PSYCHOLOGY

Overall grade boundaries

Higher level

Grade:	1	2	3	4	5	6	7
Mark range:	0 - 9	10 - 20	21 - 29	30 - 43	44 - 55	56 - 69	70 - 100

Standard level

Grade:	1	2	3	4	5	6	7
Mark range:	0 - 10	11 - 22	23 - 33	34 - 46	47 - 58	59 - 70	71 - 100

Higher level internal assessment

Component grade boundaries

Grade:	1	2	3	4	5	6	7
Mark range:	0 - 2	3 - 5	6 - 9	10 - 13	14 - 17	18 - 21	22 - 28

The range and suitability of the work submitted

It seems that the level of guidance was generally good, with a large number of centres submitting appropriate topics to investigate although there were issues of experiments using ingestion which is not allowed according to the psychology guide. The research topics were as usual related to cognitive psychology and this approach is more likely to provide good results at this level of education. There were, as usual, a number of replications of Bartlett and experiments on the Mozart effect. Research based on Bartlett typically use time delay as IV and repeated measures design. It is not possible to randomly allocate participants to the experimental conditions or counterbalance in this case where all participants basically have the same experience. The vast majority of candidates were aware of ethical issues and included a copy of informed consent in the appendices.

In general, the weaker reports shared the following characteristics.

- Weak and imprecise explanation of background research in the introduction (this affected the discussion section as well).
- Results were not always clearly related to the aim of the study and inferential tests were not properly justified or applied.
- Discussions were superficial and did not discuss own results in the light of the background research.
- Referencing was poor.

Candidate performance against each criterion

The main problems seem to be in the introduction and discussion sections. The introductions were in some cases very well written with a clear focus but it generally seems to be difficult for some candidates to have an exclusive focus on the relationship between particular research and the candidate's own research hypothesis. The introduction is important and the background research should be explained and analysed in sufficient depth (for example, aim, procedure, findings) to allow for formulation of a clear research hypothesis in the introduction. If this is done, discussion of own results in the light of the background research in the discussion section can be better developed.

Criterion A – Introduction

The analysis of background research was often superficial and based on a summary of a study found on a website or a study guide. This gave some problems in terms of justifying and formulating clearly operationalized hypotheses. In some cases, only one background study was cited and this is not sufficient at HL. However, a considerable amount of candidates were able to create a logical link between the background research and their own hypothesis. For some candidates, it is difficult to state a clear and operationalized experimental hypothesis. It is important that the introduction is clear and focused on relevant background research so that it logically leads towards the candidate's own research hypotheses and that the background research is explained and analysed in sufficient depth (for example, aim, procedure, findings) to allow for formulation of a clear research hypothesis in the introduction and to stimulate discussion of own results in the light of the background research in the discussion section. It should be noted that it is not required to make an exact replication of an experiment: a partial replication with only two conditions is enough.

Criterion B – Method: design

Most candidates demonstrated knowledge of experimental design but choice of design was not always properly justified, e.g. by reference to strengths and limitations of respective designs (repeated measures and independent designs) in relation to their own experiment. The ethical guidelines were mostly addressed in the design section but sometimes in procedures instead, which is fine. What is essential is that ethical procedures are addressed appropriately somewhere in the report. Most candidates included an informed consent from participants. As in previous years there were candidates who used many experimental conditions but it is recommended to use two conditions only as the IB experiment is a *simple* experiment.

Criterion C – Method: participants

There were often problems in identifying relevant characteristics of the participants. It may be difficult to define relevant characteristics in relation to a specific study but candidates could always choose characteristics such as age, sex, number of participants and nationality. The target population was not always appropriately identified and at times candidates made a point in saying that the target population was the participants. This is not likely to earn full marks in criterion C. All samples were based on a student population in the candidate's own school. Sampling technique was mostly clearly identified as opportunity sample or self-selected sample and it was mostly explained or justified. Both ways are fine but if it is not done, one mark is lost here. Some candidates did not specify how they allocated participants to the experimental conditions but were not be penalized for this since it is not mentioned in the assessment criteria.

Criterion D – Method: procedure

There were quite often problems where the description of procedure was not sufficiently detailed to allow replication (especially with reference to material in the appendices or information on how participants were randomly allocated to conditions) or the procedure included a lot of redundant information (e.g. reference to research before the experiment). Many candidates lost one mark because they wrote that they randomly allocated participants to the conditions but forgot to mention how this was done or they simply used the word "random" in the wrong way. Some did not include the stimulus material (e.g. reference to a video clip on YouTube or the kind of music used in the experiment). Many candidates used bullet points in the procedure section but there is a tendency to write too shortly and in some cases this made it difficult to replicate the study.

Criterion E – Results: Descriptive

Most candidates described the results in a narrative form in the results section. Not all included standard deviation as measure of dispersion even if their data allowed it. The graphs were at times very poorly labelled or not labelled at all. Not all reports included tables. There were quite a few examples of candidates graphing the standard deviation besides the mean in a bar chart. This is not the right way to do it, as standard deviation would be graphed along the x-axis and not the y-axis. Only one measure of central tendency is required in descriptive results but many candidates included mean, mode and median. This is redundant, as only one relevant measure of central tendency is needed. Calculation of descriptive statistics should be in the appendices and not in the results section.

Criterion F – Results: Inferential

Most candidates used the correct inferential statistical test but a few did not. Some had problems calculating the test correctly and therefore made the wrong conclusions. The choice of test was not always justified. In some reports, there was no statement of statistical significance. The calculations were mostly in the appendices but they were not always complete and in some cases totally missing.

Quite a few candidates seemed confused about levels of measurement of data. Some claimed their data was nominal (and it was not) and therefore used the chi-squared test.

Criterion G – Discussion

Many papers had a short and superficial discussion of own results in the light of previous research but a long description of limitations of own design, which were not always relevant in terms of their own study but rather a more general list. In quite a few reports there was also reference to strengths of the methodology used, which is no longer relevant according to the current guide. One of the popular suggestions for modification was to use a random sample and more participants in future research, which does not really address relevant limitations of the candidate's experiment.

Criterion H – Citation of sources

Candidates did often not include all the references they mentioned in the introduction and many candidates did not use secondary referencing appropriately. There are still problems with references from the internet where candidates tend to think that the URL is enough and sometimes the background study could not be found in the reference section. Referencing did not often follow a standard way of referencing, such as APA.

Criterion I – Report format

Formatting was generally excellent but some candidates lost marks because they were missing parts of the abstract or missing parts of the appendices. A few candidates did not include a table of contents.

Recommendations for the teaching of future candidates

- Most important is to help candidates find appropriate background research and theoretical framework because such resources will enable them to analyse the background research in some depth. It is recommended to find relatively simple experiments to replicate. This would help candidates to make relatively simple experiments themselves based on real scientific experiments. It is much easier for candidates to replicate an experiment if they have access to readings about experiments or summaries of them so that they can read about hypotheses and other important details.
- When doing research based on Bartlett it is perhaps worthwhile to consider how the original findings could be used to make a true experiment instead of replicating Bartlett since that study was not a true experiment either. As for the Mozart effect, no study has ever been able to replicate the findings of the original study and the work presented in November was not able to do that either. Since the Mozart effect is controversial, teachers should perhaps consider discouraging experiments based on the Mozart effect.
- There should be more focus on the relationship between the aim of the candidate's study and the background research so that these can be integrated in the introduction and the discussion of the results. The background research should be analysed in sufficient depth in the introduction so that the aim of the candidate's own research is clearly justified and the experimental hypothesis should be clearly linked to background research. Candidates should be trained in how to properly operationalize the hypotheses (and the IV/DV).
- The design must have a clear description of the experimental conditions and teachers should ensure that there are only two conditions (either two treatment conditions or one treatment condition and one control) so that there is a possibility to compare the outcome of the manipulation of the IV on the DV in the two conditions. This is in line with the IB recommendations for simple experimental studies in psychology.
- Sampling should be done according to IB rules, i.e. identification of target population, then relevant characteristics of participants and description of sampling method as well as explanation of the use of the chosen method (or justification). Most candidates use a convenience sample but they should still explain (or justify) the sampling method. The number of participants in the experiment does not need to exceed 20 (independent design) or 10 (repeated measures design).
- In the descriptive statistics section, the graphs and tables should have proper titles. It should be emphasized that graphing the results is mandatory and that a table must be included. This section includes *summarized* data not raw data or individual scores. It is recommended that candidates don't include several measures of central tendency but only the one which is relevant for their data, and that a measure of dispersion is present.

- In the inferential statistics section, candidates should be careful in choosing an appropriate statistical test and justify why this test was chosen. It is relevant to relate choice of statistics to the level of measurement of data and the design of the experiment. It is also important to instruct candidates in how to report the results of the test in the report, e.g. how to state statistical significance.
- The explanation of the empirical studies and theoretical framework from the introduction must be referred to in the discussion section. New studies or theories should not be introduced here. Candidates should be trained in making a discussion section (perhaps by reading a couple of research articles to become familiar with the idea and style) and they should consult the checklist to be sure that all the IB requirements are met, e.g. in terms of discussing statistics from the results sections. Understanding of relevant limitations of own research and suggestions for modification should be tied together and it is not enough to say that a particular study should use random sampling and more participants to be better. The limitations should be explicitly relevant to the candidate's own experiment.
- It is generally recommended that candidates are familiar with the scientific research method, which includes references to previous studies and integration of these in their own research. Likewise, the use of proper background readings must be encouraged. It is recommended that candidates be trained in critical use of internet resources and how to reference those properly (e.g. using the APA).

Standard level internal assessment

Component grade boundaries

Grade:	1	2	3	4	5	6	7
Mark range:	0 - 2	3 - 5	6 - 8	9 - 10	11 - 12	13 - 14	15 - 20

The range and suitability of the work submitted

Overall the range and suitability of the work submitted was generally of a very good standard and only a few non-experimental studies were submitted this year. Popular research topics were mostly from cognitive psychology and these choices are suitable for candidates at this level of education. The majority of candidates were aware of relevant ethical considerations; they followed them while planning and conducting research and most included a copy of informed consent in the appendices. Some candidates did not obtain parental consent for participants under 16 years. In a few rare cases participants included small children (aged 10 to 11 years) as participants. Even with parental consent children so young should not be participants in experiments.

In general there was an increase in the overall standard of the reports. Also, reports from the majority of centres were appropriately marked by teachers. Procedural aspects of the experiments were generally well done; however, choice, justification and explanation of design, sampling method and data analysis tended to be weaker. Additionally, discussions tended to be written in a superficial and general manner and did not discuss their own results in relation to the original study.

Lower quality reports tended to reflect the following problems.

- Weak introductions were provided.
- In the results section, several problems were identified but most often no calculations for descriptive statistics were included and candidates provided no justifications for descriptive statistics.
- Discussion lacked relevant information and tended to be written in a very vague manner.
- References were not presented in a systematic and standardized format.

Candidate performance against each criterion

Criterion A – Introduction

In the majority of cases introductions had clear aims and relevant research well described. However, many candidates did not explain all relevant aspects of the original study clearly. Often, candidates tended to overemphasize procedural aspects of the study while omitting other relevant information about the original study (who the participants were, what sampling method was used, how the data was analysed, what were the findings). Also, although most candidates wrote an aim, many failed to state the aim clearly by indicating the IV and DV.

In a few rare cases the introduction provided an analysis of background research – in these cases there was a review of more than one study and some reference to related theoretical background. Often these introductions tended to be too long and provided information that wasn't necessary for SL reports

Weaker candidates failed to identify and explain the original study. General introductions providing definitions and explanations of psychological concepts (i.e. operant conditioning, Stroop effect) accompanied by theoretical accounts are not acceptable as they do not meet the requirements of criterion A. Also, candidates taking this approach cannot discuss their results in relation to the original study.

Criterion B – Method: Design

Most candidates correctly identified their experimental design but the choice of the design was frequently not justified. Justifications should be presented by stating why a specific design is a "good choice" for the specific topic and study conducted. Weaker candidates still have problems in understanding the difference between a method (experiment) and design (repeated measures, independent samples, matched pairs design).

Most reports included some description of the independent and dependent variable. However, operationalization of these two variables was not always clear. Also, some weaker candidates confused the IV and DV.

Ethical guidelines were usually clearly followed and evidence of this was provided within the report. Only a few candidates did not attach a blank copy of the consent form and debriefing letter in the appendices. Candidates from some centres are still unaware that parental consent is also needed where participants are under 16 years of age. In several cases some form of deception was used within the study. When deception takes place, some justifications should be included.

Criterion C – Method: Participants

Description of characteristics of participants usually included adequate information. All samples were based on a high student population usually coming from the candidate's school. Most candidates correctly identified the sampling technique but often forgot to justify the chosen method of sampling. In many cases explanations of sampling methods were not detailed enough. Weaker candidates tended to have problems understanding and applying the term "random" – some candidates tended to frequently describe opportunity or convenience sampling as a random sample.

Some candidates had quite large samples in their study – there is no need for a very large sample. A sample of 15 – 20 participants is an adequate size for a standard level study.

Criterion D – Method: Procedure

This criterion produced the least difficulty for candidates. In the majority of cases reports reflected clear, itemized and easily followed procedures. However, some candidates did not include all relevant materials which they used when conducting their study (standardized instructions, informed consents, lists of words, and debriefing notes were not always included in the appendices). In addition, some candidates included material in their appendices but this material lacked clarity (e.g. material photocopied in black and white although colour was an important factor – this meant that replication would be difficult). Candidates should also be encouraged to include information about the timing and location of their procedure.

Criterion E – Results

The results section varied considerably in quality of presentation and the information that could be obtained. At their best they were both clear and informative, but there were occasions when the results were omitted or gave so little information that they were hardly worth presenting. Presentation of tables and graphs needs to be clearer so that data can be readily understood by the reader. Graphs should have a title, legend, and a label for each axis.

In some cases candidates presented and explained findings that were not part of the original aim. Descriptive statistics in these cases did not match the choice of the IV and DV defined in the design section. Usually the descriptive statistics chosen were mean, median, mode, standard deviation and range. It is highly recommended for candidates to choose only one measure of central tendency (the one which is appropriate for the scale of measurement used when obtaining the results). In addition, an appropriate measure of dispersion should be included. The use of descriptive statistics should be justified or explained. There is no need for SL candidates to include a presentation and calculation of inferential statistics as examiners cannot award marks for this extra work.

In several weaker reports only raw data was included in the results section.

Criterion F – Discussion

Quality of the discussion ranged from superficially written with general evaluation of limitations to well-balanced with an appropriate conclusion and offering modifications and improvements for further research. Most candidates had difficulty relating their findings back to the original work stated in the introduction, and were seldom perceptive in criticizing their own research. Another common problem was that often some relevant information was provided in the discussion section but it wasn't fully explained by providing a clear connection

between the method applied in the study and the results obtained. Furthermore, related ideas for future research were often undeveloped.

In addition, some reports reflected that candidates had problems with presenting a relevant conclusion. Although there were some concluding remarks embedded within the report a final concluding statement was rarely seen.

Criterion G – Presentation

Presentation in many reports was in accordance with requirements of criterion G – a lot of reports were impressively presented and scored full marks for this criterion. Unfortunately, there are still reports in which the maximum number of words permitted is exceeded or in which no reference section is provided. In some cases, although standard citation methods were used there was a lack of in-text citations. Another common mistake was failure to label appendices and reference them in the body of the report. Also, raw data were sometimes incorrectly included in the results section.

Recommendations for the teaching of future candidates

- Teachers should emphasize the importance of actively manipulating an independent variable.
- Although an understanding of the meaning of "design" is important, this seems to be a problem in some centres. It would be best to think about the design of the experiment right after identifying and operationally defining the independent and dependent variable. Candidates should ask themselves how they want to actively manipulate the independent variable. If they decide to have two separate groups of participants and they present the independent variable to one group while the other group is a control, then they will choose the independent samples design. If they decide that they will take the same group of participants and once measure their behaviour without introducing the IV and the second time measure their behaviour when the IV is present then they will choose and justify the repeated measures design. When explaining their decision, candidates should explain in the report why the chosen design is a good choice.
- Candidates should provide a thorough explanation of the original study – this explanation needs to include not only the aim and the findings but also relevant procedural aspects of the original study. This information should be selected and presented wisely so that it can be used later in the discussion section when a comparison of the two studies is provided.
- All candidates should provide information on how they followed ethical considerations – this information can be addressed within two sections of the report (design, procedure). If participants are under the age of 16 a parental consent form is necessary and should be included in the procedure and documented in the appendices.
- A clear justification of sampling method should be included in the participants section. Candidates should also describe relevant characteristics of participants – the characteristics presented should be important for the specific experiment and also include information such as age, sex, nationality, knowledge of English.
- Only one measure of central tendency and one measure of dispersion should be included for criterion E. Candidates need to practise identification of optimal

measures depending on the type of data gathered. Teachers should encourage candidates to check their calculations, clearly write them and include them in the appendices. For full marks for criterion G, calculations should be included in the appendices.

Higher and standard level paper one

Component grade boundaries

Higher level

Grade:	1	2	3	4	5	6	7
Mark range:	0 - 5	6 - 10	11 - 14	15 - 20	21 - 26	27 - 32	33 - 46

Standard level

Grade:	1	2	3	4	5	6	7
Mark range:	0 - 5	6 - 10	11 - 14	15 - 20	21 - 26	27 - 32	33 - 46

General comments

Many candidates showed an impressive knowledge of psychological research and theory. Many candidates, however, seemed to have a rather formulaic response, addressing principles and evaluating studies when this was not required. This often led to a less focused response which did not always address the demands of the command term.

Many candidates wrote very long responses to the short answer questions, giving several examples to answer the question. If only one example is requested, only the first one is assessed. The rest of the response earns no credit and only limits the time for writing the extended response.

In addition, many candidates made use of asterisks, arrows and comments in the margins of the paper. This often made for confusing reading. Candidates should be reminded that they should not write in the margins of their exam papers.

The areas of the programme and examination that appeared difficult for the candidates

Candidates often did not focus on the demands of the question, including much information that was not relevant. Critical thinking in the extended responses was often limited or not linked to the requirements of the question. Several candidates also were unable to correctly identify research methods. For several candidates a firm understanding of research methods was lacking.

The areas of the programme and examination in which the candidates appeared well prepared

Overall, candidates showed a good range of understanding with a wide breadth of research to support claims. There were examples of outstanding levels of organization and analysis.

The strengths and weaknesses of the candidates in the treatment of individual questions

Section A

Biological level of analysis

Question 1

Many candidates were able to correctly identify two neurotransmitters and provided an explanation of their effect on behaviour. Some candidates identified a hormone; this received no marks. Several candidates failed to identify the link the neurotransmitter to human behaviour. Simply stating that animal studies can imply human behaviour was not sufficient.

Several candidates wrote out very long descriptions of the process of neurotransmission or they discussed the role of hormones as well. This was not the aim of the question and often led to lower marks as the response lacked focus. Candidates should also be careful of neurotransmitters that may also act as hormones. In these cases, marks were only awarded if the candidate identified the neurotransmitter functions, and not the hormonal functions.

The question ("Explain how **two** neurotransmitters affect human behaviour.") was slightly beyond the remit of the course but this was taken into account at standardization and the question is to be amended when the question paper is published.

Cognitive level of analysis

Question 2

Several candidates answered this question quite well. Sometimes the principles were not appropriate, or they were poorly stated. Candidates needed to outline the principle. Often this was not the case, but instead an example was used. If the example was explicitly linked to the principle, full marks were awarded. In cases where it was only implied, only mid-range marks were awarded. Many candidates wrote very long essay-length responses. For short answer questions only one piece of evidence is necessary.

Sociocultural level of analysis

Question 3

Several candidates struggled to write precise descriptions of the study, and many did not use precise psychological vocabulary when discussing the design and procedure. Strong responses clearly linked the findings to the theory. The majority of responses used Bandura's 1961 "Bashing Bobo" study, but there were also several strong responses that made use of less complicated studies.

Several candidates did an in-depth evaluation of the study. This was not required by the command term ("describe") and often distracted from the demands of the question.

Section B

Question 4

The question asked candidates to discuss research methods used at the biological level of analysis. When the plural is used, two or more research methods are expected. Many candidates did up to six research methods, making for a highly superficial response that attracted few marks. It is recommended that only two or three research methods be discussed.

Several candidates demonstrated a very limited understanding of research methods. Often candidates described several studies with minimal reference to the method itself. In addition, several "non-methods" were chosen – for example, autopsies, lesioning, and animal research. Animal research is not in and of itself a method. Animals are the sample. Just as "student research" would not be acceptable, candidates must focus on the method and not the sample. Often candidates wrote long essays on the ethics of animal research. This attracted low marks.

Many candidates had difficulty with case studies. Many candidates argue that case studies can never be generalized – and this is incorrect. Very few candidates addressed the holistic nature of the case study method or the fact that it uses triangulation. Instead of discussing how case studies are used, examples were simply given. The most common example was Phineas Gage, which is over 160 years old. It is recommended that candidates focus on research done in the last fifty years. Though it is not required, it is does give candidates a better understanding of the use of the method in modern psychology. Basing arguments on old and often outdated studies often leads candidates to draw incorrect conclusions.

Question 5

This was a popular question on the exam. Several candidates did not take the time to clearly outline schema theory. Many candidates also wrote long and often irrelevant introductions that distracted from the argument.

As noted above, many candidates evaluated individual studies, not focusing on the theory itself. Many also discussed only very dated research, and concluded that schema theory is not applicable. This ignores the large quantity of modern research on the topic. Though critical thinking was evident, it needs to be related to the question and should not simply be a formulaic approach to the question.

Question 6

There was an impressive number of very strong essays for this question. As with question five, there were several candidates who evaluated individual research rather than the theory itself. There were also several essays where the candidates simply described a number of studies. In many of the essays limitations were simply identified in the conclusion and not discussed in any detail.

Recommendations and guidance for the teaching of future candidates

Writing skills should be reviewed. Too many candidates write very long paragraphs that drift from idea to idea, losing focus. Long lead-in paragraphs of a general nature that do not address the question should be avoided. Introductions were often poorly constructed, leading to lower marks with regard to the focus of the response. Candidates should **not** automatically outline principles of the level of analysis for each question.

Candidates should define terms and be sure that theories and studies are clearly explained, not assuming too much knowledge on the behalf of the reader. It is also important that psychological terms be clarified – for example, "triangulation is a strength" is not enough. The candidate should clearly demonstrate understanding of the term.

Evaluation of studies alone is not enough for "critical thinking." Presenting conflicting theories/studies, discussions of ethical or cultural considerations, or questioning the assumption upon which a theory is based make for more sophisticated critical thinking. It is important that the evaluation not be formulaic – and that there be a clear explanation of any evaluations made. Simply writing "the study is not ecologically valid" is not a good example of critical thinking.

Candidates should avoid anecdotal explanations of behaviour and focus on empirically based support.

Teachers should devote time to deconstructing the command terms with candidates. Special attention should be paid to the higher-level command terms to make sure that their meaning is clear.

Finally, though classic studies from the nineteenth century may help to create a historical framework to help candidates understand the development of psychological ideas, it should be recognized that studies that are over 100 years old rarely play a significant role in our modern understanding of behaviour. It is recommended, though not required, that candidates focus on research carried out in the last fifty years.

Higher and standard level paper two

Component grade boundaries – higher level

Grade:	1	2	3	4	5	6	7
Mark range:	0 - 4	5 - 9	10 - 13	14 - 19	20 - 25	26 - 31	32 - 44

Component grade boundaries – standard level

Grade:	1	2	3	4	5	6	7
Mark range:	0 - 2	3 - 4	5 - 7	8 - 10	11 - 13	14 - 16	17 - 22

The areas of the programme and examination that appeared difficult for the candidates

The overall quality of responses tended to be rather good but varied greatly from answers that provided focused and clear knowledge and understanding relevant to the question to those providing general answers for certain topics from the learning outcomes without referring to the specific command term or context of the question. Many answers tended to reflect good descriptive knowledge and understanding of the required topic but were weaker on criterion B, critical thinking.

Answers scoring in the lower ranges reflected a tendency toward anecdotal comments or generalized responses lacking in specifics. Some candidates found it difficult to support ideas with relevant psychological research (theories and/or studies).

There were no particular areas that appeared to present specific difficulties for candidates. Candidates had difficulty with the question using “compare and contrast” as the command term. Many responses to this command term contained two separate descriptions of theories, models, strategies, etc., with a weak paragraph at the end in which a few evaluative comments were provided. Candidates were better at identifying differences than they were at identifying similarities.

The areas of the programme and examination in which candidates appeared well prepared

In general, even weaker responses included an introduction and definition of key terms. In the most popular option, abnormal psychology, candidates seemed to be well prepared in terms of content knowledge. Some of the candidates who answered in Spanish were well prepared in the area of developmental psychology, particularly the question on attachment. The responses to the psychology of human relationships questions were well structured and included empirical research in the formulation of good arguments.

Evidence of analysis and evaluation was clearly present in the top essays. Some responses were exceptionally well written – these responses reflected that knowledge of relevant material was clearly applied to the questions and the answer was well organized.

In general, most responses provided some reference to psychological research, although precise and focused knowledge of research was not always present.

The strengths and weaknesses of the candidates in the treatment of individual questions

Question 1

This question was the least popular choice within the option. As noted above, “compare and contrast” is a very demanding command term.

Good responses identified a specific dysfunctional behaviour and compared individual and group approaches to treatment, usually within the framework of the cognitive and sociocultural levels of analysis.

Lower quality essays spent too much time describing the chosen dysfunctional behaviour and offered limited knowledge of individual and group approaches to treatment and provided vague comments partially relevant to the “compare and contrast” demand. A significant number of candidates *evaluated* the approaches, identifying strengths and limitations rather than identifying similarities and differences.

Some candidates knew much more about individual approaches to treatment than about group approaches. Some candidates discussed the use of drugs as an individual approach to treatment.

Question 2

This question was a popular choice within the option.

Most responses to this question offered descriptive (and sometimes inaccurate) accounts of studies. Popular choices were: Rosenhan's participant observation (often wrongly described as an "experiment") and Nicholls' research at Great Ormond Street Hospital. Unfortunately, many candidates ignored the command term in the second part, to evaluate the research. Some answers focused on general issues in diagnosis such as the definition of abnormality, the subjectivity of diagnosis, and the problem of labelling, without reference to research. The distinction between reliability and validity is not well understood.

Question 3

This question was another popular choice and was answered very well by many candidates. The best answers tended to identify either depression or bulimia and after a brief description focused on etiology clearly linked to therapeutic approach. Quite often, the description of the disorder was followed by a description of etiology, followed by a description of therapeutic approaches. A significant number of candidates clearly identified possible biological, cognitive, and sociocultural etiologies.

In some responses the link between etiology and therapeutic approach was implicit rather than explicit. Lower quality responses lacked a clear focus on an affective or eating disorder but instead provided a general response. A few candidates discussed schizophrenia and a few others discussed an anxiety disorder. Other answers receiving low marks focused on symptoms, prevalence, or difficulty of diagnosis.

Question 4

The best responses included a brief definition of attachment followed by a description and evaluation of two studies. These were usually Ainsworth's research study on stranger anxiety, Bowlby's internal working model, Hazan and Shaver's study of adult relationships or Harlow and Harlow's studies with rhesus monkeys. A few candidates linked Ainsworth's research with Hazan and Shaver's studies, using these studies as supporting research evidence for Ainsworth's findings. This approach produced a quality response since the latter study was used as evaluative commentary rather than as simply an additional study. Appropriate evaluation of each study followed the description with contextual discussion of resilience in some responses.

Lower quality responses indicated that candidates were well-versed on a variety of topics related to attachment but they had difficulties focusing on only two studies or theories. They described everything they knew about attachment.

Question 5

There were relatively few strong answers to this question. Better answers tended to use research such as Mead's anthropological studies and contrast it with biological research to show "to what extent" (a difficult command term) sociocultural factors influence the development of gender roles.

Less prepared candidates offered general and somewhat anecdotal answers with stereotypical gender roles justified using a "light" version of evolutionary theory. In these answers, evolutionary theory was not used as a counterpoint to sociocultural factors but as a way to seemingly fill in for the lack of information on sociocultural factors.

Question 6

The vast majority of answers (in English) used Piaget's theory of cognitive development as the focus. Many answers were quite good, giving a reasonable description of Piaget's stages and providing a good range of the limitations of this theory.

For answers written in Spanish, Piaget's theory was relatively well explained but evaluation was often weak and often based on inaccurate information. The better responses used Vygotsky and Brunner's theoretical positions to criticize Piaget's theory.

Question 7

Some candidates were very well prepared to answer this question and produced very strong responses. Two examples of research were clearly identified, described and evaluated, then other studies were used effectively to evaluate the two selected studies. Most candidates were able to at least describe two pieces of research, even if evaluation was limited. The weakest responses were mostly descriptive or deficient in describing and evaluating a second piece of research.

Question 8

Most candidates focused on treatments for addictive behaviours, most frequently addiction to tobacco. Answers tended to be descriptive and general. Often candidates provided a review of medical treatment and CBT. Usually more detailed knowledge and a thorough discussion was present for medical treatment while knowledge of CBT tended to be less detailed.

Question 9

Many answers referred to sedentary life style, eating habits and family traditions but provided little reference to theories or studies. Very broad generalizations were offered: for example, advertisements and fast food are responsible for obesity. Many candidates provided a superficial account of several sociocultural factors. Stronger responses usually focused on only two factors and provided an in-depth discussion of how these factors were related to overeating and the development of obesity.

Question 10

Question 10 was a very popular choice. Many candidates were well prepared to write a strong response to this question.

Well prepared candidates discussed Latané and Darley's research on the role of the number of people available to help in a situation (diffusion of responsibility) as well as informational social influence (pluralistic ignorance). Piliavin's cost versus benefit model as well as cognitive dissonance and arousal were also discussed by some candidates. Cultural factors were introduced in the form of evaluative comments.

Some responses tended to provide a detailed description and evaluation of studies on bystanderism instead of clearly addressing the question about factors influencing bystanderism. Some candidates spent too much time describing the Kitty Genovese case in great (and sometimes inaccurate) detail.

In general, responses that discussed a relatively smaller number of factors influencing bystanderism in greater depth tended to gain more marks.

Question 11

This question produced many responses in the mid-range of marks. Most candidates who responded to this question could accurately identify and describe a biological explanation and a psychological or social explanation, but the contrast tended to be somewhat superficial.

Question 12

This question was not a very popular choice within the option.

Most responses provided several theories related to sociocultural explanations of the origins of violence such as social identity theory, social learning theory, deindividuation and conformity explanations. Some candidates focused on prejudice rather than violence. In the majority of responses there was more focus on detailed description than on providing evidence of critical thinking.

Higher quality responses clearly addressed the command term "evaluate". This focused approach was achieved by discussing strengths and limitations of research by using empirical evidence that supports or contradicts the research.

Questions 13, 14 and 15

In regard to questions 13, 14 and 15 on sport psychology, there were so few responses that no useful generalizations can be made.

Recommendations and guidance for the teaching of future candidates

- Candidates need practice in writing high quality responses. Many candidates seem to have a lot of knowledge of key terms and explanations relevant for psychology but fail to apply their knowledge to the specified question. After writing their responses candidates should check whether they have addressed all aspects of the question, and most importantly, whether they have explicitly addressed the requirements of the command term.
- A good idea might be to ask candidates to go through the exercise of marking their own or someone else's response. This exercise can help candidates understand how the examiner is marking an essay.
- Candidates should be reminded that when two studies are required in the question, discussing additional studies does not add marks to their response since only the first two will be considered. An exception to this rule is when candidates are using additional studies to clearly and explicitly evaluate the two studies that are the focus of the question. In this case, candidates are showing clear evidence of critical thinking.
- Some candidates presented detailed explanations in their responses but failed to support their knowledge with relevant research (theories and/or studies). This use of research is a general requirement for paper two responses and is indicated in the general instructions for this examination. Candidates should routinely be reminded to include the use of research in their responses.
- There should be a stronger emphasis on the organizing principle of levels of analysis within the teaching of the options.

Higher level paper three

Component grade boundaries

Grade:	1	2	3	4	5	6	7
Mark range:	0 - 2	3 - 5	6 - 7	8 - 11	12 - 14	15 - 18	19 - 30

General comments

Candidates seemed to have a good understanding of the stimulus material. There was some spread in the marks awarded and candidates scored all along the mark range with few in the low range, most in the middle and quite a few in the higher range. This is an indication that many candidates are fairly well prepared to answer paper three questions.

The areas of the programme and examination that appeared difficult for the candidates

The most difficult question for candidates appeared to be question three that involved knowledge and understanding of reflexivity in qualitative research. Some candidates had accurate knowledge and understanding of reflexivity as it is used in qualitative research. However, many did not seem to be familiar with the term “reflexivity” as it is applied in relation to qualitative research but rather talked about “reflections” in relation to methodology in general. The lack of understanding of reflexivity as a way of increasing credibility/trustworthiness in the qualitative research process resulted in a number of responses that scored zero or very low.

As for question one it appeared that some candidates did not have exact knowledge of what a non-participant observation is. Many assumed that the researchers were not present and just watched the video recordings.

Some candidates referred to participation of children in question two on ethical considerations in the study. These candidates often addressed whether this was ethical and therefore did not really answer the question.

There was a tendency in many papers to speculative analysis rather than analysis based on knowledge of qualitative research applied to the stimulus material.

The areas of the programme and examination in which candidates appeared well prepared

Many candidates showed a good understanding of ethical considerations in question two and were able to give at least two explanations of relevant ethical considerations that were applied to the study in the stimulus material.

Most candidates knew what an overt observation is and wrote that it is more ethical than a covert observation.

The strengths and weaknesses of the candidates in the treatment of individual questions

Question 1

Most candidates demonstrated some knowledge of overt non-participant observation but many were not able to effectively address the command term “evaluate”, i.e. refer to strengths and limitations of non-participant observation and relate this directly to the stimulus material. Stronger responses defined the two terms and referred to strengths and limitations of its use in relation to the stimulus material. Weaker responses referred to general strengths and limitations of the method and did not link this to the stimulus material.

Quite a few candidates referred to “the Hawthorne effect” or “reactivity” as a limitation. Some referred to the ethical aspects of non-participant observation and that this was the most ethical method compared to covert participant observation as informed consent could be obtained. Other strengths mentioned were that the researchers were able to observe and make notes more freely as they were not part of the research process.

A large number of candidates used the term “experiment” interchangeably with “study” or “observation”. This is not recommended.

Question 2

Ethical considerations were generally well addressed. Most candidates could refer to ethical considerations like informed consent and link this to participants knowing they were observed in the stimulus material. The issue of confidentiality and problems of keeping participants’ anonymity were also often referred to in the responses. Quite a few said it was unethical to conduct research on sensitive topics like EBD and parenting skills but this argument was mostly based on personal opinion and not on ethical considerations in the study. Quite a few candidates said it was unethical to conduct research like this with children although children were not part of the study. Some candidates misunderstood the question and rather tried to make an evaluation of the study in the stimulus material.

Question 3

The question seemed particularly difficult to a number of candidates who scored very low marks in this question. However, some candidates seemed very well prepared and were able to explain reflexivity in qualitative research and in relation to the study in the stimulus material in a way that demonstrated understanding. Those candidates could explain the role of reflexivity in general and give examples from the study, for example that reflexivity could be applied in the study to prevent researcher bias because of the researchers’ work at a centre for children with behavioural problems. This point in the stimulus material was not generally reported as a factor that could be a reason for applying reflexivity but some candidates did mention this in their responses.

Many candidates referred to personal and epistemological reflexivity. The latter was sometimes explained as considering if observation was a useful method in this particular context. The strong responses could relate this to knowledge issues but the weaker responses just pointed at strengths and limitations of the method without reference to why a particular method could be useful or not in the context of the study or if it would have been better to use a different method with regard to the knowledge the researchers hoped to gain from the study.

Recommendations and guidance for the teaching of future candidates

Paper three in the present syllabus requires that candidates integrate knowledge of qualitative research methods with a specific stimulus material. It is not enough to describe what is in the stimulus material as it is intended to serve as a starting point for analysis of how qualitative research methods could be applied to a study. Candidates should practise “what it is like to be a qualitative researcher” so that they are able to apply relevant methodological considerations to the stimulus material. This means that teaching should include exposure to qualitative studies. It is important that candidates come to understand the difference between quantitative and qualitative methods so that they avoid using too much terminology from quantitative methods. For example, many candidates seemed unaware that terms like “experimenter”, “experiment” and “Hawthorne effect” are not really appropriate to use in the context of a qualitative study.

The questions based on an introduction with an outline of a study or scenario (the stimulus material) should give candidates an opportunity to demonstrate how to apply relevant knowledge and understanding of qualitative research methods in the context of that study or scenario. This means that candidates should be prepared in such a way that they have both a general knowledge of qualitative research methods mentioned in the guide as well as competence in applying this knowledge in relation to a stimulus material. Using previous examination questions could do this or teachers could prepare material and questions as they teach qualitative research. It is also recommended to train candidates to make balanced evaluations and discussions instead of claims and speculations that are not supported.

It is advised to instruct candidates in what it means to address the command term in relation to paper three, for example what “explain” means. Too many candidates just speculated on what they thought would be relevant but failed to explain why it could be relevant in relation to the stimulus material. Many candidates also assumed that because certain points were not made explicitly in the stimulus material (for example, in relation to ethical considerations) these were non-existent.

Every fifth line in the stimulus material is numbered so that candidates may refer to the lines without having to use extensive quotations. This could be used more effectively in the responses.