Dr Smith into believing patients have adhered to medicine where in reality they have not.

Weaker, one-mark responses were often brief and did not link the problem to Dr Smith measuring adherence of his patients and instead explained the problem the patient would have in measuring their own adherence which was not answering the question. Candidates found it difficult to explain a problem with blood or urine tests. Many attempted to explain that these tests are unethical which was not creditworthy. This is because Dr Smith can explain the reason for these tests and get informed consent before the medication is prescribed. He can explain how the blood/urine test will support the patient's health and show whether the medication is working.

Question 12

(a) There were a significant number of responses that achieved Level 3. Many were able to identify features of the sample (e.g., all women, all under 70 years old and had suffered breast cancer). These responses also identified the three conditions, gave some details of the procedure (e.g., what the participants did in the relaxation and imagery condition), one or more of the questionnaires used in the study and a result such as mood disturbance score improved the most for the relaxation plus imagery group. This was particularly found with women age 55 plus. Women who had high anger scores at the start of the study did not respond well to either relaxation or relaxation plus imagery.

Weaker responses gave fewer details of the study or gave incorrect information e.g., wrong number of participants/saying it was a repeated measures design. A significant number of responses thought that stress was measured and quoted that as a variable in the results section. If no creditworthy result was given in the response the maximum mark that the candidate could achieve was 5 marks.

Many of the responses described details of the study in part **(b)** that should have been included in this part of the question. This meant that these candidates were awarded a low mark despite knowing the study well.

A significant minority of responses appear to have used the title of the study to create a study. These responses sometimes achieved Level 1 as they could give a brief description of relaxation and/or imagery.

(b) There were a small number of responses that could achieve Level 3 or above. These types of responses started with the named issue, which was field experiments, and were able to give some strengths and weaknesses. Candidate responses achieved Level 3 and over due to giving clear and detailed reference to the Bridge et al. study from part **(a)**. The strongest responses also included some analysis. This was rarely well explained and therefore often Level 3 was awarded.

An example of analysis for field experiments was that this method has the strength of taking place in the natural environment which for this study was a hospital which was a very familiar for the participants as they had been receiving treatment for cancer in a hospital. In addition, the participants in the two intervention conditions were told to practice at home which also has good ecological validity. The women may find it easier to do relaxation while at home. However, a weakness of a field experiment is that they lack control. This study did have some controls in place such as being given standardised training in relaxation/imagery. However, it is possible that some of the participants have been doing other activities to manage their mood such as exercise, talking to a friend or family member, etc.

Other issues discussed included generalisability, individual versus situational, reliability, validity, strengths and weaknesses of questionnaires.

Level 1 and Level 2 responses did superficial evaluation/discussion of the key study. Evaluation points were often creditworthy but completely devoid of any context from part **(a)**. It was common for responses to cover 5–8 evaluation points where the candidate identified the issue and then stated whether it was a strength or a weakness of the study or which side of the debate the study/results from the Bridge et al. study supports without any example or explanation about why the study supports this side of the debate.

There were quite a few responses that attempted an application to everyday life issue. Though most responses could not give any specific clear application and how the application will be carried out. Most responses simply mentioned it can be done in hospitals or other patients with cancer without further elaboration.

24 per cent of candidates who answered the questions in the health option offered no response here or they evaluated a study that they had invented themselves which was not creditworthy.

Organisational Psychology

Question 13

Responses covered the full range of the mark scheme. Good responses were able to identify one or two of Heifetz's six principles in meeting adaptive challenges and link this to helping worker to adjust to the new/change in working hours, tiredness or mistakes made at work. Common principles included get on the balcony, regulate distress and give work back to the people. Weaker responses were often able to identify the principle but then did not link the principle to the stem. Many did not know the principles and therefore were unable to access the marks.

Question 14

(a) There were many good, full mark responses to this question. These were able to give a clear definition of the nature versus nurture debate. Weaker responses were usually when attempting to outline the nature side, some simply mentioned we are 'born that way' or gave a vague statement such as 'a natural occurrence that can't be changed' which were not creditworthy.

(b) There were some full mark responses to this question. These were able to outline one reason why Scouller's levels of leadership theory supports the nurture side of the nature versus nurture debate. For example, responses often explained how the leader interacts with people (publicly or privately) and learns through these interactions what leader characteristics work well with larger and/or smaller groups. These types of responses were able to outline a supporting answer in context with the 3 levels of leadership, outlining how it is learnt as we interact across the 3 levels. Weaker responses were able to explain why levels of leadership support nurture (e.g., leader interacts with others at the public level and learns from it) for 1 mark but struggled to access to second mark. There were several responses that did not receive any credit for this question. This was because they gave generic statements about how leadership is learnt with no mention made of Scouller's levels of leadership.

Question 15

(a) (i) Good, full mark responses were able to identify a relevant Belbin Team role and explain how Margot fits the role appropriately in the context of checking the leaflet. Completer-finisher and monitor-evaluator were the common team roles selected. Weaker responses could identify an appropriate team role but could not justify their response with a clear link to Margot checking the leaflet. Non-creditworthy responses identified an inappropriate team role or invented a team role which was not from Belbin for Margot.

(ii) Good, full mark responses were able to identify a relevant Belbin Team role and explain how Tricia fits the role appropriately in the context of designing the leaflet. Plant and implementer were the common team roles selected. Weaker responses could identify an appropriate team role but could not justify their response with a clear link to Tricia designing the leaflet. Non-creditworthy responses identified an inappropriate team role or invented a team role which was not from Belbin for Tricia.

(b) There were a number of full mark responses to this question. Most common problems included that the concept is reductionist as it does not take into account how members can be more than one team role or that some companies may be too small, thus the company may not be able to fulfil the nine roles. Weaker responses could identify a problem but did not link it with an example from Belbin's team roles. It would have been helpful to use an example to illustrate the problem e.g., a company assigns person A as a plant, but person A may be both a plant and an implementer. Some responses attempted to identify a problem such as stating it was culturally biased or deterministic with no explanation or example and therefore were not creditworthy.

Question 16

(a) There was a range of responses to this question covering the full range of the mark bands. There were some good responses outlining extrinsic motivators at work, and Deci and Ryan's self-determination theory. Level 3 responses were able to identify and outline several types of extrinsic motivators as well as the three needs from self-determination theory, often with reference to the Landry et al. study. Level 1 and 2 responses gave fewer examples of the extrinsic motivators

identified and either would just identify the three needs or give brief definitions for each. Very weak responses would often just identify (and sometimes outline) one or two extrinsic motivators.

(b) The marks for this question tended to be between Level 1 and Level 3. Those that demonstrated good knowledge of the theories in part **(a)** achieved higher marks in this question. Most attempted the named issue of individual and situational explanations. Higher mark responses gave specific examples of how extrinsic motivators and Deci and Ryan's self-determination theory support individual or situational explanations. Level 4 and above responses, of which there were a few, were able to explain how both offered both individual and situational explanations. Other commonly used evaluation points included reductionism versus holism, application to everyday life, and idiographic versus nomothetic.

Weaker responses lacked evaluation points that provided any depth of explanation. Each point tended to be very brief, and reference was frequently made to what had been outlined in part **(a)**, but this was merely identifying which side of the debate extrinsic motivators and Deci and Ryan's self-determination theory was on or whether it did have applications but only occasionally was this explained. Sometimes analysis might be given but this was also just identified and not explained. Frequently this was structured by one paragraph for extrinsic motivators and the second paragraph for Deci and Ryan's self-determination theory. This led to a lot of repetition and no analysis was given. Non-creditworthy responses just continued to outline extrinsic motivators and Deci and Ryan's self-determination theory without any evaluation.

PSYCHOLOGY

Paper 9990/32

Specialist Options: Approaches, Issues and Debates

Key messages

Questions 1, 3(a), 5, 7(a)(i), 7 (a)(ii), 9, 11(a), 13(a), 13(b), 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, 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 the theory/technique/self-report 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 could ask the candidate to outline a theory, study, technique/treatment or self-report used by psychologists that is named in the syllabus or outline one of the issues and debates, possibly with an example from the syllabus content. 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 two 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 two studies or a key study, theories, explanations of a disorder identified in the specification under the appropriate bullet point. For this exam, some of the answers used the incorrect topic area in the syllabus or the description was brief. 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, 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, 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/explanation rather than by the issue which often led the response to be quite superficial and repetitive. 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 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. Responses that achieved 3–4 marks focussed on identifying triggers of pyromania, teaching muscle relaxation, imagining or visualising the process, walking away from it, and then practising this process using recordings/doing homework. Weaker responses were often brief and only stated that Fazli should imagine walking away from the fire setting situation so achieved fewer marks. A common error was mistaking imaginal desensitisation with covert sensitisation, negative images, disapproval or shame which was not creditworthy.

Question 2

(a) There were many good responses that were able to give a clear and concise outline of systematic desensitisation. Many clearly outlined fear hierarchy (from low to high fearful stimuli) and gave an example of this hierarchy for a phobia; plus, a mention of relaxation. Some responses referred to reciprocal inhibition with a few outlining what this means. Weaker responses sometimes just identified the fear hierarchy without any further details of how it would be used to treat a specific phobia. A common error made was to outline covert sensitisation or applied muscle tension rather than systematic desensitisation which was not creditworthy.

(b) Many responses were able to achieve one mark for this question by outlining what is meant by situational. Those that achieved full marks defined 'situational' because of environmental factors and gave the example of how exposing the patient to the phobic stimulus in the hierarchy involved putting the patient into a situation with their phobia in order to help them to overcome it. A typical error was to focus on imagination and/or muscle relaxation which was repeating the answer from **Question 1** and was not creditworthy.

Question 3

(a) Responses to this question covered the full range of the mark scheme. Those that achieved 3–4 marks correctly identified id/ego conflict, explained the defence mechanism of repression/displacement, and were able to explain the Electra complex to some degree, always connecting it back to Lina's snake phobia. Weaker responses showed some level of understanding, but often with less detail on how it links to Lina's case. Many referred to the Little Hans study but failed to extrapolate to Lina. Marks tended to be given for conflict between id and ego, reference to displacement of fear of mother to fear of the snake. A significant number of responses outlined a behavioural explanation of Lina's phobia such as Lina's mother screaming at the sight of snake causing Lina to become scared which was not creditworthy.

(b) There were some full mark responses to this question. The most common weaknesses given were a lack of scientific/empirical evidence, lack of generalisability from the Little Hans study and the difficulty in measuring Lina's unconscious conflicts and defence mechanisms due to relying on subjective interpretation. Weaker responses tended to be brief and just gave a brief outline of a weakness from one of the ones just mentioned without any explanation. Those candidates who were unable to gain credit in **part (a)** often did not gain any credit in this part of the question. A significant number of responses explained that it is reductionist (which it is not) by simply stating it ignores biological or behavioural explanation, which was not creditworthy. Other responses stated that a weakness was that the explanation is deterministic, without explaining why determinism is problematic. Arguably determinism is usually seen as a strength because it allows us to make predictions.

Question 4

(a) Responses varied for this question and covered the full range of the marks available. Good responses outlined included Beck's cognitive triad, negative schemas, automatic thinking, processing errors, faulty reasoning (with examples), and provided an explanation of how they all come together to produce depression. There were also some good examples given of cognitive errors and the three parts of the triad. Good responses around learned helplessness included clear explanation of loss of autonomy, sense of powerlessness or 'giving up' in face of repeated negative events and included attributional styles and good explanation of what these mean and how they contribute to produce depression. Many also included a brief outline of Seligman's study correlating attribution styles with scores on the Beck Depression Inventory.

Weaker responses often lacked detail for one of the explanations or both (more common in Level 1 responses). Many responses gave a brief outline of Beck's cognitive triad (sometimes with no link to the symptoms of depression) and then just stated that learned helplessness was a person facing a difficult experience and then 'giving up'. These types of responses achieved Level 1. For learned helplessness there were many candidates describing the Seligman study on dogs, which was not creditworthy. In addition, for Beck many outlined features of the Beck Depression Inventory which also was not creditworthy.

(b) The marks for this part of the question did cover the full range of marks available with the most frequent levels awarded being Level 2 and 3. Those that achieved Level 3 and above structured their response issue by issue and often started with the named issue **part (a)** of determinism versus free-will, along with clear and specific examples from the explanations and analysis. Apart from the named issues, other popular issues covered were application to everyday life, individual versus situational and reductionism versus holism. Popular examples for the named issue included how Beck's model allows for change via CBT (suggesting free will), while learned helplessness implies passivity and determinism. Level 4 and 5 responses were able to explain how both explanations had elements of determinism and free-will with clear contextualised examples for both sides or the argument.

Weaker responses achieving Level 1 or Level 2 did not contextualise their response. While some of the evaluation points were valid, the lack of context prevented responses from achieving a higher band. There were several generic definitions and a very weak attempt to include context from one of the explanations given in **part (a)** or why the issues apply to this topic area. Responses that evaluated the Beck Depression Inventory or Seligman's study on dogs were not creditworthy.

Consumer Psychology

Question 5

There were several good responses to this question, and some achieved full marks. Full mark responses were able to outline two suggestions that Company B's staff could using competitor-focused sales techniques to increase the sales of its bicycles. Full mark responses outlined a technique such as lowering prices and/or additional features on the bicycles compared to the competition and then briefly outlined how this would increase sales. Weaker responses gave one or more correct suggestions but either did not explain how this could be done or the effect of bicycle sales at Company B. Some responses failed to refer to the competitive-focused sales technique and instead their suggestion could just as easily have been about a customer-focused sales technique so was not creditworthy. Some responses were too vague to receive credit such as 'make the product better'.

Question 6

(a) Full mark responses showed they had a clear understanding of what is meant by 'post-purchase cognitive dissonance'. These types of responses often clearly defined the concept with an example of what this might look like in the real world. For example, 'the customer cares about the environment but bought a car that is not fuel efficient' or explanation of what the conflict is such as 'believing a certain way however acted differently during the purchase that is against their belief'. Many could correctly define post-purchase cognitive dissonance, though some gave overly simplistic definitions such as 'buyer regret' without discussing conflicting cognitions or beliefs which achieved one mark. Non creditworthy responses included outlining cognitive dissonance as 'expectation, and/or 'satisfaction'.

(b) Some responses were able to gain full marks. Some were able to explain one way post-purchase cognitive dissonance could be reduced. Good responses identified that either the seller can do X, or the consumer can do Y with a clear suggestion of why this would help reduce post-purchase cognitive dissonance. Most chose a consumer-related way such as looking at reviews or exaggerating good aspects of the product after purchase. Weaker responses just outlined a way but omitted the 'how' or 'why' it would reduce post-purchase cognitive dissonance. Those that did not receive credit often did not know what the term meant which was evident in **part (a)** of the response.

Question 7

(a) (i) Some responses were able to achieve full marks by identifying either utility theory or the take-the best heuristic and provided an example of how this can be seen in the stem based on what Steven says. Weaker responses often just identified utility theory but did not justify from the stem. Common errors included giving unrelated methods such as compensatory and non-compensatory which were not creditworthy.

(ii) Strong answers selected satisficing or anchoring theory and then justified these using the idea of buying the first suitable product found or the first piece of information being the most important. Better candidates included website placement or ranking in search engines as relevant factors

which clearly linked to the stem. Weaker responses sometimes gave a definition with an incorrect explanation, e.g., identifying anchoring correctly but giving justification for satisficing which was partially creditworthy. Many identified the theory but did not justify it. Non-creditworthy responses either identified the incorrect theory/model or gave a heuristic which was not creditworthy.

(b) Good full mark responses were able to give an outline of one problem psychologists have when deciding which model of consumer decision-making is being used by customers. Stronger responses acknowledged that consumer decisions may involve more than one model or noted the lack of a valid way of measuring insight into internal decision processes. Use of examples (e.g., product returns or mixed strategies) was common in better responses. Weaker responses gave a basic outline of the problem with no reason why it is a problem. Poor responses were generally too vague or made little sense to the question and seemed to just be guesses. A significant number of candidates did not respond, with 13 per cent of candidates providing no attempt to answer this question.

Question 8

(a) There were some good, Level 3 responses to this question. These types of responses included an outline of a study on any type of advertising media, and Lauterborn's 4 Cs marketing mix model. Commonly chosen studies on advertising in media were Ciceri et al. or Auty and Lewis. Descriptions of Lauterborn's 4Cs were often well done, especially where examples of each 'C' were provided. Responses that described Snyder and DeBono did not usually achieve Level 3 as the outline was not clear, giving vague details of the study. Weaker responses for Lauterborn's 4 Cs often either just identified and outlined 1 or 2 of the Cs or identified all 4 Cs with no outline given. A significant number of candidates simply made up a study (clearly believing that is what the question was asking for). This is not creditworthy as such a 'design a study' question would be found on Paper 4.

(b) There were some Level 3 and above responses to this question. Some did evaluate using the named issue of validity and were able to explain why the study given in **part (a)** had good validity and then poor validity (with contextualised example(s) from the study). These types of responses were credited with providing analysis. Other common evaluative issues that scored higher level marks were ecological validity, generalisations, applications to everyday life and reliability. Generally, candidates tended to evaluate the study and the 4Cs separately, but the more able candidates were able to combine them particularly with an issue like generalisations.

Many responses for this question achieved Level 1/Level 2. This was due to writing superficial evaluation points. The issue was named and the 4 Cs or the study was either stated to support the issue or which side of the debate it supported. This question had several very brief responses often with just a few sentences within which numerous issues were identified with little else provided. It was also noted in a significant number of candidates did not attempt the question with 19 per cent leaving the answer space blank.

Health Psychology

Question 9

There were some strong responses to this question achieving 3–4 marks. Good responses gave a detailed answer with clear understanding of how stress inoculation therapy could help to reduce Isha's stress at work. They recognised the 3 steps of SIT and expanded on it through included 'conceptualisation – discovering the issues', 'skill acquisition- finding coping mechanisms to work through them during stressful situations' and finally 'practice in the stressful environment.' Some used examples such as helping Isha to cope better with work deadlines or the quality of her work which gained the 4th mark. Weaker responses identified the three SIT stages (conceptualisation, skills acquisition, application) but did not explain these or apply them to Isha's stress at work. Several responses failed to understand this question and therefore gave general ways Isha could reduce stress at work which was not creditworthy.

Question 10

(a) Full mark responses identified a clear delay type (e.g., appraisal delay) and explained this using specific symptoms or examples (e.g., headaches not recognised as illness). Common reasons for delay were utilisation delay, appraisal delay and barriers to adherence usually citing cost. Basic responses (worth one mark) often provided a brief response and said 'cost' or 'fear of treatment'

with no further reasoning. Some were able to give a correct reason for delay, but explanation/example did not match so the second mark could not be achieved.

(b) There were some good responses to this question. Good responses connected their answer to **part (a)** correctly and provided a reasonable explanation related to it. These often were able to compare cultural groups to provide explanation and show understanding. The best linked to specifics e.g. how free health insurance in some countries means that cost is not a barrier to seeking treatment. A significant number of weaker responses gave a basic problem, lacking detail of how the reason can/cannot be applied to different cultures. Non-creditworthy responses gave strange reasoning which did not make sense, or which were too vague, e.g., 'No it wouldn't apply to other cultures because other cultures have different systems.'

Question 11

(a) There were several good responses to this question achieving 3–4 marks. Good responses identified doctor or direct style of consultation in contrast to shared style and described what these meant and with a study backing them up. The results of the Savage and Armstrong study were commonly referred to in good answers. Highest marks were given to those who remained close to the question, i.e., explaining how these changes would impact satisfaction. Weaker responses were vague or only described changes in communication without any more specifics. There was also often no connection made to patient satisfaction which limited the marks awarded. A common, non-creditworthy response was a suggestion that patients fill out a form so that Dr Brown could establish what style of doctor consultation they prefer which did not address the question.

(b) Most responses attempted to explain a weakness of using a questionnaire to measure satisfaction. Good responses identified self-report questionnaires as having weaknesses due to bias, social desirability, lack of reasoning/qualitative data, and then explained why this impacts the doctor's ability to successfully assess satisfaction. Weaker responses often failed to link to satisfaction so could not achieve the 2nd mark. Some responses gained fewer marks due to a confusion between demand characteristics and social desirability. Non-creditworthy responses often just identified an issue with no explanation.

Question 12

(a) There were a number of responses that gave clear details of a study on promoting healthy eating in schools and a study on promoting health and safety in worksites. Tapper et al. and Fox et al. were described by the vast majority. Good Level 3 responses often outlined the sample, procedural details and at least one result. Most of these covered both studies and referred to the school/workplace contexts clearly. Overwriting was evident in several responses, mostly giving many but correct details of the three studies by Tapper et al. Many candidates in this category did not finish the exam, probably because they spent so much time describing the study. Weaker responses often gave fewer details of one or both studies or provided some inaccurate details. Responses that did not receive credit did not include any accurate information about either study or, at worst, simply made studies up. A few attempted to design their own study, suggesting a lack of preparation for this type of question (which is frequently asked).

(b) There were a small number of responses that could achieve Level 3 or above. These types of responses started with the named issue which was experiments and were able to give some strengths and weaknesses. Responses that achieved Level 3 and above was due to clear detailed reference to the studies from **part (a)**. The strongest responses also included some analysis. This was rarely well explained and therefore often Level 3 was awarded. If the analysis was explained, such as explaining why you are stating that the two studies are similar due to controls used during the study, then Level 4 and above could be awarded. Other common evaluation issues included longitudinal studies, generalisability, ethics, determinism and application to everyday life.

There were a significant number of weak responses to this question where many received a Level 1 or 2. This was due to the response including superficial or basic evaluation. This frequently involved candidates identifying several issues and then stating, for example, how each study has good ecological validity. Weaker responses outlined a number of issues but in a superficial manner, simply stating that issues did or did not apply to studies but not why. These types of responses often achieved the Level 1 mark band.

If no creditworthy studies were described in **part (a)**, it was common for no marks to be awarded in this part of the question.

Organisational Psychology

Question 13

(a) Good full mark responses identified a correct need from McClelland's theory of achievement motivation and then provided an explanation as to why Jin was meeting this need linked to the stem. Responses that identified achievement were better able to outline how Jin was meeting this need rather than those who identified power. Those who identified power often stated that it was being met as Jin was allowing the workers to take responsibility for a project from start to finish or just stated that the workers had power over others, but in this context the workers did not have any power over other workers. Instead, this need is about having influence over the outcomes within the organisation. The workers need for this is met by Jin as due to their responsibility over the project they are influencing its outcome. Weaker responses often did not link the need identified to the stem. Responses that did not achieve marks were unable to identify a correct need from McClelland's theory.

(b) Good full mark responses identified the need for affiliation from McClelland's theory of achievement motivation and then provided an explanation as to why Kasem was meeting this need linked to the stem. More candidates were able to achieve marks for this question than for the **part (a)** and more made a clear link to the stem. Weaker responses were similar to those in **part (a)** as these often did not link the need identified to the stem. Responses that did not achieve marks were unable to identify the need for affiliation from McClelland's theory.

Question 14

(a) There were some good, full mark responses to this question. These were able to give a clear definition of what is meant by 'valence' from Vroom's VIE theory of motivation. Better responses defined valence as the value placed on rewards and explained how this affects motivation to complete tasks. Some used examples such as pay rises or promotions. Weaker responses often were able to give a basic outline of valence as how much a worker values a reward but did not link this to motivation. Responses that did not receive credit often outlined what is meant by motivation rather than valence. There were a significant number of candidates who did not know what the term meant so wrote incorrect or vague responses. 11 per cent of candidates did not attempt this question.

(b) The responses to this question varied quite a bit with some giving a clear explanation of one reason why Vroom's VIE theory of motivation could be considered holistic. Strong responses explained VIE as holistic as it considers motivation to be affected by three factors (valence, expectancy and instrumentality). There were some good definitions of holism, but these were rare, usually mentioning looking at the whole person. Weaker responses that achieved one mark frequently stated that the theory considered three factors which were then identified but no further explanation was given. A significant number of responses failed to show why this theory is holistic. It was rare for candidates to know what holism was (or at least be able to define it). 12 per cent of candidates did not attempt to answer this question.

Question 15

(a) The marks received by candidates for this question covered the full range of the mark scheme. Those who achieved 3–4 marks gave two clear suggestions each linked to how Lucy could provide her workers with 'motivators' in her restaurant. Good responses identified motivators clearly and then contextualised them to the restaurant, e.g., complimenting/praising chefs or waiters for their service, giving more responsibility to chefs creating new recipes, worker of the month or giving staff at the restaurant the opportunity to learn new skills (often suggesting that the staff work in other areas of the restaurant).

Weaker responses could suggest one or two ways for Lucy to provide her workers with motivators but frequently just identified it (e.g., worker of the month/praising workers) which was a basic suggestion and was awarded one mark per way. Some responses did not put either of the suggestions into the context of a restaurant so were unable to achieve full marks. A significant

number of responses suggested increased pay or bonuses, which are hygiene factors rather than motivators. These types of responses were not creditworthy.

(b) There were several good, full mark responses to this question that were able to provide a clear explanation of one weakness of Hertzberg's two factor theory of job satisfaction. Common weaknesses described in many stronger responses included culturally biased, not generalisable to different organisations and does not consider individual differences. Weaker responses were often brief such as stating that the theory does not consider individual differences as workers are motivated by different things which was a basic explanation of one weakness so achieved one mark. Many candidates did not provide a coherent weakness or did not seem to know Herzberg's two factor theory of job satisfaction, so could not provide a correct weakness and were not awarded any marks. 11 per cent of candidates who did answer all or most of the other organisation questions, did not attempt to answer this question.

Question 16

(a) There was a range of responses to this question covering the full range of the mark bands. Many achieved Level 3 by describing one (or more) of the hypotheses, the sample, details of the procedure including the conditions and at least one result. Very few responses gave details about the self-report/data collection tool. Those that did describe it identified that it had 14 adjectives and some gave examples of the adjectives, but these were rare. Weaker, Level 2 or Level 1 responses included fewer details of the study. These responses often described a hypothesis, the sample and a result. It was common for weaker responses to omit any procedural/task details or provide a muddled or inaccurate description. Some of these responses incorrectly stated that the study was a field study and then outlined the procedure as the participants observing in a hospital rather than reading a passage. A significant number of candidates were not prepared for this question, instead giving some background of masculine and feminine leadership styles without any reference to the study. 10 per cent of candidates did not attempt this question.

(b) The marks for this question tended to be between Level 1 and Level 3. Those that demonstrated good knowledge of the study in **part (a)** achieved higher marks in this question. Most attempted the named issue of self-reports. Better responses that achieved Level 4 and above for this issue were able to give strength(s) and weakness(es) of self-reports and provided clear examples of the self-reports used in the study by Cuadrado et al. Many candidates had not outlined any of the self-reports used in the study in **part (a)** so although they were able to evaluate self-reports, their response lacked context and therefore achieved either Level 1 or Level 2. Other commonly used evaluation points included ecological validity, laboratory setting, generalisability and application to everyday life.

Weaker responses often identified an issue and attempted to give examples and mainly just described the content from **part (a)**. Quite a significant number of responses tended to briefly outline strengths and/or weaknesses of the study, rather than structuring their response issue by issue. This frequently meant that each point was not developed and there was some repetition of issues. A significant number of responses gave details of the Cuadrado et al. study that they had not included in **part (a)** (and therefore were not credited) in this part of their answer.

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 study in **part (a),** but this was merely identifying which side of the debate the theory was on or whether it did have applications or good generalisability but only occasionally was this explained. Sometimes analysis might be given but this was also just identified and not explained. If the candidate did not know the Cuadrado et al. study and achieved no marks in **part (a)**, they often achieved no marks or occasionally one mark for guessing at some correct evaluation point(s) in this part of their answer.

PSYCHOLOGY

Paper 9990/33

Specialist Options: Approaches, Issues and Debates

Key messages

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

These questions in this exam asked candidates to apply an area of the syllabus (theory, diagnostic criteria, technique/treatment, self-report, etc.) to explain how it is relevant 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 the suggestion such as the theory/technique that was outlined in the candidate's response to part **(a)** of the question. In this exam, these types of questions asked the candidate to evaluate the theory/technique outlined in part **(a)** such as with a problem with the theory/technique and explain why this is a problem in the scenario given in the stem of the question. It would be helpful for candidates when doing revision to learn strengths and weaknesses/problems of the theories, techniques, using diagnostic criteria, treatments, etc. that they have learned and put these into their revision notes. They should also practice explaining the evaluation point in the context of the question.

Questions 2, **6**, **10 and 14**

Part (a)

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

Part (b)

This part of 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 might ask candidates to explain a strength or a weakness of the issue/debate or the syllabus content outlined in part **(a)**. It would 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 question. 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 two 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 two treatments, two theories (possibly with reference to research that investigated the theory), and a key study identified in the specification under the appropriate bullet point. For this exam, some of the answers used the incorrect topic area in the syllabus or the description was brief. 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. 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 study/studies, theories, 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, in order 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 study/studies, theories, 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, 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. Several of the responses did do analysis. Candidates should be aware this question is worth 10 marks and they 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 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. Responses that achieved 3-4 marks were able to explain the different components of CBT such as creating a therapeutic alliance, identifying faulty thinking, reality testing (often with an example of either a delusion or a hallucination), keeping a diary and an example of homework that Mahmoud could do outside of sessions. Weaker responses were sometimes brief and just identified that CBT can challenge thinking without any reference to schizophrenia and how CBT can assist with positive symptoms. A small number of responses outlined how CBT could help with a negative symptom which was not creditworthy. Some responses included a detailed outline of the symptoms of

schizophrenia which was not creditworthy. These responses often then outlined how CBT could be applied to a positive symptom, but it was often very brief.

Question 2

(a) There were some good responses that were able to clearly outline Miller's feeling-state theory of impulse control disorders. Full mark responses outlined the different components of the feeling state and that the memory of the positive feelings lead to the individual wanting to perform the impulsive behaviour again. Other responses outlined an example of the feeling-state for a specific impulse control disorder such as gambling or pyromania which was also creditworthy. Weaker, 1-mark responses outlined a link to positive feelings and specific behaviours (e.g., gambling). There were also many responses that focused on the negative rather than the positive experience that are felt when either engaging in or after the compulsive behaviour which was not creditworthy.

(b) Some responses were able to achieve one mark for this question by explaining one reason why Miller's feeling-state theory could be considered holistic. These often explained that it was due to the theory considering emotions, physiological arousal and cognition (memory) to explain impulse control disorder. A small number of responses achieved the second mark for providing a definition of holism with reference to it being about the bigger picture or how components work together to cause behaviour. Credit was given to genetic definitions or holism or in the context of Miller's feeling-state theory. Candidates found it difficult to define holism and frequently stated that it involved multiple factors which was not creditworthy. In addition, many were unable to explain one reason why Miller's feeling-state theory could be considered holistic. These types of responses often incorrectly identified the different parts of the feeling state or just identified one or two.

Question 3

(a) (i) Most responses were able to achieve at least one mark by suggesting one reason why Emi may be diagnosed with pyromania. This was achieved by citing an example from the stem such as feeling better after setting a fire, fascination with fire or feeling excited when thinking about starting a fire. Responses that scored full marks were able to relate the example to one of the diagnostic criteria in the ICD-11. It was not a requirement of the question to specifically identify the ICD-11 but the candidate needed to make it clear how the example given of Emi's behaviour/thoughts fit a symptom/diagnostic criterion of pyromania. A few responses did not receive credit, and this was often because they made a vague link to the stem such as 'she felt anxious about the situation'.

(ii) Those responses that received credit either referred to the fact that the factory fire had been done out of revenge or that Emi had only set one (or two, including the matches) fire. A few candidates linked this to ICD-11 criteria that there is no motive for setting the fire or there are multiple acts of fire-setting. Some responses suggested that Emi had not had these impulses/fire-setting behaviour for very long which was not creditworthy as the ICD-11 does not state a time period for the symptoms. Some responses suggested that she would not be diagnosed with pyromania because either Emi did not act on her thoughts to start a fire when she was in a new job or that it has not impacted her life significantly enough. Neither of these are in the diagnostic criteria so these types of responses were not creditworthy.

(b) There were some full mark responses to this question. Common problems included Emi being dishonest with the doctor about the fire setting due to legal reasons or comorbidity with other disorders such as anxiety. Full mark responses gave a clear link to pyromania whereas the weaker, 1-mark responses outlined a generic problem. Some responses just repeated the question in their answer e.g. Emi does not meet all of the diagnostic criteria, which was not creditworthy.

Question 4

(a) Responses varied for this question and covered the full range of the marks available. Level 3 responses were less common with most responses achieving either Level 1 or Level 2. Those that did achieve Level 3 kept their response focussed on SSRIs and exposure and response prevention (ERP) as a treatment/way to manage obsessive-compulsive disorder (OCD). Level 3 descriptions of SSRIs went beyond that it just blocks the re-uptake of serotonin. These responses were able to fully describe that it blocks the reuptake into the presynaptic neuron therefore leaving more serotonin available in the synaptic gap. They were also able to talk about increasing serotonin receptors and then finally linking it to specific symptoms of OCD. ERP was often described in more depth and with more accuracy of the steps involved in this therapy and how this would lead to a

reduction in obsession(s) and compulsion(s) of the client with OCD. Many responses included an outline of Lehmkuhl et al.'s study on Jason with a focus on ERP and Rapaport's study on Charles with the focus on the SSRI used and its effect on Charles.

Weaker responses showed a poor understanding of SSRIs and how they work to help with symptoms of OCD. Some included dopamine and oxytocin which SSRIs do have an indirect effect on but that is not the main neurotransmitter and neither does it link to OCD, so this was incorrect. There were a number of incorrect responses that attempted to describe what 'blocking the re-uptake of serotonin' means. Many stated it was blocked in the brain or in the post-synaptic neuron which was not creditworthy. Level 1 responses were often very brief for both SSRIs and ERP. Some responses failed to provide a clear outline of ERP and gave a generic description of talking therapy or mixed it up with other treatments such as systematic desensitisation (linked to phobias). Neither of these types of responses were creditworthy.

(b) The marks for this part of the question did cover the full range of marks available with the most frequent levels awarded being Level 1 and 2. The responses that did achieve Level 3 and above structured their response issue by issue and often started with the named issue of the use of children in research, along with clear and specific examples from either Lehmkuhl et al.'s study on Jason and/or Rapaport's study on Charles and analysis. Frequent points raised regarding the use of children included application to everyday life, lack of demand characteristics, communication issues and ethics as both a strength and a weakness. Good responses were able to discuss how the research on children is ethical as consent was obtained from parents and the child's OCD symptoms reduce. Weak responses stated that the studies were unethical as consent is not given by the child. This is not a creditworthy point as consent from the parent makes the study ethical and, depending on the age of the child, consent could be given by them as well. In addition, many attempted to argue that both ERP and SSRIs are harmful. Credit was given to points regarding the side effects of SSRIs as harmful to children. If the candidate stated that ERP is harmful as it causes anxiety; this was only credited if it was made clear in the response that the anxiety was short-term and the therapy helped reduce anxiety in the long term due to treating the child's OCD symptoms. Apart from the named issue, other popular issues covered were determinism versus free-will, situational versus individual, generalisability, idiographic and nomothetic, application to everyday life and reductionism versus holism.

Weaker responses achieving Level 1 or Level 2 did not contextualise their response. While some of the evaluation points were valid, the lack of context prevented responses from achieving a higher band. There were several generic definitions and a very weak attempt to include context by identifying either SSRIs or ERP as the example. Some responses took the approach to evaluate the two treatments one after the other. This usually resulted in very superficial points with a lack of context. For example, the candidate would state that SSRIs are unethical as they have side effects without giving any examples or explaining that side effects will lessen the longer the patient takes them for. It was also very common for responses to state that SSRIs reduce symptoms immediately which is incorrect as they will take a few weeks to start to reduce symptoms of OCD.

Consumer Psychology

Question 5

There were a few good responses to this question, and some achieved full marks. Full mark responses were able to suggest two ways that the shop could change their virtual store layout that could reduce customer complaints. Strong responses were often concise. Grid layout was seen as easiest and freeform as most entertaining. There was a clear link to reducing complaints, particularly when referring to the current racetrack layout. Weaker responses did not relate their suggestion to reducing complaints. Some responses mixed up the effect of a grid layout for a freeform layout or vice versa which was not creditworthy.

Question 6

(a) There were a number of full mark responses that gave a clear definition of free-will and linked this to how system 2 uses free-will. Most were able to outline a definition of free-will. Some candidates gave a vague definition of free-will such as 'being able to express yourself' which was not given credit. In addition, many responses could not give an example of how system 2 uses free-will although some of these did understand what system 2 thinking involves. There were some responses that compared system 2 thinking to system 1 which did not address the question and therefore was not creditworthy.

(b) Some responses were able to gain full marks by explaining one problem a psychologist may have when they investigate whether consumers use free-will when deciding to purchase products. A common response was that it is difficult to see/measure what the consumer is thinking and determine whether they are using free-will. In addition, there may be external factors such as advertising/promotions/packaging that could influence the consumer's choice without the consumer being aware of this. Therefore, the consumer would tell the psychologist that they used free-will when in fact the purchase was determined by external factors. Few responses were able to relate the problem to purchasing a product to achieve the second mark. Some responses explained that external factors could affect purchase without linking this to a problem that a psychologist may have when investigating free-will. These types of responses were not creditworthy.

Question 7

(a) (i) Some responses were able to achieve full marks by identifying the anchoring heuristic and then provided an example from the stem on how Sally used this heuristic when making the decision of which running shoes to purchase. These responses often stated Sally was influenced the most by the first piece of information she received. Weaker responses could not explain from the stem how Sally used the anchoring heuristic. Many candidates simply mentioned it was the first item she saw without explaining how the anchoring heuristic comes into effect. A significant number of candidates identified the incorrect heuristic or restated what had already been outlined in the stem e.g., Sally bought the shoes from the first shop she went in or she bought them as they were first. These types of responses were not creditworthy.

(ii) Compared to part **(a)(i)**, a higher proportion of responses correctly identified the recognition heuristic and referred to Dan's brand loyalty due to familiarity, knowing the brand well or trusting the brand due to past experience to achieve full marks. Weaker/non-creditworthy responses often provided an incorrect heuristic such as 'take-the-best' and/or were not able to explain from the stem how Dan used the recognition heuristic. Many candidates simply mentioned he knew the running shoe brand but did not elaborate about Dan trusting the brand/long history with the brand. These types of responses were not creditworthy.

(b) A few responses were able to achieve one or two marks for this question. Creditworthy problems outlined included that consumers may use multiple heuristics or it is difficult because researchers do not know what a consumer is thinking. Some explained that an interview or questionnaire would need to be used, and this might lead to invalid results as consumers may lie/produce socially desirable answers. Alternatively, consumers may not be aware of which heuristic they used as these are mental shortcuts and are often unconscious. The best responses were able to give an example in the context of deciding on a product to purchase. Some responses used an example from the STEM, although this was not required for full marks.

Weaker responses frequently gave a brief outline of a problem without any reference to choice heuristics used when purchasing a product so could only achieve one mark. Some responses attempted to explain that other variables might influence the choice of product such as pricing/advertising. These types of responses were not creditworthy as consumers still use heuristics with these factors. There was a poor understanding in many responses of what is meant by a choice heuristic which meant that they failed to achieve any marks. 15 per cent of candidates did not attempt to answer this question and often achieved 0 marks for their responses to part **(a)(i)** and **(a)(ii)**.

Question 8

(a) There were a few Level 3 responses to this question. Responses that obtained this mark band were able to describe example studies to support their response. The best described clearly the Porublev et al. study on gift-wrapping. Atalay et al. on attention and shelf position was sometimes well-described by candidates who described one of the three studies. The best responses had an outline of participants, procedure, and the results from the example studies. Some responses outlined either the Howard study or the Rixom et al. study on gift-wrapping. However, candidates who described one of these studies frequently had very limited knowledge of the study and either gave a result or a conclusion from the study which was awarded Level 1. Some responses attempted to outline the three studies done by Atalay et al. but frequently none of them were described in any depth and candidates often confused the procedure and result between the three studies.

There were a significant number of weaker responses that often gave anecdotal/generic descriptions of gift-wrapping including beliefs of giver and recipient. These included comments about buying and receiving presents and how we would like them wrapped or not wrapped. There was lots of inclusion of cultural differences for colour that had no real psychological evidence to back it up. These types of responses often achieved Level 1. Similarly, responses then attempted to outline what psychologists have discovered about attention and shelf position but did not know a relevant study so simply outlined where products should be placed on shelves in supermarkets and shops which was not creditworthy.

(b) There were a very small number of Level 3 and above responses to this question. Those candidates who had outlined at least one study in part **(a)** were able to discuss the named issue of generalisations from findings. Candidates often used cultural difference in their discussion about generalisations, mentioning that different cultures may wrap gifts differently thus traditional wrapping/non-traditional wrapping preferences may not be generalised across different cultures. Weaker responses for the named issue frequently lacked knowledge of the sample from the studies outlined in part **(a)** so just identified that the studies were done in the 'West' and therefore could not be applied to other cultures. These types of responses were awarded Level 1 for this issue as this is basic evaluation.

The best responses covered this question by structuring it by issues, along with supporting examples from the studies in part **(a)** and analysis. Apart from the named issue, other popular issues covered were deterministic versus free-will, application to everyday life, validity and idiographic versus nomothetic. Some responses attempted to use applications to real life but simply stated that the findings will 'improve sales' without elaborating how or providing evidence from the studies. Better responses were able to outline how stores should include gift wrapping services to improve customer service or put their best valued products at the middle of the shelf to improve profits.

Most responses for this question achieved in Level 1/Level 2. This was due to writing superficial evaluation points. The issue was named and either the theory or the study was either stated to support the issue or which side of the debate it supported. This question had several very brief responses often with just a few sentences within which numerous issues were identified with little else provided. It was also noted in a significant number of candidates who had chosen consumer as one of their options, did not attempt the question with 20 per cent leaving the answer space blank.

Health Psychology

Question 9

There were some strong responses to this question which achieved 3–4 marks. The best responses were able to outline what is meant by false positive and provide several reasons why a doctor might make a false positive diagnosis. Common suggestions included patients giving false information either deliberately or due to shyness/embarrassment, the patient might exaggerate their symptoms, or language barriers. Other creditworthy suggestions included problems of clinical tests such as blood tests, lack of time during consultation and lack of expertise by doctor. Some wrote about how the doctor's skill level caused them to misinterpret symptoms/medical readings which led to a false positive.

A few responses confused false positive diagnosis with a false negative, which was not creditworthy. However, responses that mentioned factors applicable to both types of diagnostic errors, such as patient misinformation or clinician inexperience, were awarded marks.

Weaker answers often lacked a definition or example of a false positive diagnosis or provided only one reason without elaboration. Some responses listed multiple similar points – such as 'patient lying,' 'patient exaggerating,' and 'patient providing false information' – which were considered a single idea and thus limited the marks awarded. Additionally, some candidates focused on why a patient might provide false information (e.g., due to conditions like Munchausen syndrome), which did not directly address the doctor's role in making a false positive diagnosis and therefore did not receive credit. Some responses suggested that that the doctor misdiagnosed the patient with a different disease, which is not a false positive diagnosis. This part of the response did not receive credit.

Question 10

(a) There were a number of full mark responses that accurately defined the nomothetic approach and then gave an example from a study on preferences for practitioner clothing. The vast majority used the McKinstry and Wang's study as their example, such as patients' preferences for male doctors in suits and ties and female doctors in white coats. Weaker responses often were able to define the nomothetic approach but unable to give an example from a study or gave an incorrect result.

Some candidates correctly stated that the nomothetic approach involves collecting quantitative data, earning partial credit. However, they often failed to link this to a concrete example from the study, limiting their marks. Merely stating that the study used closed questions without elaboration was too vague to be credited.

A number of responses lacked a clear definition of the nomothetic approach or confused it with other concepts like reductionism, which did not receive any marks. Additionally, some candidates described the study's aim or procedure without explaining how it exemplified a nomothetic approach. These examples from the study were not creditworthy.

(b) There were a few good responses to this question. The most common weakness given was that the nomothetic approach does not consider individual or cultural differences. Full mark responses frequently gave an example of a difference such as the older participants in the McKinstry and Wang study felt more strongly about the importance of dressing formally compared to younger participants. The candidate did not need to refer to a result in a study and some gave an example of a possible, specific cultural difference in preferences.

Some responses explained the weakness was due to a lack of in-depth/qualitative data. These often achieved 1 mark as the candidate was unable to elaborate or just briefly explained that we do not know the reason for the choice of practitioner clothing which was a basic outline of this weakness.

A common error was that the nomothetic approach uses qualitative data instead of quantitative and an outline of a weakness of qualitative data was given that did not receive credit. Some responses outlined a result from the McKinstry and Wang study but did not link this to the weakness identified. These type of responses were awarded one mark.

Question 11

(a) This question was well answered by candidates and many achieved full marks. Common suggestions included pill counting, TrackCap, and blood/urine tests. Many responses were able to achieve two marks for each suggestion by explaining how the measure works or how it will show that Dr Smith's patients are taking their medication correctly. Weaker, one-mark responses often just identified the method but were unable to suggest how it would measure patient adherence. Some responses suggested that a Trackcap can count the pills removed but this is incorrect as it tracks the time the bottle was opened. In addition, some candidates who suggested a blood or urine test did not explain that Dr Smith will need to check the results to see if their medication is present in the blood or urine so were unable to achieve full marks.

(b) There were a number of full mark responses to this question where the problem was explained and why this might lead Dr Smith to an inaccurate measure of his patients' adherence. Common responses for this question used pill counting and TrackCap. Most candidates explained that patients could possibly remove the pills on their own where neither method would be able to track this. Another common problem was that social desirability may take place when an interview is used. Candidates achieved full marks by explaining that this will mislead Dr Smith into believing patients have adhered to medicine where in reality they have not.

Weaker, one-mark responses were often brief and did not link the problem to Dr Smith measuring adherence of his patients and instead explained the problem the patient would have in measuring their own adherence which was not answering the question. Candidates found it difficult to explain a problem with blood or urine tests. Many attempted to explain that these tests are unethical which was not creditworthy. This is because Dr Smith can explain the reason for these tests and get informed consent before the medication is prescribed. He can explain how the blood/urine test will support the patient's health and show whether the medication is working.

Question 12

(a) There were a significant number of responses that achieved Level 3. Many were able to identify features of the sample (e.g., all women, all under 70 years old and had suffered breast cancer). These responses also identified the three conditions, gave some details of the procedure (e.g., what the participants did in the relaxation and imagery condition), one or more of the questionnaires used in the study and a result such as mood disturbance score improved the most for the relaxation plus imagery group. This was particularly found with women age 55 plus. Women who had high anger scores at the start of the study did not respond well to either relaxation or relaxation plus imagery.

Weaker responses gave fewer details of the study or gave incorrect information e.g., wrong number of participants/saying it was a repeated measures design. A significant number of responses thought that stress was measured and quoted that as a variable in the results section. If no creditworthy result was given in the response the maximum mark that the candidate could achieve was 5 marks.

Many of the responses described details of the study in part **(b)** that should have been included in this part of the question. This meant that these candidates were awarded a low mark despite knowing the study well.

A significant minority of responses appear to have used the title of the study to create a study. These responses sometimes achieved Level 1 as they could give a brief description of relaxation and/or imagery.

(b) There were a small number of responses that could achieve Level 3 or above. These types of responses started with the named issue, which was field experiments, and were able to give some strengths and weaknesses. Candidate responses achieved Level 3 and over due to giving clear and detailed reference to the Bridge et al. study from part **(a)**. The strongest responses also included some analysis. This was rarely well explained and therefore often Level 3 was awarded.

An example of analysis for field experiments was that this method has the strength of taking place in the natural environment which for this study was a hospital which was a very familiar for the participants as they had been receiving treatment for cancer in a hospital. In addition, the participants in the two intervention conditions were told to practice at home which also has good ecological validity. The women may find it easier to do relaxation while at home. However, a weakness of a field experiment is that they lack control. This study did have some controls in place such as being given standardised training in relaxation/imagery. However, it is possible that some of the participants have been doing other activities to manage their mood such as exercise, talking to a friend or family member, etc.

Other issues discussed included generalisability, individual versus situational, reliability, validity, strengths and weaknesses of questionnaires.

Level 1 and Level 2 responses did superficial evaluation/discussion of the key study. Evaluation points were often creditworthy but completely devoid of any context from part **(a)**. It was common for responses to cover 5–8 evaluation points where the candidate identified the issue and then stated whether it was a strength or a weakness of the study or which side of the debate the study/results from the Bridge et al. study supports without any example or explanation about why the study supports this side of the debate.

There were quite a few responses that attempted an application to everyday life issue. Though most responses could not give any specific clear application and how the application will be carried out. Most responses simply mentioned it can be done in hospitals or other patients with cancer without further elaboration.

24 per cent of candidates who answered the questions in the health option offered no response here or they evaluated a study that they had invented themselves which was not creditworthy.

Organisational Psychology

Question 13

Responses covered the full range of the mark scheme. Good responses were able to identify one or two of Heifetz's six principles in meeting adaptive challenges and link this to helping worker to adjust to the new/change in working hours, tiredness or mistakes made at work. Common principles included get on the balcony, regulate distress and give work back to the people. Weaker responses were often able to identify the principle but then did not link the principle to the stem. Many did not know the principles and therefore were unable to access the marks.

Question 14

(a) There were many good, full mark responses to this question. These were able to give a clear definition of the nature versus nurture debate. Weaker responses were usually when attempting to outline the nature side, some simply mentioned we are 'born that way' or gave a vague statement such as 'a natural occurrence that can't be changed' which were not creditworthy.

(b) There were some full mark responses to this question. These were able to outline one reason why Scouller's levels of leadership theory supports the nurture side of the nature versus nurture debate. For example, responses often explained how the leader interacts with people (publicly or privately) and learns through these interactions what leader characteristics work well with larger and/or smaller groups. These types of responses were able to outline a supporting answer in context with the 3 levels of leadership, outlining how it is learnt as we interact across the 3 levels. Weaker responses were able to explain why levels of leadership support nurture (e.g., leader interacts with others at the public level and learns from it) for 1 mark but struggled to access to second mark. There were several responses that did not receive any credit for this question. This was because they gave generic statements about how leadership is learnt with no mention made of Scouller's levels of leadership.

Question 15

(a) (i) Good, full mark responses were able to identify a relevant Belbin Team role and explain how Margot fits the role appropriately in the context of checking the leaflet. Completer-finisher and monitor-evaluator were the common team roles selected. Weaker responses could identify an appropriate team role but could not justify their response with a clear link to Margot checking the leaflet. Non-creditworthy responses identified an inappropriate team role or invented a team role which was not from Belbin for Margot.

(ii) Good, full mark responses were able to identify a relevant Belbin Team role and explain how Tricia fits the role appropriately in the context of designing the leaflet. Plant and implementer were the common team roles selected. Weaker responses could identify an appropriate team role but could not justify their response with a clear link to Tricia designing the leaflet. Non-creditworthy responses identified an inappropriate team role or invented a team role which was not from Belbin for Tricia.

(b) There were a number of full mark responses to this question. Most common problems included that the concept is reductionist as it does not take into account how members can be more than one team role or that some companies may be too small, thus the company may not be able to fulfil the nine roles. Weaker responses could identify a problem but did not link it with an example from Belbin's team roles. It would have been helpful to use an example to illustrate the problem e.g., a company assigns person A as a plant, but person A may be both a plant and an implementer. Some responses attempted to identify a problem such as stating it was culturally biased or deterministic with no explanation or example and therefore were not creditworthy.

Question 16

(a) There was a range of responses to this question covering the full range of the mark bands. There were some good responses outlining extrinsic motivators at work, and Deci and Ryan's self-determination theory. Level 3 responses were able to identify and outline several types of extrinsic motivators as well as the three needs from self-determination theory, often with reference to the Landry et al. study. Level 1 and 2 responses gave fewer examples of the extrinsic motivators

identified and either would just identify the three needs or give brief definitions for each. Very weak responses would often just identify (and sometimes outline) one or two extrinsic motivators.

(b) The marks for this question tended to be between Level 1 and Level 3. Those that demonstrated good knowledge of the theories in part **(a)** achieved higher marks in this question. Most attempted the named issue of individual and situational explanations. Higher mark responses gave specific examples of how extrinsic motivators and Deci and Ryan's self-determination theory support individual or situational explanations. Level 4 and above responses, of which there were a few, were able to explain how both offered both individual and situational explanations. Other commonly used evaluation points included reductionism versus holism, application to everyday life, and idiographic versus nomothetic.

Weaker responses lacked evaluation points that provided any depth of explanation. Each point tended to be very brief, and reference was frequently made to what had been outlined in part **(a)**, but this was merely identifying which side of the debate extrinsic motivators and Deci and Ryan's self-determination theory was on or whether it did have applications but only occasionally was this explained. Sometimes analysis might be given but this was also just identified and not explained. Frequently this was structured by one paragraph for extrinsic motivators and the second paragraph for Deci and Ryan's self-determination theory. This led to a lot of repetition and no analysis was given. Non-creditworthy responses just continued to outline extrinsic motivators and Deci and Ryan's self-determination theory without any evaluation.

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 A Level component. For example, at AS candidates learn about methodology, such as experiments, which also apply to A Level.
- 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 for '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 the heading instructions for each question section carefully.

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

Candidates are not required to re-write the question before beginning their answer.

Some candidates apply the terms 'demand characteristics' and 'social desirability' to every answer when in most cases these terms do not apply.

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 not always related to the question, and so restricted marks.

In question **part (c),** when answering the 'in this study', the 'relating/linking' part of the question, simply rewriting the words in the question is not sufficient and no marks can be awarded for it. This merely shows that candidates can copy a question. Candidates need to add more to show understanding of the study in question. See examples below, such as **Question 5a**.

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, 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 stated 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 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 are advised not to 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, which is an essential element of the plan.

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

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) Despite the question stating 'other than getting a phlebotomy' (which means that this will receive no credit because it is part of the question) many candidates overlooked the instruction. There were nine other situations included in the SUDS hierarchy candidates could use and any two of these nine received credit. Candidates often stated 'getting a finger prick' which was too vague to be awarded marks.

(ii) Most candidates could not provide an accurate result for any one of the nine possible situations. Candidates achieving full marks gave answers such as 'wife taking my blood pressure reduced from 50 before, to 45 during and 20 after'. Candidates were awarded 0 marks for stating results for 'getting a phlebotomy' which could not be used as this was a continuation of question **part (a)**.

(b) Candidates frequently were awarded 1 mark for stating 'T may not have been truthful' which is correct, but without stating what T was not being truthful about. Adding an additional comment like 'the SUDS score he gave for taking my own blood pressure' would link the point about being untruthful to the study and an additional mark would be awarded.

(c) Candidates could often give two strengths of quantitative data but rarely related the strengths to the study by Chapman and DeLapp and 'T'. To be awarded full marks, candidates must relate/link their strength (or weakness if applicable) to the study in question. Stating 'quantitative data is numerical' is not a strength; that quantitative data can be statistically analysed is.

Question 2

(a) To be awarded marks candidates needed to explain how irrational beliefs are changed by REBT. This means that the D and E (disputing and 'effects') of the Ellis ABCDE model were essential inclusions. Some candidates focused on the ABC, the theory, and this could achieve 1 mark. Many candidates merely stated that irrational beliefs are changed into rational ones, and this was awarded 0 marks because it showed no knowledge of Ellis's REBT.

(b) Many candidates stated 'I would give a questionnaire' without elaboration of technique or format. A2 requires more than just 'questionnaire'. Candidates who stated 'I would give a BDI' without elaboration were also awarded 0 marks. To achieve marks, candidates should state what a BDI is and give some detail about it (closed questions, 21 items, rating scale, etc.) and say how it would measure the effectiveness of REBT. Answers achieving top marks constructed their own questionnaire, stating the format and technique, gave an example of a question, and how this would be scored.

(c) The most common answer was that REBT does not involve biochemical treatments (1 mark) with many candidates going on to state, 'such as anti-depressants which can be addictive' and so earning one more mark. Some candidates gave the same strength again but with a different example, such as 'so there are no side effects'. However, the question states two strengths and so the strengths must be different.

Question 3

(a) (i) Very few candidates knew the questions that participants were asked in the key study by Robson et al. There were many incorrect responses, and some candidates knew a few relevant words. For example, candidates might know the word 'comfortable' but not be able to correctly place it in the question asked by Robson et al. which was 'Sitting at this table would make me feel comfortable'.

(ii) Very few candidates knew the rating scale that Robson et al. used. The questionnaire is a fundamental aspect of the key study. Marks were awarded for stating 'Likert-type scale', 'seven-point scale' and 'labels of strongly disagree (1) via 'neither agree nor disagree' to strongly agree (7)'.

(b) There were some excellent answers provided by candidates who could suggest an alternative way in which the table spacing could be studied. The strongest answers suggested a method and linked it to the measurement of table spacing. Many candidates either stated a method in detail but not what they were measuring, achieving 1 mark, or stated nothing more about the method than 'I would interview people' with no link to anything.

(c) Like other **part (c)** questions, candidates often gave a strength and weakness without relating it to the question. One strength is that a 7-point scale allows a wider range of responses than a 3-point of 5-point scale. However, most candidates did not relate this to the key study, when full marks could have been awarded by stating that 'Robson et al. used a 7-point scale' (strongly agree to strongly disagree) when asking 'Sitting at this table, I would feel comfortable'. Weaknesses mentioned were often good but were rarely related to the Robson et al. study.

Question 4

(a) All candidates were awarded 1 mark for their answers on brand awareness, but only those providing some elaboration in their answer were awarded the second available mark. Reference to brand name, brand logo or brand slogan is sufficient.

(b) Similar to answers to other questions, many suggestions included nothing more than 'I would do an observation/questionnaire/interview' with no elaboration about the format or technique. Marks could not be awarded for such answers. Candidates providing some elaboration were awarded 1 mark with those linking it to how the method would be used to investigate brand awareness being awarded full marks.

(c) The lack of inclusion about the features of different methods evident in **4(b)** was also evident here. The syllabus states 'describe *the main features* of an observation'. It also distinguishes between questionnaire/interview technique and format. This meant that candidates stating 'I would do an observation' struggled to provide a strength or weakness, and when this was not linked to brand awareness no marks were awarded.

Question 5

(a) There were seven different features of the sample that candidates could identify to be awarded marks, for example 1167 participants at the outset; 1038 at the end; aged 11 – 14 years; two middle schools; mostly Jewish. Most candidates could give at least one feature.

(ii) Many candidates suggested reasons why participants withdrew from the study. However, this was an AO1 question meaning that the only correct reasons for withdrawal were those stated in the original study. These were that participants withdrew due to absence (34 intervention, 22 control) and withdrew due to refusal to complete questionnaires (2 intervention, 4 control).

(b) Most candidates suggested 'to do a comparison/baseline' which was sufficient for 1 mark to be awarded. However, candidates could not elaborate beyond this. To clarify, the wait-list control condition meant that schools and appropriate candidates were 'ready and waiting' should the

selected school dropout and six school principals with similar features to the intervention school agreed to be on the wait-control list.

(c) Candidates frequently assumed that a longitudinal study automatically gathered both quantitative and qualitative data and that this data is always in-depth, which is incorrect. Participants might only have to complete a short, closed question questionnaire with only 5 items. No candidate referred to what data was gathered from the participants in this study at the 'one-year-later' follow up.

Question 6

(a) Identifying any two of the ten behavioural categories from the UAB pain behaviour scale earned 2 marks. Included in the ten categories were things like 'vocal complaints non-verbal' and 'facial grimaces'. Marks could be awarded for including examples that were listed on the UAB, such as 'groans, moans, gasps' but not for any example not on the UAB list.

(b) Candidates had to suggest a psychological treatment, such as attention diversion, non-pain imagery or cognitive redefinition which earned 1 mark. Biological or alternative treatments such as acupuncture were not creditworthy. In addition, candidates had to say how acute pain would be treated, and this aspect was only addressed by a very small number of candidates because they failed to answer the question set.

(c) Answers to this question often provided brief weaknesses such as 'the pain behaviour may be misinterpreted' with no elaboration. Answers like this will often be awarded 1 mark, but not more. Candidates should always add a link to the study, such as giving an example. As **6(a)** is about the UAB, an example from that should be used, for example 'facial grimaces by the patient could be misinterpreted as being in pain or upset or angry'.

Question 7

(a) (i) Candidates were often awarded 1 mark for correctly stating that the relevant ethical guideline was confidentiality. However, candidates rarely referred to sabotage, ignoring the question which stated, 'Outline why this was necessary *in this study'.*

(ii) Many candidates were awarded 1 mark for correctly stating that the relevant ethical guideline was deception. However, like **Question 7(a)(i)**, candidates rarely referred to sabotage, ignoring the question which stated, 'Outline why this was necessary *in this study'.*

(b) There were some excellent answers which linked their suggestion to the key study. The weaker responses did not relate to the study, for example, 'he knew all about the organisation because he worked there'. This statement is true but is very general and could apply to anything. Candidates should ensure that there is a comment about the study in the question in their answer.

(c) In response to this question candidates often gave 'pre-prepared answers' which did not apply to the situation of sabotage at work. For example, stating that the sample size was too small is not appropriate because the key study (a small sample) produced 29 sabotage forms and 11 reasons for sabotage. Stating that the type of industry might not generalise would be perfectly acceptable, or that there are cultural differences with sabotage never happening in some countries.

Question 8

(a) Some candidates failed to see the word 'accident' in the question and wrote about errors of commission in accounting. To clarify, an error of commission is where a worker misuses or misapplies standard/procedural information when working with a machine and performs an action which is incorrect or inappropriate, or performs an incorrect additional action which should not have been done.

(b) Some suggestions were 'give the workers training', but often trained workers still have accidents. Top marks were earned by candidates suggesting applying psychological strategies/studies such as using a token economy system as suggested by Fox et al.

(c) In response to this question, like in other **part (c)** questions, candidates often gave very brief, single sentence answers. These were sometimes worth 1 mark but, in many cases, they were too general or vague and therefore not creditworthy. For example, stating that 'the operator may not tell

the truth' is too vague and could apply to any question. Candidates should always give sufficient detail to make the strength or weakness unambiguous and ensure that it is related/linked to the question.

Section B

Question 9

(a) If candidates know very little about a topic area they should answer the question from their other chosen option. Some answers stated nothing more than the words 'attribution style' which was insufficient. With no further knowledge, only Level 1 marks will be awarded even if the method is very good. Answers at the top end of the mark range used Seligman's styles in designing their own questionnaire and some candidates modified or used questions from Seligman's ASQ. Knowledge of postal questionnaires was often weak, with nothing more than the use of the words 'postal questionnaire'. Similarly, many candidates used both closed and open questions to gather data when using only closed questions might increase the return rate and make analysis of the responses much easier. Simple plans that are done well nearly always achieve higher marks.

(b) (i) Following from **part (a),** many candidates could not answer this question part and did not demonstrate any knowledge of attribution styles, or they wrote what they knew about depression which was often Beck's cognitive triad or the Beck Depression Inventory. Candidates achieving top marks wrote about Seligman's styles where internal, stable and global attributions may result in a person feeling helpless and depressed. Some candidates wrote about the Attributional Style Questionnaire (ASQ) which also received credit.

(ii) Candidates achieving full marks for this question explained how they had used the knowledge (outlined in **(b)(i)**) about Seligman's styles to devise questions used in their postal questionnaires. Some candidates were awarded 0 marks because they didn't demonstrate knowledge of attribution styles, and so could not apply them, or wrote about general methodological things instead even though the question requires knowledge.

(c) (i) Many candidates applied both quantitative and qualitative data but found it difficult to give a reason why they had used both. There were answers such as 'I chose to use quantitative and qualitative data to increase reliability and validity'. An answer such as this was awarded no marks because it did not explain how reliability or validity is improved. Candidates are advised to choose one type of data. Using quantitative data, for example, would allow statistical analysis and comparison.

(ii) Most candidates were awarded 1 mark for correctly stating that postal questionnaires are often not returned by participants. Almost all candidates were unsuccessful in answering the 'in this study' part of the question and relate the weakness to their study and so could not be awarded the second available mark.

(iii) Many candidates chose an opportunity sampling technique and the stated reason for this choice was that people with depression who were in a clinic could be invited to participate. For candidates choosing to use a volunteer sample, candidates frequently stated that 'it is so consent is automatic'. Volunteering does not automatically give informed/valid consent. Another weakness was that there was no explanation of how the participants knew about the study. There was also often no link to the plan and the answers could apply to any study.

Question 10

(a) Many candidates who chose to attempt this question did not demonstrate knowledge about the psychology of food names, often resulting in Level 1 marks. Another weakness was that some candidates focused on the words 'menu item choice' and wrote whole answers on eye magnets, ignoring entirely the 'food name' component. Other candidates wrote excellent answers using food name as the IV and menu item choice as the DV.

(b) (i) For psychological knowledge the most appropriate research would be that by Lockyer (2006). This research found a positive effect of descriptive words such as 'tender', 'golden' and 'natural'. Some candidates described the research by Wansink et al. (2005). Although this study is no longer on the syllabus, it is still creditworthy as an alternative 'e.g.' study. Wansink compared food menu items with 'regular names' such as 'grilled chicken' with 'descriptive names' such as 'tender grilled chicken'.

(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, those candidates able to describe food names given by Lockyer or Wansink in **(b)(i)** explained how they used the same names, or similar examples of their own, in their plan.

(c) (i) Many candidates stated 'I chose quantitative and qualitative data' without giving a reason for their choice. They often constructed interviews to ask open questions simply to gather qualitative data which complicated their plan and often made it incoherent in attempting to include two methods. As the method was a field experiment, having 'menu item choice' gathering quantitative data as their DV would be simple and coherent. It would also save time and still be creditworthy.

(ii) By choosing both quantitative and qualitative data in **(c)(i)**, for this question part candidates were unsure what to write as a weakness of both. Candidates are advised to choose only one type of data (such as quantitative) because this **(i)** simplifies their **part (a)** plan, **(ii)** means that they can give a clear reason for one of these in **(c)(i)** and **(iii)** means that there is then a weakness for **(c)(ii)** (such as no qualitative data being gathered).

(iii) The choice of DV for many was simply 'participant choice of menu item' or 'the number of each food item chosen from the menu' because it gave quantitative data which could be statistically analysed to allow comparison of different food names. Such answers earned full marks for giving a reason and relating it to their DV.

Question 11

(a) Most candidates planned an appropriate way to investigate whether biochemical treatments are more effective than stimulation therapy/TENS for chronic pain, and included IV, DV, controls, and an experimental design. However, a number of answers suffered from errors such as muddling IV and DV or focusing on acute rather than chronic pain. Crucially, if a question invites candidates to consider effectiveness, then it is essential that their plan addresses this.

(b) (i) For psychological knowledge the most appropriate research was to consider biological treatments and/or stimulation therapy/TENS. This was done successfully by many candidates although sometimes candidates incorrectly thought that acupuncture was a stimulation therapy.

(ii) Candidates being awarded full marks wrote about how they used their knowledge of stimulation therapy/TENS and/or biological treatments to inform their plan. This was often done through details of a treatment programme such as the dosage of a drug and the duration of the treatment followed by participants. Top marks linked their **(b)(i)** answer with what they had done in their **part (a)** plan.

(c) (i) Many candidates wrote 'I would have a semi structured interview in order to obtain both quantitative and qualitative data'. This statement is incorrect and could not be awarded any marks. A semi-structured interview is when there are some fixed questions but with the option to ask additional questions. Questions can be entirely closed, or they could be entirely open. Quantitative and qualitative data doesn't have anything to do with the choice of structured/semi-structured or unstructured interviews.

(ii) Most candidates stated the generic 'the participant may not give honest answers when face-to-face' and were awarded 1 mark. However, most candidates failed to say what participants may not be honest about. Were the participants dishonest about whether TENS was effective, or whether their pain levels were better, or worse? A link to the study is essential.

(iii) A variety of sampling techniques were used, and some answers were clearly linked to their plan. However, many answers stated nothing more than 'an opportunity sample is quick and easy' which was awarded 0 marks.

Question 12

(a) There weren't many strong responses to this question. There were three major weaknesses in the candidate responses. First, candidates assumed that lack of sleep was the only health effect that needed to be investigated, when shiftwork has many other effects on health such as peptic ulcers and coronary heart disease. Second, many candidates asked questions about accidents, which was not part of the question, and so was irrelevant. Third, candidates knew very little about rapid

rotation shiftwork, showing no knowledge of the metropolitan or continental rotas for example. If candidates do not have knowledge about a topic area they should answer the question from their other chosen option.

(b) (i) Relevant psychological knowledge here could include two things: knowledge of rapid rotation shifts such as the Metropolitan rota and the Continental rota and also the effects of shiftwork on health such as the study by Gold et al. (1992). Many candidates followed the contents of the Gold study too closely and did not adapt the knowledge to plan their study. For example, Gold wrote about 'rotators', about sleep quality and about accidents. Strongest answers included health effects such as peptic ulcers, cardiovascular disease and reproductive effects.

(ii) Candidates focusing specifically on the Gold study found it difficult to explain how they applied knowledge of that study to plan their study, other than to state 'I did what Gold did'. Top answers explained why they had chosen the Metropolitan rota or the Continental rota and why they had asked specific questions that clearly linked to health effects.

(c) (i) Candidates frequently misunderstand this question. The question asks how answers to questions are scored or rated in order to draw conclusions from the data. This could be done if closed questions are asked about health and answered with yes/no, or on a rating scale, such as 0 – 4. The number of 'yes' or 'no' answers can then be found, or the average rating for each group calculated health data from both groups can be compared. Analysis of open questions could also be done, but this would be more complex. Having both quantitative and qualitative data would take far too long to explain, so is not advisable.

(ii) If a yes/no or rating scale is used to score/interpret the data, an obvious weakness is that there is no 'why' associated with the yes/no, or why a particular number is chosen. However, as many candidates included both quantitative and qualitative data, they struggled to find a weakness of doing both. Candidates are advised to select either quantitative or qualitative data.

(iii) Comments from **Question 11(c)(i)** above also apply here. For this question, a structured interview would have been apposite with the same questions asked from workers on both a rapid-rotation and a non-rapid-rotation shiftwork pattern so that their responses could be compared.

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 A Level component. For example, at AS candidates learn about methodology, such as experiments, which also apply to A Level.
- 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 for '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

Candidates are not required 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 these terms do not apply.

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. Many answers included strengths and weaknesses, but these were often not related to the question, and so restricted marks.

Question **part (c)** when answering the 'in this study', the 'relating/linking' part of the question simply rewriting the words in the question is insufficient and no marks can be awarded for it. Doing this merely shows that candidates can copy a question. Candidates need to add more to show understanding of the study in question. See examples below, such as **Question 5a**.

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, 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 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 are advised not to 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, which is an essential element of the plan.

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

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) Candidates were awarded 1 mark for stating positive comments such as 'it was nice when they smiled' and negative comments such as 'sometimes appeared hostile'. Often candidates did not finish the comment, which would have given them full marks. For example, 'it was nice when they smiled, *made me feel welcome*'. Similarly for negative 'sometimes appeared hostile, *sometimes rude*'.

(b) Some candidates misunderstood the question and suggested conducting a semi-structured interview or using some other method. Such answers were awarded 0 marks because the question asked about how the comments (already collected) could be analysed. Some candidates suggested analysing them by categorising them into positive and negative, but this was awarded 0 marks because this was stated in **Question 1(a)**. Stronger answers suggested finding which comments were the most frequent (1 mark) and giving an example from the study, such as finding that 'some were intimidating' was the most frequent comment (achieving 2 marks).

(c) A strength of qualitative data given by most candidates was that 'participants can answer the 'why' question' without reference to what the 'why' was referring to. Candidates struggled to give a weakness, often stating that 'qualitative data does not give numbers' without stating why that is a weakness. Candidates are advised to ensure that they always support their strength(s) or weakness(es) with an example from the study in the question. Without this, full marks can't be achieved.

Question 2

(a) This question asked about the *scale* used on the Y-BOCS, so answers referring to a 5-point scale, 0 – 4 range or the scale descriptors from none (0) to extreme (4) were awarded credit. The *scale* does not include actual questions or how the total score is assessed: 0 – 7 (subclinical through to 32 – 40 (extreme).

(b) The reliability of the Y-BOCS can be assessed through test-retest where the same person takes the same test on different occasions. It could also be assessed through split-half. It cannot be assessed through inter-rater reliability and, for those who confused terms, it cannot be assessed through any form of validity.

(c) There were three types of answer which covered the whole mark range: **(i)** candidates who appeared not to know what a psychometric test was (even though **parts (a)** and **(b)** were about Y-BOCS) and who could state nothing more than 'a weakness is social desirability'; **(ii)** candidates who could give a generic strength and weakness but who could not relate these to any psychometric test for OCD; **(iii)** candidates who provided an appropriate strength such as 'a

diagnostic tool used by therapists' to assess OCD indicating those with a score of 32 or more being 'extreme'. A common weakness was that psychometric tests such as the Y-BOCS and MOCI provide numbers but do not explore what the person's specific obsession-compulsion is, for example excessive hand washing.

Question 3

(a) (i) Nearly all candidates could state that most money was spent when classical music was being played and could be awarded 1 mark. The question asked about total spend and candidates stating that total spend for classical music was £32.51 compared to pop music at £29.46 were awarded an additional mark.

(ii) Very few candidates answered this question correctly. The question asked about spending on wine. The introduction to this key study led the authors to *expect* that more wine would be purchased with classical music, but this was not the case. Spending on wine was *found* to be the greatest in the no music condition £5.05 compared to £4.87 in the classical music condition and £4.08 in the pop music condition.

(b) There were some excellent answers provided by candidates who could suggest an alternative way in which the effect of classical music could be studied. Most frequent suggestion was to time how long people stayed in the restaurant in each of the three conditions. Another suggestion was to observe the non-verbal behaviour of diners, such as whether they 'nodded their head' in time to the music or sang along, showing their approval/enjoyment of it.

(c) This 'generalisations' question was answered well by most candidates who mentioned the restaurant in the study being 'upmarket' and so could not be generalised to 'down-market' restaurants. Cultural differences also featured often in relation to types of music where some countries do not play classical or English 'pop' music.

Question 4

(a) The two components of wayfinding 'knowing where we are' and 'knowing where we are going' were included in nearly all candidate answers which were awarded full marks.

(b) Some very inventive answers placed participants in railway stations, theme parks, unknown cities and laboratories (wearing VR goggles) and then gave a task where participants were asked to navigate to some destination. Top marks were awarded for answers which addressed the question specifically.

(c) This question was answered very well with candidates using 'signs' and 'you are here' maps as examples. The 'e.g.' study by Dogu and Erkip (2000) was often referred to. Strongest answers referred to things like gathering information about the effectiveness of existing signage, whereas some candidates referred to 'lots of people being available' which was vague and could apply to any study.

Question 5

(a) Many candidates did not know the dependent variables (DV's) of this key study and some stated the independent variables (IV's) instead. There are three DV's: the number of target children (1 mark) receiving one or more inoculations at the clinic (2 marks); the number of target children attending the clinic (1 mark) (for any reason) (2 marks); and the total number of inoculations/vaccinations/immunisations (1 mark) received by target children (2 marks).

(b) Most candidates suggested 'by sending a text message' which was sufficient for 1 mark to be awarded. However, candidates often added 'about the importance of inoculations' which was not creditworthy because this merely repeats the words of the question. Candidates should show their knowledge of the study or elaborate by adding 'the clinic is open for inoculations at these times and dates' for example.

(c) There were three types of answer: **(i)** candidates who could not provide any strength of objective data because they did not know what the term objective meant; **(ii)** those who knew the term and could provide strengths, such as 'objective data is fact that is derived from a source that has no subjectivity' but who could not give any examples and **(iii)** those who knew the term and gave

examples. Examples such as 'the number of children receiving inoculations/attending the clinic' or 'the number of children receiving vaccinations'.

Question 6

(a) Most candidates were awarded full marks for their answers to this question by stating that a visual analogue scale (VAS) is where a person marks or points to a spot on a continuum from no pain to extreme pain. Many correct variations led to the awarding of the second mark and included the use of colours or the use of smiley to crying faces.

(b) Candidates often confused validity with reliability and were awarded 0 marks. Correct answers often identified criterion validity and explained that a recording from the VAS could be compared with a recording from a different measure, such as the MPQ.

(c) Some candidates gave a definition of acute pain, then a definition of VAS without explaining how the VAS is used to help manage acute pain. Other candidates gave two strengths but did not relate them to acute pain as the question required. Given that a VAS was outlined in **6(a)** these examples t could have been used to support the strengths. For example, 'a child in acute pain, such as a broken leg can point to a 'crying face' on a piece of paper to show how much pain they are in', an answer which would be awarded 2 marks.

Question 7

(a) (i) Candidates could be awarded marks for stating any of the following: 106 participants, 65 female and 41 male, average age of 20.5 years, all undergraduates/candidates, all at University in the United States of America. Many candidates were awarded 2 marks, but some could provide only one of these features.

(ii) There were many incorrect answers, and some vague answers, such as 'they were all volunteers'. However, the question asked, '*how* were participants recruited', not *identify* the sampling technique. The correct answer was that they recruited through a psychology experiment website at a large university in the southeastern United States.

(b) An alternative location would be anywhere outside a laboratory and many candidates correctly suggested an actual workplace. This would increase ecological validity, as was often stated, but the question asked for the effect this would have on the results. In order to address this, candidates needed to know about the Claypoole and Szalma (2019) study which looked at electronic performance monitoring, which strongest answers did to achieve full marks.

(c) Candidates could give two strengths of random allocation, often explaining why there would be no experimenter or participant bias. However, the '*in this study*' part of the question was rarely addressed. Simply adding 'the experimenter does not know whether a participant is in the electronic performance or control condition' would show a little knowledge of what the study is about and be credited.

Question 8

(a) Many candidates confused the terms social loafing with social inhibition. Social inhibition is where the performance of an individual is restricted because of the mere presence of other people watching them (which can apply to an individual person). Social loafing is the decrease in effort when performing in a group (with no-one watching) compared to when they perform alone. A few candidates incorrectly assumed social loafing was concerned with people in groups making decisions, i.e. 'groupthink'.

(b) Suggestions here were aimed at reducing social inhibition or groupthink, all of which were incorrect. Social loafing can best be reduced in a meeting if managers clearly outline all the tasks required of each individual and where they recognise individual contributions and reward each member.

(c) In response to this question candidates again struggled to correctly apply *social loafing*. One strength of studying social loafing is that it can identify guilty individuals who can be re-trained to work at appropriate levels; another is that if social loafing is eliminated, all members of a team can

work happily with the confidence that every team member is 'pulling their weight'. There are also benefits for the organisation in terms of productivity.

Section B

Question 9

(a) Four different errors reduced the marks of many candidates: **(i)** not using the correct method (such as using a questionnaire) and not an observation; **(ii)** not referring to any ethical guidelines at all (essential inclusions in any plan); **(iii)** not answering the question set which was to find which compulsion is most common (often focusing on just one compulsion), and **(iv)** assuming incorrectly that OCD and ICD (impulse control disorder) are the same thing and writing about pyromania, for example. Many candidates who planned an observation decided to add a questionnaire or an interview to gather qualitative data. This was irrelevant to the named method of observation and was awarded no additional marks.

(b) (i) For psychological knowledge candidates should have focused on what the syllabus described as 'types of obsessions and compulsions' and provided an outline of the common types such as: ritualistic acts; excessive hand washing; cleaning items; checking items; counting, ordering or arranging; doubting; touching repeatedly. Some of these could appear in response categories in a structured observation. Some candidates focused on the case study of 'Charles' which was creditworthy but often restricted answers to just handwashing.

(ii) Candidates achieving full marks for this question explained how they had used the knowledge (outlined in **(b)(i)**) such as the different types of compulsions (checking, cleaning, etc.) to form their response categories. Some candidates used the Y-BOCS to assess their participants had diagnosed OCD which was also creditworthy.

(c) (i) Giving a reason for choice of structured observation was straightforward for candidates using behavioural/response categories and full mark answers gave examples of the types of compulsion they had used. Some candidates used an unstructured observation 'to see what they would find' often because they could not identify any type of compulsion either in **9(a)** or **9(b)(i)**.

(ii) A typical answer was to state that 'there might be a behaviour that was not on the list of response categories' but without any example of what this might be.

(iii) Some candidates gave a reason for including both quantitative and qualitative data (and included using a questionnaire in their plan), but often the reason for qualitative data was 'to tell the severity of the disorder' or 'to have a better understanding of the disorder' both of which were irrelevant to the question set and were awarded no marks.

Question 10

(a) This question saw many candidates achieving no more than Level 1 or 2 because their plans were overcomplicated and lacked coherence. A simple approach would be to interview people (structured and face-to-face) about the taste of one food when in an outdoor market with different noise levels. However, many candidates played music through headphones; conducted the study in a laboratory; applied too many types of food such as sweetness, saltiness, and crunchiness; and conducted complex interviews simply to gather both quantitative and qualitative data. Candidates are advised to design a simple plan with features in a little detail rather than trying to apply too many things that are not specifically answering the question.

(b) (i) For psychological knowledge the most appropriate research would be that by Woods et al. (2010) which is an 'e.g.' study so an alternative study could be substituted. Woods et al. found that sweetness and saltiness was significantly lower in the loud sound conditions and crunchiness was reported to be more intense. Some candidates looked at the reasons why sound influences taste, which was also creditworthy.

(ii) Candidates often replicated in their plan what Woods et al. did in theirs, which led to the overcomplication in plans. Candidates should base their plan on the knowledge gained from a study, not replicate that study. For example, Woods et al. used headphones in a laboratory, whereas this question required the study be conducted in an outdoor market where there would be different levels of natural noise and headphones would be inappropriate. Woods et al. investigated

sweetness, saltiness and crunchiness, but that does not mean that candidates also had to investigate all three. Answers to this question require an explanation, not single sentences such as 'Woods et al. used crunchiness and so did I'.

(c) (i) Many candidates incorrectly claimed that they used a semi-structured interview so they could use both open and closed questions and gather both quantitative and qualitative data. Doing this also complicated their plans, when a few simple, closed questions with yes/no answers would have been sufficient.

(ii) Candidates making incorrect assumptions about the interview format in **(c)(i)** struggled to answer this question correctly because they could not give a weakness. If they had chosen a structured interview for **(c)(i)**, for example, then a weakness of that is that participants could not add any additional information about their noise/taste experience.

(iii) There were answers which merely stated 'I asked people at random' which achieved 0 marks, and those who achieved 1 mark for 'I used an opportunity sample because participants were available at the time' (with no link to the study) to those who achieved full marks for adding 'as they left the outdoor market after tasting the food (linking to the study).

Question 11

(a) A minor weakness in many answers was to state nothing more than 'my online questionnaire' with very little detail of how this would be implemented with 'I would post it online' being common. A major weakness was that many candidates did not know what was meant by rational non-adherence and throughout their plan referred to no additional detail that showed they knew anything about it. Some candidates wrote about delay in seeking treatment and some focused on adherence in general, neither of which answered the question. Another problem in many plans was that the questions asked were often nothing more than whether the cost of the medicine was restrictive rather than asking about a range of adherence decisions.

(b) (i) For psychological knowledge the most appropriate research were the studies by Laba et al. (2012) and Bulpitt (1998). Bulpitt focused on the negative side effects for the rational decision to stop taking medication and Laba et al. looked at immediate and long-term medication harm and benefit, cost, and treatment regimen.

(ii) Top mark answers used the knowledge described in **(b)(i)** to design questions for their plan. For example, Laba et al. referred to 'treatment regimen' which led to questions such as 'do you always take your medication at the same time each day' and 'do you always take the correct number of pills even though they might make you feel unwell'.

(c) (i) Candidates had little difficulty in giving a reason for their use of closed questions. Reasons given were for example 'same format for all participants', 'easy to score/analyse', 'large numbers questioned relatively quickly'. What was often lacking was an example to link the reason to their plan even though questions that could have been used appeared in the plan.

(ii) Following from **(c)(i)** the weakness focused on a lack of qualitative data in many answers, the 'why' question, and others correctly focused on the rating scale they had used. However, although the weakness was that a five-point scale is restricting, there was no example from the plan of the five-point scale that had been used. Candidates are advised that for all **part (c)** questions examples should be given from their plan.

(iii) It is important when conducting an online questionnaire that candidates know what their sampling technique is. Many candidates simply 'posted their questionnaire online' without stating where or how participants would see the study advertised. Candidates were often undecided whether the technique they used was volunteer (because people choose to complete the questionnaire) or random (because it is online 'everyone has an equal chance of seeing it') or opportunity.

Question 12

(a) There were some excellent plans which coherently investigated the reduction on the number of accidents (a DV) when a reward system like token economy was introduced or not (an IV). Appropriate experimental designs were applied as were controls. A number of candidates incorrectly planned a study using motivational rewards which are not linked to accidents.

(b) (i) Relevant psychological knowledge here was that by Fox et al. (1987) who studied the use of a token economy system at two open-cast mines. Employees could earn stamps/tokens to gain rewards for working without time lost for injury; not being involved in accidental damage to equipment; and behaviour that prevented accidents or injuries. Stamps were lost for unsafe behaviour that could cause accidents. Maslow's work on motivation was not relevant.

(ii) Top answers explained how they adapted and applied parts of the Fox et al. strategy (which was conducted on mine workers) to their study on workers in a clothes factory where the types of accidents would be very different. Sometimes methodological comments were made, such as sampling technique or actual method used which receive no credit.

(c) (i) This question was relevant because candidates could use either repeated measures (same participants do both conditions) or independent measures (different participants do different conditions). A reason was given and frequently this reason was applied to the plan.

(ii) The weakness was dependent on the design chosen in **(c)(i).** Common generic answers were provided: with an independent design there is no control over participant variables; a repeated measures design would not work because the same participant would do two (or more) conditions and the effect of the first might influence the second. Either was creditworthy but without a link to the candidate's plan no more than 1 mark could be awarded.

(iii) Many candidates were awarded full marks for stating that the type of experiment was a field experiment and relating it to their plan. Other candidates were confused because they believed that conducting a study in a clothes factory meant that it was a natural experiment which is incorrect.

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 A Level component. For example, at AS candidates learn about methodology, such as experiments, which also apply to A Level.
- 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 for '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 the heading instructions for each question section carefully.

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

Candidates are not required to re-write the question before beginning their answer.

Some candidates apply the terms 'demand characteristics' and 'social desirability' to every answer when in most cases these terms do not apply.

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 not always related to the question, and so restricted marks.

In question **part (c),** when answering the 'in this study', the 'relating/linking' part of the question, simply rewriting the words in the question is not sufficient and no marks can be awarded for it. This merely shows that candidates can copy a question. Candidates need to add more to show understanding of the study in question. See examples below, such as **Question 5a**.

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, 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 stated 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 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 are advised not to 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, which is an essential element of the plan.

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

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) Despite the question stating 'other than getting a phlebotomy' (which means that this will receive no credit because it is part of the question) many candidates overlooked the instruction. There were nine other situations included in the SUDS hierarchy candidates could use and any two of these nine received credit. Candidates often stated 'getting a finger prick' which was too vague to be awarded marks.

(ii) Most candidates could not provide an accurate result for any one of the nine possible situations. Candidates achieving full marks gave answers such as 'wife taking my blood pressure reduced from 50 before, to 45 during and 20 after'. Candidates were awarded 0 marks for stating results for 'getting a phlebotomy' which could not be used as this was a continuation of question **part (a)**.

(b) Candidates frequently were awarded 1 mark for stating 'T may not have been truthful' which is correct, but without stating what T was not being truthful about. Adding an additional comment like 'the SUDS score he gave for taking my own blood pressure' would link the point about being untruthful to the study and an additional mark would be awarded.

(c) Candidates could often give two strengths of quantitative data but rarely related the strengths to the study by Chapman and DeLapp and 'T'. To be awarded full marks, candidates must relate/link their strength (or weakness if applicable) to the study in question. Stating 'quantitative data is numerical' is not a strength; that quantitative data can be statistically analysed is.

Question 2

(a) To be awarded marks candidates needed to explain how irrational beliefs are changed by REBT. This means that the D and E (disputing and 'effects') of the Ellis ABCDE model were essential inclusions. Some candidates focused on the ABC, the theory, and this could achieve 1 mark. Many candidates merely stated that irrational beliefs are changed into rational ones, and this was awarded 0 marks because it showed no knowledge of Ellis's REBT.

(b) Many candidates stated 'I would give a questionnaire' without elaboration of technique or format. A2 requires more than just 'questionnaire'. Candidates who stated 'I would give a BDI' without elaboration were also awarded 0 marks. To achieve marks, candidates should state what a BDI is and give some detail about it (closed questions, 21 items, rating scale, etc.) and say how it would measure the effectiveness of REBT. Answers achieving top marks constructed their own questionnaire, stating the format and technique, gave an example of a question, and how this would be scored.

(c) The most common answer was that REBT does not involve biochemical treatments (1 mark) with many candidates going on to state, 'such as anti-depressants which can be addictive' and so earning one more mark. Some candidates gave the same strength again but with a different example, such as 'so there are no side effects'. However, the question states two strengths and so the strengths must be different.

Question 3

(a) (i) Very few candidates knew the questions that participants were asked in the key study by Robson et al. There were many incorrect responses, and some candidates knew a few relevant words. For example, candidates might know the word 'comfortable' but not be able to correctly place it in the question asked by Robson et al. which was 'Sitting at this table would make me feel comfortable'.

(ii) Very few candidates knew the rating scale that Robson et al. used. The questionnaire is a fundamental aspect of the key study. Marks were awarded for stating 'Likert-type scale', 'seven-point scale' and 'labels of strongly disagree (1) via 'neither agree nor disagree' to strongly agree (7)'.

(b) There were some excellent answers provided by candidates who could suggest an alternative way in which the table spacing could be studied. The strongest answers suggested a method and linked it to the measurement of table spacing. Many candidates either stated a method in detail but not what they were measuring, achieving 1 mark, or stated nothing more about the method than 'I would interview people' with no link to anything.

(c) Like other **part (c)** questions, candidates often gave a strength and weakness without relating it to the question. One strength is that a 7-point scale allows a wider range of responses than a 3-point of 5-point scale. However, most candidates did not relate this to the key study, when full marks could have been awarded by stating that 'Robson et al. used a 7-point scale' (strongly agree to strongly disagree) when asking 'Sitting at this table, I would feel comfortable'. Weaknesses mentioned were often good but were rarely related to the Robson et al. study.

Question 4

(a) All candidates were awarded 1 mark for their answers on brand awareness, but only those providing some elaboration in their answer were awarded the second available mark. Reference to brand name, brand logo or brand slogan is sufficient.

(b) Similar to answers to other questions, many suggestions included nothing more than 'I would do an observation/questionnaire/interview' with no elaboration about the format or technique. Marks could not be awarded for such answers. Candidates providing some elaboration were awarded 1 mark with those linking it to how the method would be used to investigate brand awareness being awarded full marks.

(c) The lack of inclusion about the features of different methods evident in **4(b)** was also evident here. The syllabus states 'describe *the main features* of an observation'. It also distinguishes between questionnaire/interview technique and format. This meant that candidates stating 'I would do an observation' struggled to provide a strength or weakness, and when this was not linked to brand awareness no marks were awarded.

Question 5

(a) There were seven different features of the sample that candidates could identify to be awarded marks, for example 1167 participants at the outset; 1038 at the end; aged 11 – 14 years; two middle schools; mostly Jewish. Most candidates could give at least one feature.

(ii) Many candidates suggested reasons why participants withdrew from the study. However, this was an AO1 question meaning that the only correct reasons for withdrawal were those stated in the original study. These were that participants withdrew due to absence (34 intervention, 22 control) and withdrew due to refusal to complete questionnaires (2 intervention, 4 control).

(b) Most candidates suggested 'to do a comparison/baseline' which was sufficient for 1 mark to be awarded. However, candidates could not elaborate beyond this. To clarify, the wait-list control condition meant that schools and appropriate candidates were 'ready and waiting' should the

selected school dropout and six school principals with similar features to the intervention school agreed to be on the wait-control list.

(c) Candidates frequently assumed that a longitudinal study automatically gathered both quantitative and qualitative data and that this data is always in-depth, which is incorrect. Participants might only have to complete a short, closed question questionnaire with only 5 items. No candidate referred to what data was gathered from the participants in this study at the 'one-year-later' follow up.

Question 6

(a) Identifying any two of the ten behavioural categories from the UAB pain behaviour scale earned 2 marks. Included in the ten categories were things like 'vocal complaints non-verbal' and 'facial grimaces'. Marks could be awarded for including examples that were listed on the UAB, such as 'groans, moans, gasps' but not for any example not on the UAB list.

(b) Candidates had to suggest a psychological treatment, such as attention diversion, non-pain imagery or cognitive redefinition which earned 1 mark. Biological or alternative treatments such as acupuncture were not creditworthy. In addition, candidates had to say how acute pain would be treated, and this aspect was only addressed by a very small number of candidates because they failed to answer the question set.

(c) Answers to this question often provided brief weaknesses such as 'the pain behaviour may be misinterpreted' with no elaboration. Answers like this will often be awarded 1 mark, but not more. Candidates should always add a link to the study, such as giving an example. As **6(a)** is about the UAB, an example from that should be used, for example 'facial grimaces by the patient could be misinterpreted as being in pain or upset or angry'.

Question 7

(a) (i) Candidates were often awarded 1 mark for correctly stating that the relevant ethical guideline was confidentiality. However, candidates rarely referred to sabotage, ignoring the question which stated, 'Outline why this was necessary *in this study'.*

(ii) Many candidates were awarded 1 mark for correctly stating that the relevant ethical guideline was deception. However, like **Question 7(a)(i)**, candidates rarely referred to sabotage, ignoring the question which stated, 'Outline why this was necessary *in this study'.*

(b) There were some excellent answers which linked their suggestion to the key study. The weaker responses did not relate to the study, for example, 'he knew all about the organisation because he worked there'. This statement is true but is very general and could apply to anything. Candidates should ensure that there is a comment about the study in the question in their answer.

(c) In response to this question candidates often gave 'pre-prepared answers' which did not apply to the situation of sabotage at work. For example, stating that the sample size was too small is not appropriate because the key study (a small sample) produced 29 sabotage forms and 11 reasons for sabotage. Stating that the type of industry might not generalise would be perfectly acceptable, or that there are cultural differences with sabotage never happening in some countries.

Question 8

(a) Some candidates failed to see the word 'accident' in the question and wrote about errors of commission in accounting. To clarify, an error of commission is where a worker misuses or misapplies standard/procedural information when working with a machine and performs an action which is incorrect or inappropriate, or performs an incorrect additional action which should not have been done.

(b) Some suggestions were 'give the workers training', but often trained workers still have accidents. Top marks were earned by candidates suggesting applying psychological strategies/studies such as using a token economy system as suggested by Fox et al.

(c) In response to this question, like in other **part (c)** questions, candidates often gave very brief, single sentence answers. These were sometimes worth 1 mark but, in many cases, they were too general or vague and therefore not creditworthy. For example, stating that 'the operator may not tell

the truth' is too vague and could apply to any question. Candidates should always give sufficient detail to make the strength or weakness unambiguous and ensure that it is related/linked to the question.

Section B

Question 9

(a) If candidates know very little about a topic area they should answer the question from their other chosen option. Some answers stated nothing more than the words 'attribution style' which was insufficient. With no further knowledge, only Level 1 marks will be awarded even if the method is very good. Answers at the top end of the mark range used Seligman's styles in designing their own questionnaire and some candidates modified or used questions from Seligman's ASQ. Knowledge of postal questionnaires was often weak, with nothing more than the use of the words 'postal questionnaire'. Similarly, many candidates used both closed and open questions to gather data when using only closed questions might increase the return rate and make analysis of the responses much easier. Simple plans that are done well nearly always achieve higher marks.

(b) (i) Following from **part (a),** many candidates could not answer this question part and did not demonstrate any knowledge of attribution styles, or they wrote what they knew about depression which was often Beck's cognitive triad or the Beck Depression Inventory. Candidates achieving top marks wrote about Seligman's styles where internal, stable and global attributions may result in a person feeling helpless and depressed. Some candidates wrote about the Attributional Style Questionnaire (ASQ) which also received credit.

(ii) Candidates achieving full marks for this question explained how they had used the knowledge (outlined in **(b)(i)**) about Seligman's styles to devise questions used in their postal questionnaires. Some candidates were awarded 0 marks because they didn't demonstrate knowledge of attribution styles, and so could not apply them, or wrote about general methodological things instead even though the question requires knowledge.

(c) (i) Many candidates applied both quantitative and qualitative data but found it difficult to give a reason why they had used both. There were answers such as 'I chose to use quantitative and qualitative data to increase reliability and validity'. An answer such as this was awarded no marks because it did not explain how reliability or validity is improved. Candidates are advised to choose one type of data. Using quantitative data, for example, would allow statistical analysis and comparison.

(ii) Most candidates were awarded 1 mark for correctly stating that postal questionnaires are often not returned by participants. Almost all candidates were unsuccessful in answering the 'in this study' part of the question and relate the weakness to their study and so could not be awarded the second available mark.

(iii) Many candidates chose an opportunity sampling technique and the stated reason for this choice was that people with depression who were in a clinic could be invited to participate. For candidates choosing to use a volunteer sample, candidates frequently stated that 'it is so consent is automatic'. Volunteering does not automatically give informed/valid consent. Another weakness was that there was no explanation of how the participants knew about the study. There was also often no link to the plan and the answers could apply to any study.

Question 10

(a) Many candidates who chose to attempt this question did not demonstrate knowledge about the psychology of food names, often resulting in Level 1 marks. Another weakness was that some candidates focused on the words 'menu item choice' and wrote whole answers on eye magnets, ignoring entirely the 'food name' component. Other candidates wrote excellent answers using food name as the IV and menu item choice as the DV.

(b) (i) For psychological knowledge the most appropriate research would be that by Lockyer (2006). This research found a positive effect of descriptive words such as 'tender', 'golden' and 'natural'. Some candidates described the research by Wansink et al. (2005). Although this study is no longer on the syllabus, it is still creditworthy as an alternative 'e.g.' study. Wansink compared food menu items with 'regular names' such as 'grilled chicken' with 'descriptive names' such as 'tender grilled chicken'.

(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, those candidates able to describe food names given by Lockyer or Wansink in **(b)(i)** explained how they used the same names, or similar examples of their own, in their plan.

(c) (i) Many candidates stated 'I chose quantitative and qualitative data' without giving a reason for their choice. They often constructed interviews to ask open questions simply to gather qualitative data which complicated their plan and often made it incoherent in attempting to include two methods. As the method was a field experiment, having 'menu item choice' gathering quantitative data as their DV would be simple and coherent. It would also save time and still be creditworthy.

(ii) By choosing both quantitative and qualitative data in **(c)(i)**, for this question part candidates were unsure what to write as a weakness of both. Candidates are advised to choose only one type of data (such as quantitative) because this **(i)** simplifies their **part (a)** plan, **(ii)** means that they can give a clear reason for one of these in **(c)(i)** and **(iii)** means that there is then a weakness for **(c)(ii)** (such as no qualitative data being gathered).

(iii) The choice of DV for many was simply 'participant choice of menu item' or 'the number of each food item chosen from the menu' because it gave quantitative data which could be statistically analysed to allow comparison of different food names. Such answers earned full marks for giving a reason and relating it to their DV.

Question 11

(a) Most candidates planned an appropriate way to investigate whether biochemical treatments are more effective than stimulation therapy/TENS for chronic pain, and included IV, DV, controls, and an experimental design. However, a number of answers suffered from errors such as muddling IV and DV or focusing on acute rather than chronic pain. Crucially, if a question invites candidates to consider effectiveness, then it is essential that their plan addresses this.

(b) (i) For psychological knowledge the most appropriate research was to consider biological treatments and/or stimulation therapy/TENS. This was done successfully by many candidates although sometimes candidates incorrectly thought that acupuncture was a stimulation therapy.

(ii) Candidates being awarded full marks wrote about how they used their knowledge of stimulation therapy/TENS and/or biological treatments to inform their plan. This was often done through details of a treatment programme such as the dosage of a drug and the duration of the treatment followed by participants. Top marks linked their **(b)(i)** answer with what they had done in their **part (a)** plan.

(c) (i) Many candidates wrote 'I would have a semi structured interview in order to obtain both quantitative and qualitative data'. This statement is incorrect and could not be awarded any marks. A semi-structured interview is when there are some fixed questions but with the option to ask additional questions. Questions can be entirely closed, or they could be entirely open. Quantitative and qualitative data doesn't have anything to do with the choice of structured/semi-structured or unstructured interviews.

(ii) Most candidates stated the generic 'the participant may not give honest answers when face-to-face' and were awarded 1 mark. However, most candidates failed to say what participants may not be honest about. Were the participants dishonest about whether TENS was effective, or whether their pain levels were better, or worse? A link to the study is essential.

(iii) A variety of sampling techniques were used, and some answers were clearly linked to their plan. However, many answers stated nothing more than 'an opportunity sample is quick and easy' which was awarded 0 marks.

Question 12

(a) There weren't many strong responses to this question. There were three major weaknesses in the candidate responses. First, candidates assumed that lack of sleep was the only health effect that needed to be investigated, when shiftwork has many other effects on health such as peptic ulcers and coronary heart disease. Second, many candidates asked questions about accidents, which was not part of the question, and so was irrelevant. Third, candidates knew very little about rapid

rotation shiftwork, showing no knowledge of the metropolitan or continental rotas for example. If candidates do not have knowledge about a topic area they should answer the question from their other chosen option.

(b) (i) Relevant psychological knowledge here could include two things: knowledge of rapid rotation shifts such as the Metropolitan rota and the Continental rota and also the effects of shiftwork on health such as the study by Gold et al. (1992). Many candidates followed the contents of the Gold study too closely and did not adapt the knowledge to plan their study. For example, Gold wrote about 'rotators', about sleep quality and about accidents. Strongest answers included health effects such as peptic ulcers, cardiovascular disease and reproductive effects.

(ii) Candidates focusing specifically on the Gold study found it difficult to explain how they applied knowledge of that study to plan their study, other than to state 'I did what Gold did'. Top answers explained why they had chosen the Metropolitan rota or the Continental rota and why they had asked specific questions that clearly linked to health effects.

(c) (i) Candidates frequently misunderstand this question. The question asks how answers to questions are scored or rated in order to draw conclusions from the data. This could be done if closed questions are asked about health and answered with yes/no, or on a rating scale, such as 0 – 4. The number of 'yes' or 'no' answers can then be found, or the average rating for each group calculated health data from both groups can be compared. Analysis of open questions could also be done, but this would be more complex. Having both quantitative and qualitative data would take far too long to explain, so is not advisable.

(ii) If a yes/no or rating scale is used to score/interpret the data, an obvious weakness is that there is no 'why' associated with the yes/no, or why a particular number is chosen. However, as many candidates included both quantitative and qualitative data, they struggled to find a weakness of doing both. Candidates are advised to select either quantitative or qualitative data.

(iii) Comments from **Question 11(c)(i)** above also apply here. For this question, a structured interview would have been apposite with the same questions asked from workers on both a rapid-rotation and a non-rapid-rotation shiftwork pattern so that their responses could be compared.