



---

# **GCE EXAMINERS' REPORTS**

---

**GCE  
PSYCHOLOGY  
AS/Advanced**

**SUMMER 2022**

Grade boundary information for this subject is available on the WJEC public website at:  
<https://www.wjecservices.co.uk/MarkToUMS/default.aspx?!=en>

### **Online Results Analysis**

WJEC provides information to examination centres via the WJEC secure website. This is restricted to centre staff only. Access is granted to centre staff by the Examinations Officer at the centre.

### **Annual Statistical Report**

The annual Statistical Report (issued in the second half of the Autumn Term) gives overall outcomes of all examinations administered by WJEC.

<b>Unit</b>	<b>Page</b>
Unit 1	1
Unit 2	5
Unit 3	10
Unit 4	16

# PSYCHOLOGY

## General Certificate of Education

Summer 2022

### Advanced Subsidiary/Advanced

#### UNIT 1 - Past to Present

#### General Comments

Candidates should be commended on producing some excellent responses, showing depth and breadth of knowledge as well as an ability to express themselves clearly and concisely under exam conditions. They were able to respond to the skill and content requirements of each question thoroughly across the paper. Areas for improvement remain focused on application skills in AO2 questions as well as contextualising responses to demonstrate AO3.

#### Comments on individual questions/sections

- Q.1 (a)** Overall, this was one of the most attempted questions on the paper (98.9%) and also received the highest facility factor (73.7). Most candidates were able to outline key features of the different areas of the brain, and many gave detailed descriptions of their links to different behaviours. The best answers were also able to use appropriate examples to illustrate this assumption such as Raine's study, areas of the brain associated with language or case studies such as Phineas Gage. However, it was not always clearly explained how these case studies or research were an illustration of this assumption and this sometimes limited the marks that could be awarded.
- Q.1 (b)** This question was generally well answered with 98.9% attempting it and a mean score on the question of 5.1. There were lots of excellent answers, characterised by use of appropriate terminology and examples. In psychosurgery responses, many described a range of different techniques accurately, including historical and modern methods. Terminology was used to add detail and the methods used were fully explained. However, it was not always clearly linked to mental illness and the use of these methods as therapy, which would have helped some answers reach the top band. In answers for drug therapy, there was some confusion of the mode of action for the drugs being described although many candidates were able to name a variety of different ones. The most successful answers fully explained how the drug worked to achieve its purpose rather than just stating it. For example, how SSRI's block reuptake of serotonin in order to increase its availability in the synapse.
- Q.2** This question received the lowest facility factor on the paper of 44.4, with a mean score of 2.7. The standard deviation of 1.8 suggests a wide range in the quality of responses. There were some outstanding answers to this question, where candidates were able to accurately state the average speed estimate for all 5 verbs in Experiment I as well as the numbers for each group who remembered seeing broken glass in Experiment II.

Where there is any discrepancy, figures should be taken from the original article rather than from other sources. It is important to highlight that if candidates used a table to display the results this did not affect the mark awarded, and usually helped add clarity to the response. On the other hand, there were many vague answers that also did not address both experiments. Given the phrasing of the question this was an important aspect of the marking criteria. Some candidates also wasted valuable time describing parts of the study not required by the question, such as procedures or conclusions.

- Q.3** As with the other assumptions question on the paper, this had a high attempt rate (98.9%) and a high facility factor (65.3). Most candidates were able to accurately state two assumptions from the behaviourist approach in their answers. The 'learning through conditioning' assumption was particularly well explained with some very detailed descriptions of relevant examples such as Pavlov's Dogs and Skinner's Rats. However, when candidates explained the 'blank slate' assumption, they were often able to use appropriate terminology but failed to fully address the 'using an example from psychology' part of the question. Therefore, they were unable to access the highest marks. Candidates should be encouraged to use theories and research from other parts of the course to illustrate each assumption. For example, the classic evidence of Watson and Rayner could be used to illustrate the learning of behaviour from the environment on a child who previously did not display any fear of the stimuli.
- Q.4** To be successful in this question, candidates needed to cover both strengths and weaknesses, using examples from the approach to illustrate why these were a strength or weakness of the approach. Many candidates were able to do this successfully, for example highlighting how the assumption of 'internal mental processes' illustrates reductionism and how this may be an overly simplistic view of behaviour. On some occasions the lack of an appropriate example led to a generic response that could have been applied to many approaches and therefore meant that answers were 'not appropriately contextualised'. This may have contributed to the mean mark of 4.8 and the facility factor of 47.8. Some candidates also limited their marks by largely evaluating Loftus and Palmer's research or one of the therapies. Whilst these can be used to illustrate some of the wider evaluation points, the focus of this question should be evaluating the approach as a whole. The key distinction between answers considered 'thorough' and those marked as 'reasonable' was in the conclusions drawn (either throughout the answer or at the end). This means candidates being able to explain why the point and example they had chosen illustrated a strength or weakness of the approach.
- Q.5** Most candidates were able to identify a relevant assumption and refer to a specific type of relationship, and the mean mark was 2.3. However, answers varied in terms of the use of terminology and the clarity of explanation. The best answers used specific terminology from the approach such as signature strengths or the good life. In addition they were able to use this to explain relationships, for example, how romantic relationships allow both the expression and development of specific character strengths such as kindness, love and honesty. Some candidates presented relevant research evidence but did not always use this successfully to explain the importance of relationships for wellbeing. Many candidates muddled their explanation with the halo effect from the cognitive approach and therefore received no credit for their response.

- Q.6** As is often the case with classic evidence questions, the examiners were very impressed with the level of detail included in many of the answers. 96.5% of candidates attempted this question and there was a mean mark of 4.2. Some candidates were able to describe, in depth, all of the different sessions of the study, including accurate information about the age of Little Albert, the stimuli used and how often each was presented. This question focused on procedure (the actions of Watson and Rayner) although some candidates focused on findings (the behaviour of Little Albert) and therefore could only access limited marks. As with Q.2 it is important to use the original article as the source of information for details about the study.
- Q.7** This question assessed AO2 skills and therefore required candidates to apply their knowledge to the scenario presented. The answer also requires similar evaluation skills to Q.4. Candidates needed to state a strength/weakness of Bowlby's research, give a contextualised example from the study and then explain why this was a strength or weakness. The best answers focused on methodological and ethical issues (as indicated by the scenario) and directly referred to the different viewpoints of Dan and his teacher. However, there was a failure by many candidates to fully engage with this question indicated by a mean mark of 4.5 and a facility factor of 45.1. Occasionally, candidates evaluated Raine's research study and therefore received no credit for their response.
- Q.8** Most candidates were able to identify a difference between these two approaches and provide a basic comparison (e.g. psychodynamic being unscientific and positive using scientific methods). Many then included context to illustrate this difference, using terminology and examples from the approach (e.g. the unfalsifiable nature of the unconscious mind compared to the use of brain scanning techniques to investigate the effects of mindfulness). However, lacking in many answers was an evaluative comment. The overall mean score of 2.0 was the lowest of the 3 questions on the paper with a maximum mark of 4. As this question assesses AO3 skills, for the top band, candidates needed to show some evidence of evaluation, for example explaining that the difference meant that the positive approach was a 'better' explanation (e.g. scientific methods more credible). The weakest answers simply described both approaches and their assumptions, which attracted no credit. Some also muddled up key terms such as reductionism and determinism when trying to explain the difference.
- Q.9** This question had the highest facility factor of the AO3 questions on the paper (53.8). The focus of this question was to discuss the ethical issues of one of the therapies from the behaviourist approach and most answers focused on issues of harm and valid consent. The best answers discussed both sides of any ethical issue to create a more discursive response. For example, for systematic desensitisation the issue of harm caused by exposure to a phobic stimulus was compared to the relaxation techniques taught by therapists to offset this potential harm. For aversion therapy many compared the potential short term harm caused by the use of Antabuse to the potential longer term harm caused by the addiction itself. Some answers wrote about effectiveness of the therapies and therefore received no credit for this. There were also times where ethical issues were discussed generically and therefore this also limited the marks available.

**Q.10** The lower attempt rate for this question (93.1%) suggests that candidates may have run out of time and been unable to finish the paper. In this question candidates received credit for discussion of both ethical issues and effectiveness, and many covered a wide range of points using appropriate examples from each therapy to illustrate each one. Answers on mindfulness were seen more frequently than those on quality of life therapy. As with other AO3 questions on this paper the discussion of why something was a strength, or a weakness of the therapy was not always clear. In particular, many candidates included details of research studies when discussing effectiveness, but the research was described, rather than being used as evaluation of the therapy.

### **Summary of key points**

The use of psychological examples in assumptions questions (e.g. Q.1 (a) and Q.3) is required to access the highest marks. Candidates should be able to do this for all assumptions in all approaches.

Candidates would benefit from more practice of AO2 skills to ensure they fully apply their knowledge to the requirements of these questions (e.g. Q.7).

In evaluation questions (e.g. Q.4, Q.10), candidates would benefit from a clearer explanation of why something is a strength or weakness.

In compare and contrast questions (e.g. Q.8), candidates should include evaluative comments as well as explaining the similarity/difference.

# PSYCHOLOGY

## General Certificate of Education

Summer 2022

### Advanced Subsidiary/Advanced

#### UNIT 2 - Using Psychological Concepts

##### General Comments

Overall, candidates were well prepared with very few non attempted questions. Areas for improvement remain focused on more challenging aspects of the specification; observational sampling techniques (Question 1), interpretation of graphical representation (Questions 6c i and ii and 7d i and ii), and ethics committees (Question 8e). It was encouraging to see a better quality of handwriting overall, however centres should encourage candidates to use the extra lines at the back, of the booklet if they find the constrained space provided it not long enough for their response. Careful labelling of this additional material is also recommended. Additionally, candidates should be reminded to use the number of marks as an indication of how much detail is required in their response. Where the question is applied to a scenario/context (AO2) and there are more than two marks available it would be prudent to ensure there is more than one link to contextualise any response given.

##### Comments on individual questions/sections

- Q.1** Whilst some candidates were able to adeptly describe 'event sampling' and whilst the attempt rate for this question was high (93.2%) this was the lowest attempt rate for any question on the whole paper, suggesting that some candidates had no idea what event sampling was. A number of candidates used their 'psychological imagination' to respond to this question and referred to attending events such as 'festivals' to gather participants for research. The mean mark was one but, where candidates did respond correctly, they frequently scored full marks. Where this did not happen, and the candidate was able to accurately describe event sampling as a form of observational sampling technique, they failed to gain the final mark due to a lack of appropriate example.
- Q.2** This question was attempted by almost every candidate (98.2%) and candidates on the whole did well (mean score of 2.3). Almost all candidates who attempted were able to score at least one mark for partially explaining the term 'deception'. Those who failed to score more than two marks did so because they lacked a second example from Milgram's research, or they failed to fully explain the two aspects of deception. This question essentially looked for the ideas that deception can be passive (omitting information) or active (lying or misleading participants). Most explained active deception (this specific terminology was not required to gain the marks) and related this to Milgram. Some went on to describe passive deception, or to give a second example, but rarely did a candidate do both and so they were capped at 3/4 marks.

- Q.3** There were some excellent examples of responses to this question. It was pleasing to see that the majority of candidates had a good grasp of research methods terminology and were able to show it off in their responses (e.g. discussing order effects, control of participant/confounding variables and specific validity issues). The majority of candidates gave a 'reasonable' evaluation of repeated measures, showing depth and range in their answer but not in equal measure. As this question requested strengths and weaknesses, those who only provided one aspect were limited in the marks they could achieve. As you would expect for this type of question the standard deviation was higher (2.0).
- Q.4** This question, and Question 7 (a), required candidates to explain terms related to location of research. A significant number of candidates used examples in their response to Q.4 which were related directly to past papers (the 2016 scenario about smiling in the classroom was a common and creditworthy example). Over 99% of candidates attempted this question and most were able to gain some marks (mean 2.3). Most responses gained either two or three of the four marks available, either through lack of relevant example or because their explanation was not thorough enough. Candidates were not required to evaluate the location in their answer.
- Q.5 (a)&(b)** Attempt rates for these questions were high (98% and 98.3%). Almost all candidates were able to correctly define secondary sources in (a) (mean 0.8); (b) was less well answered. In Q.5 (b), many candidates claimed that secondary sources are unreliable, when there is no reason to assume this is the case. Moreover, where candidates suggested secondary sources were outdated, this was only credited where an appropriate justification could be given, as one cannot assume all secondary sources will be outdated. Where candidates did offer a correct response, they frequently often only scored one of the two marks available because they failed to explain why the point given was a disadvantage. For example, they may identify that 'secondary sources may not suit the aim of the existing study', but then fail to briefly explain why this is a problem to gain the second mark.
- Q.6 (a) (i)** Almost all candidates were able to select a method that could be used to collect participants' responses in the given context (mean 0.8). In Q.6 (a) (ii) almost all candidates gave an appropriate justification (mean 0.8) but failed to contextualise for the second mark. As a general rule, in scenarios questions, if candidates are unsure whether they need to contextualise or not, they should do so. They will not be penalised if they contextualise, and it was not required by the question.
- (b)** Levels of data have always been a trickier aspect of the specification, but it was pleasing to see that a much larger proportion of candidates were able to respond correctly to this question than similar questions on past series. Almost all those who attempted this question identified a 'level of data', even if the one they chose was incorrect. This type of question is one that has clearly benefitted from the principle of 'practice makes perfect'.

- (c) (i) Whilst almost all candidates identified that the correlation was positive, a significant proportion either omitted the strength of the correlation, or misunderstood the meaning of the question and gave a positive evaluation point (advantage). This meant the mean out of a possible two marks was 1.3. In part (ii) the average mark was 2.2. Candidates are still giving inferential conclusions, despite the fact that we are unable to fully credit these at AS level. Where candidates superficially gave an inferential conclusion but then related it specifically to the data, full marks were awarded in the spirit of positive marking. To gain full marks the candidates need to use the 'data from the diagram'. Many failed to do this limiting them to two out of the three marks.
- (d) Well answered, lots of correct responses, but contextualisation is still missing from far too many responses. Some candidates gave detailed weaknesses but did not contextualise and so we could not credit full marks. This meant the mean score was 1.0, which is frustrating as the majority of candidates knew the answer but fell at the final hurdle.
- (e) This is becoming a firm favourite of candidates. A mean mark of 3.6 meant that it was fairly accessible to the majority of candidates. Candidates should be to show their working for their calculations of the mean and  $n-1$ , where a few candidates did all calculations correctly but failed to include one or both of these aspects and lost a mark as a result. Where an error is made in the calculations we continue to mark positively, so that one error early on is not carried over to the next step in the calculations. This can mean a candidate gets the wrong final answer but could still get 5 of the 6 marks available assuming all other calculations are present and correct based on the numbers they have provided.
- Q.7** (a) Whilst this question was quite similar in style to question 4, except an example was not required, it was much better answered. It had the highest attempt rate on the paper (99.7%) and almost all candidates were able to gain at least one mark out of the three available. Ideally, it would have been nice to see some more research methods terminology, to lift the quality of responses, as the majority talked only about control and failed to explain the impact that this has on the standardisation of procedures (replicability) and extraneous variables etc. However, overall this question was well answered.
- (b) This is an example of where the number of marks available should give an indication of the depth of response and detail needed within contextualisation. Whilst almost all candidates were able to explain how a self-selected (volunteer) sample is collected, the mean score was 1.2, indicating that contextualisation is lacking in the majority of responses. Contextualisation was credited for talking about advertising 'in the university' or asking for 'adults who slept 6-9 hours' on average to come forward. Any links to sleep or the location (university sleep lab) were creditworthy.
- (c) There were generally two subtypes of responses to this question. Candidates tended to either gain full marks, or no marks. Those who gained no marks, but responded, often did so by referring to the validity of the sample/sampling technique. As the question asks about why 'sleep studies lack external validity' referring to population validity is not creditworthy, as it is not specific to the method or location of a sleep study (one might argue population validity is a problem in any study, it is not specific to sleep studies).

Candidates were credited by talking about the location of 'sleep' being invalid, or any other factor that could have impacted sleep that was not mundane, e.g. they were aware they were being watched/monitored and so could not sleep as usual, they are sleeping in a sleep lab not their own beds, which may be more/less comfortable meaning the sleep more/less than usual etc.

- (d) (i) A surprising number of candidates fell at the first hurdle and did not select a histogram, with the most common incorrect answer being 'bar chart'. Even fewer still gained marks in (d) (ii) as for the minority who stated that the data displayed was continuous, they failed to contextualise (time slept is continuous data) and so gained no marks. This question was the least accessible on the paper with the lowest facility factor.

**Q.8** (a) Both parts (i) and (ii) were well answered with the highest levels of accessibility (facility factors). The mean score for both parts was 1.8 out of the possible two marks for calculating the mean and median scores from the data provided.

- (b) When asked to write a directional hypothesis, candidates made a reasonable attempt, with some misinterpreting the scenario. Candidates generally scored one out of two (mean 1.1), where only one condition was identified e.g., they said belief ratings would increase, but didn't identify a comparison or they compared the beginning and end of the study but failed to include the measure of belief. It would be best practice to ensure both conditions of the I.V. are included and that the D.V. is operationalised, thus even if operationalisation is not required, they will have included all elements required of a hypothesis.

- (c) This question had the second lowest attempt rate, but for those who did attempt most gained one mark (mean 0.8). The second mark was most often lost because of a lack of contextualisation, so quite complex and creditworthy arguments were made that reduce demand characteristics (single/double blind procedures, use of deception etc.), but then they were not applied to the scenario, which is a real shame. A few misunderstood the purpose of the research (to see if repetition of fake news makes it more believable), suggesting the use of independent groups instead of repeated measures, which in the context would not make sense.

- (d) This was one of the best answered questions on the paper with a 98.4% attempt rate and a mean score of 1.6. Almost all candidates got one mark and the majority contextualised to gain the second mark. This question was an example of what candidates can do when responding to all aspects of the question.

- (e) Many were able to give reasonable explanations of the role of an ethics committee, but a large number just explained ethical guidelines that they might use, limiting their score to 1 (mean 1.5). This is the first time this aspect of the specification has been tested. Candidates would benefit from thinking about the processes involved in psychological research from inception through to completion of the study, and how the committee fits in with this. There were rare glimpses of genius, mentioning their role in socially sensitive research, or talking about how the committee might suggest amendments to proposals, but these were in the minority, possibly because this was near the end of the paper where candidates might well be 'running out of steam'.

## **Summary of key points**

Candidates are to be commended on their attempt rate; very few scripts had unanswered questions.

Areas of strength include descriptive statistics, experimental design and standard deviation. Areas of improvement include the need for contextualisation, to allow access to full marks, through application of responses to the scenario. Candidates should use the number of marks available as a guide for how much detail is needed, particularly in extended writing responses (4 marks and higher).

Some evidence of increased understanding of research methods terminology on previous series is evident and centres are to be commended for their and their candidates hard work and efforts.

# PSYCHOLOGY

## General Certificate of Education

Summer 2022

### Advanced Subsidiary/Advanced

#### UNIT 3 - Implications in the Real World

##### General Comments

There were some outstanding answers to many of the questions, with candidates demonstrating an impressive depth and range of knowledge. Specifically, the knowledge of research studies and use of key terminology was very impressive.

Most candidates seemed well prepared for this unit and have refined the skills needed to access marks across all assessment objectives (AOs). AO1 stood out as the best prepared for AO, with some impressive illustrations of knowledge and understanding of theories/therapies across a range of behaviours. There were clear improvements on standalone AO2 questions, with many candidates showing a clear application and accessing top marks. Additionally, there were improvements seen in AO3 preparation - with candidates showing a thorough level of evaluation. However, at times there was a lot of unnecessary describing in some AO3 questions which was not creditworthy.

With regards to the combined AO questions – some candidates were not able to write appropriately for the AO balance required, often writing far too much for one skill and neglecting the other. Scenario questions were a particular concern with many candidates merely name dropping rather than engaging with the scenario.

Unfortunately, the controversies questions were generally not completed well this year. Candidates were not using much research evidence or included lengthy outlines of the research procedures rather than using the study to address the controversy; the best responses did this and used the key words in the questions to shape their response. As with previous years there was some relying on pre-planned answers that missed the focus of the question and instead tended to be a generic 'evaluate culture bias/ethical issues' answers. It was also evident from some responses that a lot of candidates did not read the question carefully.

##### Comments on individual questions/sections

###### Q.1 Addictive behaviours

- (a) This question had a mean score of 5.5. This question required both AO1 and AO3, however sometimes AO3 was left out. Some candidates also struggled to gauge how much to include for 5 marks, often writing far too much and including conclusions which were not necessary on this lower mark answer. 'Peer pressure' and 'role of the media' were most commonly chosen. At the highest level, individuals were able to discuss the explanations in detail with examples of research and displayed an understanding of addiction within psychology.

Many used Bandura's Social Learning Theory when discussing 'peer influences' and how alcohol and smoking were glamourised in films for the 'role of the media'. There was some confusing of social explanations and individual differences explanations and at times generic outlining of social explanations with weak or no links to addiction. However, there were some exceptional responses who not only described and evaluated explanations but used research studies to add depth.

- (b) This question had a mean score of 5.9. This question also required both AO1 and AO3, however, the same issues arose as with Q.1 (a) in that sometimes AO3 was left out and some candidates struggled to gauge how much to include for 5 marks. Stronger answers tended to explain how the therapy modified addiction including links to Antabuse and/or rapid smoking and used key words to explain the process through conditioning. With regard to AO3, stronger answers included fewer points that were explained and clearly linked to aversion therapy for addiction as opposed to lots of generic evaluation.
- (c) This question had a mean score of 2.6. AO2 skills were being assessed here. The focus needed to be on how biological explanations could modify addictive behaviour. This was done better this year with most candidates discussing how the dopamine hypothesis is linked to agonist and antagonist substitution and its impact on modifying addictive behaviours. Some candidates tended to just describe how the therapy worked as opposed to applying it to the biological approach, which did not gain credit. Some answers merely mentioned a biological explanation before going onto the method of modification, which would receive little marks.

## **Q.2 Autistic spectrum behaviours**

- (a) The question required the display of AO1 skills, to describe any two social explanations of autistic spectrum behaviours. For some candidates, there was some confusion seen between the social explanation, emphasising-systemising theory and the individual differences explanations, Theory of mind and Weak central coherence theory. Most candidates discussed empathising-systemising theory and the refrigerator mother, at times with one response was better than the other. Many students were able to explain these with the required depth and range with clear links to autistic spectrum behaviours. However, where the refrigerator mother explanation was used there was a lot of focus and discussion on Bettelheim's book.
- (b) Unfortunately, the combination of application and evaluation was clearly an issue for some candidates struggling to evaluate a method of modifying autistic spectrum behaviours and applying this to the scenario. Rather there was a lot of simply describing the method of modification with reference to the scenario or evaluation with little to no reference to the scenario. A minority of candidates were able to apply the evaluation of RDI or PECs the scenario by linking research, social implication or validity discussions to concerns from Dion's teachers/parents, Dion's struggle to communicate or his struggle with social situations. There was also evidently an issue with depth and range of AO3 and inclusion of conclusions.

### Q.3 Bullying Behaviours

This was the least popular option on Section A, with only 0.5% attempt rate for (a) and (b).

- (a) The mean mark was 6.7 for this question. AO2 was needed in this question by linking the Olweus Bullying Prevention Programme to the scenario- the best answers identified the ways of using Olweus Bullying Prevention Programme for online bullying (e.g. due to anonymity of online platforms its often hard for students who are cyber bullied to respond, O.B.P.P. tackles bullying at a number of levels and bully can take place in different places due to online platforms). Disappointingly, there were a number of candidates that did not do this and instead gave a generic outline of how O.B.P.P. works which was classed as AO1 only. The AO1 was done very well, overall, on this question with top band answers explaining the basis of O.B.P.P. and its different levels and how it hopes to reduce bullying. They also drew on research to illustrate working. Lower band answers had a lack of depth and range for 10 marks and missed the focus on how the programme reduces bullying.
- (b) This was an AO3 question that required evaluation, unfortunately, a few candidates outlined the individual difference explanations instead, giving an uncreditable answer. There were also a few candidates that only included one explanation instead of two and some forgot to add a conclusion. The best responses discussed the strengths and weaknesses of both explanations by making reference to specific research, discussing key debates such as the usefulness of the explanation and comparison with explanations.

### Q.4 Criminal behaviours

- (a) Generally, this question was done very well across the board and some impressive answers were seen. This was a 5+5 mark AO1 question which asked for one biological explanation and one individual differences explanation. For the biological explanation, most candidates discussed the role of the amygdala or inherited criminality, including lots of key terminology and studies/research in the field. At times, some candidates did not link the explanation to criminal behaviour. With regards to the individual differences explanation, Eysenck's criminal personality was most often discussed, but a few candidates failed to move their depth and range beyond a simple discussion of extraversion, neuroticism and introversion personality types and failed to link them to criminal behaviour.
- (b) This question assessed both AO2 (5) and AO3 (10). Again, the application element was not always made apparent, where some candidates failed to move beyond repetitive statements, 'so will reoffend'- to apply to the scenario. With regards to the evaluation, some good answers were seen. Inclusion of research evidence demonstrating its effectiveness, ethical/social implications and even comparison with alternative methods or types of crime. The best answers balance depth and range appropriately for 10 marks and included a conclusion which added something valuable to the answer.

Some candidates did, however, miss the requirement to include both effectiveness and social implications. Unfortunately, a similar issue was seen again on this answer, where some candidates missed the command word and instead outlined the methods of modifying criminal behaviour rather than evaluate.

#### **Q.5 Schizophrenia**

This was the most popular option in Section A, with an attempt rate of (a) 73.6%; (b) 73.3% and (c) 69.1%.

- (a)** This question asked for the description of the characteristics of schizophrenia. Most candidates had excellent knowledge of the characteristics and were able to identify and describe them in some detail. However, this was a 5 mark question and many candidates wrote a response sufficient for a question worth 10 marks, describing at some length most of the key characteristics. Weaker answers tended to list the characteristics without explaining their meaning or a specific example of how it could be seen.
- (b)** This question had a mean score of 6.5. This question required both AO1 and AO3, however many candidates left out AO3. Some candidates also struggled to gauge how much to include for 5 marks, often writing far too much. For AO1, most used the dopamine hypothesis and explained in detail the key aspects involved with reference to a variety of key words. Where genetic inheritance was used, at times research was included but it was not clear if this was being used to illustrate or evaluate. For AO3, some candidates were well prepared to use AO3 points, and where this knowledge was used well these points were detailed and linked to both the explanation and schizophrenia. Lower band answers tended to use generic evaluation.
- (c)** This question had a mean score of 5.0. This question required both AO2 and AO3. Unfortunately, a large proportion of candidates purely explained CBT and how it works as opposed to evaluating it. For those who did evaluate, there was plenty of opportunity to apply to the scenario with a variety of concepts given regarding Siwan. However, answers tended to lack threaded AO2 and instead there was a lot of 'name dropping' in an attempt to apply. Also, some candidates missed the fact that AO3 would account for only 5 marks and tended to include a large range of points, possibly indicating that they believed it to be 10 marks for AO3.

#### **Q.6 Stress**

- (a)** This question asked for the description of the characteristics of stress. There was a large variety of ways this was approached, i.e. stress as a response; a stimuli; types of stress etc. The better answers were able to balance between depth and range, given detail and illustrated with examples. Lower band answers were those that tended to give a long list or included a very restricted range and then did not go into much detail.

- (b) This was an AO3 only question, which required two social psychological explanations. Most candidates focussed on daily hassles and life changes. Higher band answers tended to give specific evaluative points that were clearly in context to the explanation that it was evaluating and give a clear conclusion that added something to the answer. Lower band answers tended to use generic AO3, and it was difficult to see which explanation they were evaluating. Again, in this AO3 question a notable proportion of candidates described explanations. In addition, a few candidates also unfortunately confused them with the individual difference explanations.
- (c) There were some improvements seen on this question in comparison to previous years. The better answers giving a clear structure of a brief outline of the biological explanation and then linking these concepts to how, for example, beta blockers then work. However, the common mistake was still seen in this answer with some not applying but describing or just mentioning 'S.A.M.' before giving a detail outline of how beta blockers work.

## Section B Controversies

After seeing a cumulative improvement in these types of questions in previous years, unfortunately this was not the case this year, with a lot of candidates seemingly unprepared for the demands of these questions.

### Q.7 Culture Bias

This controversy was less frequently attempted with a 26.0% rate. The mean mark was 11.5 and there were a few disappointing things seen with this answer, notably pre-learnt essays that did not address the question; long descriptions of studies without engagement with the question; answers with no reference to research/theories or approaches at all. Some answers tended to just outline different biases and failed to link these to how they affect our understanding of behaviour, missing the key point of the question.

In addition, too many answers tried to include topics from other controversies to pad out the answer. Answers often included different genders or sexualities. While these may be cultures in their own rights the discussion was weak which made it obvious that the arguments came from sexism. The discussion was generic in many cases and did not relate to 'improving understanding'. Many answers did not address both sides of the argument and instead was very one-sided, focussing on how cross-cultural research does not improve our understanding.

Better essays had selected studies/evidence from across the course to support and develop arguments that linked to the question. Most often research such as Buss (1985), Myers and Diener (1995), Kohlberg (1968) and Milgram (1963) were discussed. They discussed studies that demonstrated these biases which led to discussion of why this could be an issue and how it can affect our understanding of behaviour. This would have led to them answering the question set. The question also asked for social and historical considerations which were not always addressed.

## **Q.8 Ethical costs of conducting research**

Similar to Q.7, there were a lot of pre-learnt essay seen that did not address the question, answers with long descriptions of studies without engagement with the question and answers without any reference to research or theories. It is disappointing that students are not including much evidence appropriately to answer the question. For this question, there was a large number of answers that did not address the question, instead misinterpreting it to be a generic 'evaluate ethical issues' question. There also needs to be more emphasis on having a balanced argument and covering both sides of the debate. The question required the candidate to 'discuss the importance of using ethical guidelines appropriately to benefit society' some answers did not refer to the 'importance' or 'the benefit to society' element. Where they have answered the question, it has mostly been done well. These answers used research/theories/approaches to indicate that ethical guidelines were either important or a cost to society. For example, by using Watson and Raynor and suggesting that guidelines would have been important to protect Little Albert as his well-being deserves the same consideration as the benefits to society. However, these guidelines could come at a cost, as this study would not be possible to conduct in today's society, but it provided beneficial information to society, such as treatments, which would be missed today.

### **Summary of key points**

Candidates would benefit from practicing responses to a variety of different question styles.

Better preparation for controversies essays, specifically with a drive to answer the question set using relevant studies/approaches/theories, rather than just producing a generic, pre-learnt response.

Practise on balancing the depth and range of 5 mark questions, and the application of knowledge in answers to the scenarios in questions with combined assessment objectives.

# PSYCHOLOGY

## General Certificate of Education

Summer 2022

### Advanced Subsidiary/Advanced

#### UNIT 4 - Applied Research Methods

#### General Comments

This year there was an impressive number of candidates applying their knowledge and understanding of research methods to the novel scenarios. It was also clear to see that most candidates had a good understanding of their personal investigations, and it was encouraging to see some innovative ways of conducting the experiment on helping behaviour. The time constraint for this paper is deemed appropriate with most candidates attempting every question. The main discriminator on this paper remains application, as this paper is applied research methods. The AO2 focus was lacking in places, particularly in questions focusing on the personal investigation on the interview on social media. Another discriminator was the use of appropriate terminology.

#### Comments on individual questions/sections

- Q.1 (a) (i)** 98.8% of candidates attempted this question with a mean mark of 1.3. Most candidates were able to operationalise their I.V. The better answers clearly operationalised the level of the I.V. and each condition. For example, the level of I.V. is position of authority operationalised by teacher (authority), student (no authority), or the level of the I.V. is gender operationalised by male, female, non-binary. The weaker answers just stated the conditions without the level of the I.V. for example the I.V. was operationalised by male and female or teacher and student. One concern was that there seemed to be a few candidates who had conducted an experiment on sharing behaviour, and not helping behaviour.
- (ii)** 97.4% of candidates attempted this question with a mean mark of 0.9. There were impressive responses to this question with candidates able to fully justify why they operationalised their I.V. in the way they discussed in part (i) with most referring to allowing to test the effect that their I.V. had upon helping behaviour. However, there were a number of candidates who failed to fully justify, and this is represented in the facility factor of 44.9%.
- (b) (i)** 98.3% of candidates attempted this question with a mean mark of 1.3. A popular choice was the Nickell's helping attitude scale. The better answers operationalised this by saying what the helping score was out of, for example, 100. Other impressive answers were those that consisted of timing in seconds how long it took a passer-by to pick up dropped papers/books. The weaker responses lacked full operationalisation for example referring to simply the score on the helping behaviour scale.

- (ii)** 97.1% of candidates attempted this question with a mean mark of 0.9. The better responses justified the way of operationalising their D.V. as it allowed them to gain a level of data (nominal/ordinal) that meant they could quantify their data and present it graphically, for example, in the form of a bar graph. The weaker responses justified their operationalisation of the D.V. as it allowed quantitative data which made comparisons easy and lost marks on the full justification.
- (c)** 98.5% of candidates attempted this question with a mean mark of 1.4. A large proportion of candidates were able to correctly identify an extraneous variable. The better responses referred to referencing keeping all conditions the same in the way of dealing with the issue. Although there were some responses that referred to confounding variables, such as the participant's mood on the day, most candidates referred to appropriate extraneous variables and seemingly understood that this meant affecting all participants equally.
- (d)**
- (i)** 96.7% of candidates attempted this question with a mean mark of 0.8. The better answers were able to justify their choice of graph with reference to their investigation. A popular choice was the bar graph due to the data being nominal and specifying the categories of the data. The weaker answers would just state a reason for the graph with minimal reference to their investigation. Appropriate terminology was seen in these responses such as the level of measurement of their data e.g. nominal, ordinal etc. and reference to continuous/non-continuous data. Some candidates referred to a scattergraph which was deemed inappropriate due to the personal investigation being experimental.
- (ii)** This question wasn't as well attempted as the other questions, and this is evidenced by the attempt rate of 89%. The mean mark for this was 0.5, suggesting that candidates, on the whole, struggled with this question. What was clear to see was that a lot of candidates didn't understand what the term descriptive statistics was as there were many responses that described the findings of the investigation. The better answers were able to justify their descriptive statistic with reference to their investigation. Answers that simply stated statistics didn't gain credit and the answers that described their findings didn't gain credit.
- (iii)** 91.0% of candidates attempted this question with a mean mark of 1.7. This question is a prime example of the link being a discriminating factor in the mark awarded. A high proportion of candidates were able to fully justify their choice of inferential statistic with no link to their research which is represented by the mean score. The better answers fully justified their choice of inferential statistic with clear links to their research. The key point here is that reference to the personal investigation is key to achieving the higher marks. Some responses referred to descriptive statistics and therefore were not creditworthy.
- Q.2 (a)** 99.2% of candidates attempted this question with a mean mark of 1.7. This question was both well attempted and answered very well with most candidates being able to correctly identify two features of the participants in their research.

- (b)** 98.5% of candidates attempted this question with a mean mark of 0.4. Unfortunately, this question was one of the questions that really lacked reference to the interview on social media and therefore wasn't answered very well on the whole which is represented in a lower mean score and facility factor. Although this question was attempted by most candidates, to gain any marks for this question there had to be a link to the investigation and a lot of answers simply explained a way of assessing reliability with no reference to their personal investigation. The better answers were able to justify their way of assessing reliability with a clear link to the research and a popular choice of this was inter-rater reliability with reference to the percentage of consistency between the interviewers. Some responses referred to a questionnaire study and not an interview, which was not creditworthy.
- (c)** 99.3% of candidates attempted this question with a mean mark of 0.4. Again, this was another question that was attempted by most candidates but was poorly answered and that is due to the lack of the reference to the interview on social media. To gain any marks of this question there had to be a link to the investigation and a lot of answers simply justified the type of interview with no reference to their personal investigation and therefore could not receive any credit. The better answers referred to elements of their interview such as referencing their questions to justify why they used that particular type of interview.
- (d)** 99.2% of candidates attempted this question with a mean mark of 1.5. On the whole, this was answered well with most candidates able to give an appropriate conclusion of their investigation. Answers that stated more than/less than without a comparison were not as good as answers that clearly stated the conclusion by giving the comparison, for example, the more time spent on social media the less amount of sleep per night, or the older generation spent less time on social media than the younger generation were better than answers such as the older generation spent less time on social media. Those who found no difference/correlation were also perfectly accepted valid conclusions.
- (e)** 99.2% of candidates attempted this question with a mean mark of 1.7. Most candidates were able to discuss one improvement that could be made to their investigation. The discriminator here for the top band was the terminology. Noticeably, there were some candidates who referred to making the study 'more accurate' instead of more valid and some candidates referred to a more representative sample as having better 'ecological validity' as opposed to 'population validity'. The better answers used appropriate terminology such as population validity/internal reliability/external validity etc.
- Q.3 (a)** 97.8.2% of candidates attempted this question with a mean mark of 1.1. Most candidates were able to access 1 mark for this question by identifying the independent groups design represented by the mean score. The better answers referenced the research and explained the two conditions of the research being homework three times a week or no homework.
- (b)** 99.5% of candidates attempted this question with a mean mark of 0.8. This was attempted by most candidates and the differentiator within this question was the application to the scenario; those who didn't apply their answer to the scenario received no credit. Within this question it was also clear that some candidates described the sampling method in relation to allocation of groups rather than the sample as a whole and this was not credited.

- (c) (i) 99.7% of candidates attempted this question with a mean mark of 4.0. This question provided some excellent responses. Candidates were able to discuss three ethical issues and apply them to the piece of research they were given in the novel scenario. Unfortunately, some candidates referred to debriefing as an ethical issue and this wasn't creditworthy as it is a way of dealing with an ethical issue. Those answers that lacked any application to the research were not creditworthy.
- (ii) 99.5% of candidates attempted this question with a mean mark of 4.1. This question, again, provided excellent responses. Candidates were generally able to explain ways of dealing with the ethical issues that they discussed in (c) (ii) the better answers clearly explained the way of dealing with the issue through specific way of dealing with the issue whereas the weaker answers simply stated the ways of dealing with the issues.
- Q.4** (a) 99.7% of candidates attempted this question with a mean mark of 1.5. Generally, this was well answered with most candidates being able to state an appropriate conclusion using the data given in the novel scenario. The better answers used the data to reach the conclusion that blue was the preferred colour preference of toy as the infants spent more time with the blue toy than the pink and green toys.
- (b) 96.4% of candidates attempted this question with a mean mark of 0.2. Unfortunately, this question was the most poorly answered across the paper with a lot of candidates not being able to correctly identify the level of data as ratio data. This question is therefore the least accessible to candidates with a facility factor of 19.5. Many attempted this question but were unable to identify the correct level of data and there were a wide range of answers not relating to level of data at all (e.g. mean, time, seconds).
- (c) (i) 96.6% of candidates attempted this question with a mean mark of 0.6. Although this question was well attempted there seemed to be a lot of candidates who were unable to identify the experimental design as repeated measures. Confusion about the term experimental design was apparent as reference to research methodology and inferential statistics were made in this response.
- (ii) 95.2% of candidates attempted this question with a mean mark of 1.1. The better answers referred to participants taking part in all/both conditions of the I.V. The weaker responses referred to participants taking part in different conditions.
- (d) 99.8% of candidates attempted this question with a mean mark of 4.8. This was a well attempted question, and most candidates were able to receive some credit on this. The better responses discussed three or more detailed improvements to the research and how it would improve the research. The basic discussion points lacked detail or range in the answer. Candidates should be careful when referring to changing the sampling method to random sampling that they avoid saying that it will make it more representative as this is not always the case; better responses to this question discussed how a random sample would make the sample less biased.

## Summary of key points

It was clearly evident that a lot of work went into the personal investigations and candidates should be commended for their efforts conducting these. One important point to make here is that the investigation should be fit for purpose as there were some responses that referred to an experiment on sharing behaviour in Q.1 instead of helping behaviour; some responses referred to a questionnaire study on social media, instead of an interview on social media.

The attempt rate was high across all questions.

This is an applied research methods paper, and the key is application as half of the marks on the paper are assessing AO2 skills. There are candidates who clearly understand the content but are lacking the ability to sufficiently apply their knowledge and understanding which is disadvantaging them, particularly in Section A.

There were clear gaps in candidates' knowledge of level of data and descriptive statistics.

Another area of improvement should be a focus on appropriate terminology. There is seemingly an increasing number of responses who are using the word 'accuracy' to refer to validity which isn't appropriate and 'ecological validity' instead of 'population validity'. A focus on appropriate terminology would be beneficial to candidates and improve the quality of answers.



WJEC  
245 Western Avenue  
Cardiff CF5 2YX  
Tel No 029 2026 5000  
Fax 029 2057 5994  
E-mail: [exams@wjec.co.uk](mailto:exams@wjec.co.uk)  
website: [www.wjec.co.uk](http://www.wjec.co.uk)