November 2012 subject reports

PSYCHOLOGY

Overall grade boundaries

Higher Level

Grade:	1	2	3	4	5	6	7
Mark range:	0 - 9	10 - 20	21 - 31	32 - 44	45 - 57	58 - 70	71 - 100

Standard Level

Grade:	1	2	3	4	5	6	7
Mark range:	0 - 10	11 - 23	24 - 33	34 - 46	47 - 59	60 - 72	73 - 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

Overall the range and suitability of the work submitted was of a good standard this session. A few examples of unethical work based on ingestion were seen although the guide is very specific that candidates should not undertake such studies (page 48). The majority of candidates were well aware of ethical issues and all candidates with a full report included a copy of informed consent in the appendices.

Most reports were based on cognitive psychology and this seems to provide good results at this level of education. Favourite experiments were, as usual, levels of processing, Stroop, Loftus and Palmer (1974) and experiments related to schema theory and reconstructive memory. A few candidates performed experiments with several conditions although a simple experiment with only two conditions is recommended in the guide.

In general, the weaker reports shared the following characteristics:

- Weak and imprecise explanation of background research in the introduction. The hypotheses were not clearly justified and operationalized.
- Results were not always clearly related to the aim of the study and inferential tests were absent or not justified. In descriptive statistics, the results were not stated in words, the use of descriptive statistics was not explained, or there was no table.
- Discussions were superficial with no discussion of own results in the light of background research and/or no reference to statistics. Identification of limitations of own procedure was not linked to suggestions for modification.
- There was poor or no referencing

It should be noted that it is not required to make an exact replication of an experiment. A partial replication will do but the candidate's experiment should be closely linked to an actual experiment and not just inspired by some kind of research.

Candidate performance against each criterion

Most candidates seemed familiar with the requirements of each assessment criterion although there were differences in achievement levels. In the stronger reports, the introduction was very well written with a clear focus and well explained background research that clearly led to the candidate's own research hypotheses. It seemed difficult for weaker candidates to explain the background research in sufficient detail so that there was a clear relationship between a particular research study and the candidate's own research hypothesis. The introduction is important in that it presents the rationale for the candidate's own experiment and uses the background research to justify the candidate's own research hypotheses. Therefore 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 and to stimulate discussion of own results in the light of the background research in the discussion section. The experiment should be simple and therefore it is acceptable to carry out a partial replication of a research study, for example, by reducing the number of variables. The introduction and discussion sections are often the most difficult to write as they require a good understanding of how research studies are linked to formulation of new hypotheses. The level of depth of the analysis of the background research was at times shallow, partly because weaker candidates used insufficient background material such as revision guides. This influenced the discussion where comparisons of the candidate's own results and those of the background studies could not be done in sufficient depth. Referencing was not always of a standard format and there were still problems with resources found on the internet.

Criterion A: introduction

Generally, the background research needed more attention in the reports. Theories and studies were often not explained in sufficient detail to justify the hypotheses. The aim of the candidate's own study was not clearly formulated in all reports. The analysis of background research was in some cases very well done but at times superficial, for example based on a summary of a study found on a website or a

revision guide. This gave some problems in terms of justifying and formulating clearly operationalized hypotheses. It was obvious that some candidates found it very difficult to state a clear and justified 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 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.

Criterion B: design

Most candidates could state an appropriate experimental design (repeated measures and independent designs) but choice of design was not always properly justified, for example by reference to strengths and limitations of respective designs. A number of candidates had problems with operationalization of the IV and the DV. The ethical guidelines were mostly addressed in the design section but sometimes in procedures instead (and that is acceptable). What is essential is that ethical procedures are addressed appropriately somewhere in the report. Most candidates included an informed consent from participants. Similarly, there was parental consent for participants under the age of 16. In the few cases where the informed consent was not included it was due to a general weak report where all appendices were missing.

Criterion C: participants

Many reports did not include a sufficient number of relevant characteristics of the participants. Candidates could focus on characteristics such as age, sex, number of participants and nationality (or in studies such as Stroop that participants were not colour blind). The target population was not always identified. 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 often explained or justified (both ways are fine). Some candidates claimed to have used a random sample, which it was not. There seemed to be confusion about the meaning of the word "random". Some candidates did not specify how they allocated participants to the experimental conditions but they cannot be penalized for this since it is not mentioned in the assessment criteria.

Criterion D: procedure

There were at times problems with the description of the procedure in sufficient detail, and therefore it would be difficult to replicate the experiment. Reference to the materials in the appendices (either in procedure or in materials) was not always done and this would make replication difficult or impossible. It is recommended to include all materials and give details on how the materials were used in the procedure.

Criterion E: results – descriptive

Most candidates described the results in a narrative form in the results section. Not all included standard deviation as descriptive statistics even if their data allowed it. Some used the range. Only the strongest candidates explained the use of descriptive statistics. There was still a tendency to include a whole range of measures of central tendency in the results section but this is not recommended. It does not affect the marks but it is a clear indication that candidates have not considered which measure would be most appropriate for their data.

In some reports there was no graph but only a table of results. Not all reports included tables. A few candidates had individual scores in the results section but most candidates had correctly placed the raw data in the appendices. A few candidates graphed their standard deviation side by side with the mean but this does not really make sense and should be avoided.

Criterion F: results – inferential

Most candidates chose and justified the inferential statistical test correctly but a few did not. Most candidates used the non-parametric tests but an increasing number of candidates used the t-test, which is also appropriate since these tests are quite robust. This session some candidates used the chi-square test even though their data was not nominal. Most candidates justified their choice of statistical test but some did not. A number of candidates did not make a statement of statistical significance.

Criterion G: discussion

A number of high-scoring candidates demonstrated competence in discussing their own results in the light of the background research and these candidates identified relevant methodological limitations and suggested relevant modifications to their own study. Those candidates who wrote poor introductions were not able to discuss their own results in the light of previous research. Many reports had a very short and superficial discussion of their own results in the light of previous research but a long description of the limitations of the candidate's own design – which was not always relevant in terms of their own study but rather a more general list. A few candidates included strengths of their own design although this is no longer necessary. The suggestions for modifications were often related to general factors rather than the candidate's own study. For example, there were a lot of candidates who mentioned the opportunity sample as a problem and suggested a random sample and more participants in future research.

Criterion H: citation of sources

Candidates often did not include all the references mentioned in the introduction. There are still problems with internet references where candidates tend to think that the URL is enough, and sometimes the background study could not be found in the reference section. Often candidates did not use a standard citation method such as APA or referencing was not complete. Secondary referencing was generally poor.

Criterion I: report format

Only a few reports did not include an abstract or appendices. A few reports were not within the word limit. Generally the appendices were well organized and labelled. Most of the time the abstract was clear and included a summary of aim, procedure and findings of own study.

Recommendations and guidance for the teaching of future candidates

- Of utmost importance is to help candidates find appropriate background research, that is, theoretical framework and appropriate experiments. Finding relatively simple experiments to replicate is recommended. This would help candidates to undertake 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.

- 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 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.

- The design section should include a clear description of the two experimental conditions. Teachers are encouraged to recommend simple experiments with only two conditions to their candidates as this will make it easier to compare the outcome of the manipulation of the IV on the DV in the two conditions.

- Sampling should be done according to IB rules, that is, identification of target population including relevant characteristics and description of sampling method as well as explanation (or justification) of the use of the chosen method. Most candidates use a convenience sample but they should still explain the sampling method or justify its use. The number of participants in the experiment does not need to exceed 20 (independent design) or 10 (repeated measures design), and it is recommended to observe this.

- In the descriptive statistics section, the graphs and tables should have proper titles. It should be emphasized that graphing the results is mandatory and a table must always be included to access the highest mark in criterion E. It is important to instruct candidates that this section includes summarized data not raw data or individual scores. It is recommended that candidates do not include several measures of central tendency but only the one which is specifically relevant for their data. Likewise, an appropriate measure of dispersion should be calculated and the results included in this section. Calculation of descriptive statistics must be documented in the appendices.

- In the inferential statistics section, candidates should take care choosing an appropriate statistical test and justify why this test was chosen. This could relate to the level of measurement of the data and the experimental design. Calculation of statistics should be documented in the appendices. If the calculation is performed online, a screen shot of the calculation could be included in the appendices as documentation.

- 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 writing the discussion section (perhaps by reading a couple of research articles to become familiar with the idea and style). Identification of limitations of own procedure and suggestions for modification should be tied together. 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 scientific standards, which include 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. Many candidates only used internet sources of a non-specialist nature as background literature. Furthermore, candidates could gain more marks in criterion H if they learn to use a standard citation method.

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

The quality of the reports submitted tended to suggest that the level of guidance provided by the teachers was generally very good, with a large number of centres submitting reports on appropriate topics. However, there were a few issues of reports where there was no clear manipulation of the independent variable. The research topics chosen were usually from cognitive psychology and this approach is very likely to provide good reports at this level of education. There were, as usual, a number of replications of Stroop effect, Mozart effect and experiments replicating a simplified version of Loftus's study on eye witness testimony. Occasionally research replicating Bartlett's study of reconstructive memory tended to pose a difficulty because the independent variable was not clearly manipulated. In these reports candidates defined the independent variable as time delay and chose the repeated measures design. However, for this study it was not possible to randomly allocate participants to different conditions or to use counterbalancing because all participants were basically exposed to the same experience.

Ethical guidelines were usually well followed and evidence of this was clearly provided within the report. The majority of candidates attached a blank copy of the consent form and debriefing letter in the appendices. Some candidates did not state ethical considerations within the design section. However, if they referred to ethical considerations in the participants section or in the procedure section moderators awarded full marks for this criterion if all other requirements had been met.

Details of the procedure of the experiments were generally well done. Candidates had more problems with choice, justification and explanation of design, sampling method and presentation and justification of data analysis. Discussions were often written in a superficial manner, candidates often didn't discuss own results in the light of the replicated study, and/or there was no reference to statistics. Also in lower quality reports referencing was poor and incomplete.

Candidate performance against each criterion

Many candidates still have problems with providing a clear link between the introduction and discussion sections. Although the introductions were in some cases well written there was a lack of clear focus on the replicated research study and candidate's own aim. A good introduction is important and the original study being replicated should be explained and

analysed in sufficient depth (for example, aim, procedure, choice of design and sampling method, findings) to allow for a thorough, well developed comparison of own results in the light of the original study in the discussion section.

Criterion A: introduction

Often the aim of the study was not clearly stated by addressing the relationship between the IV and DV. Instead some candidates stated that their aim was to replicate the original study. Occasionally a hypothesis was stated rather than the aim. Most candidates did well identifying a relevant study and describing the findings of the study being replicated, but not all reports clearly explained all relevant aspects of the original study (aim, type of design used, target population and sampling method, procedure, results obtained and conclusions reached).

Occasionally candidates attempted to write a review of literature but failed to clearly identify and explain in detail one study that was actually replicated. Teachers should not encourage candidates to provide several studies in the introduction since this is not required according to criterion A. Instead, focus should be made on only one relevant study.

Criterion B: design

In some cases candidates had problems with applying their knowledge of experimental methodology to a specific aim that they had chosen. In addition to this some reports reflected candidates' lack of knowledge of psychological research terminology. In this section candidates should identify the design of their research study and not the research method chosen. Since candidates are required to select and perform an experiment of their choice they need not justify why the study is an experiment. In some reports candidates tended to select an appropriate design, but still had difficulty justifying its use. When justification was attempted this referred to vague and general justifications that were not linked to relevant methodological issues. Justification of design asks the candidate to explain why they have chosen a specific type of design. Candidates should select the most appropriate design for a specific aim and the conditions in which the study will be performed.

On the whole most candidates correctly stated the independent variable and dependent variable for their experiments, but these definitions often lacked clarity and operational definition.

Criterion C: participants

Candidates did a good job of identifying and describing relevant characteristics of the participants and identifying the sampling technique, although a few reported choosing a random sample when they actually used opportunity sampling to recruit participants. Although the use of opportunity samples is appropriate, the justification for its use needs to be more thorough and meaningful.

Criterion D: procedure

In general this section was very well written. Most candidates provided a clear description of the procedures used; however, some reports only included a vague and incomplete listing of steps, which reduced the ability to replicate the study. Occasionally, standardized instructions, questionnaires or debriefing letters were mentioned in the procedure but were not made available in the appendices. In lower quality reports candidates tended to omit important aspects of the study needed for replication (for example, the list of words used for a memory task or the type of task used before giving a memory test).

Criterion E: results

The quality of the results sections varied considerably. In high quality reports the results section was both clear and informative, but there were occasions when the results were not related to the aim or candidates provided minimal chunks of information that lacked clarity. Presentation of tables and graphs needs to be clearer so that data can be readily understood by the examiner. All graphs should have a title, a legend and a label for each axis.

In some cases candidates presented and explained findings that were not part of the original aim. In these cases the measures of central tendency were not matched to the choice of the IV and DV stated in the method section. Usually the descriptive statistics chosen were mean, median, mode, standard deviation and range. Teachers should encourage candidates to choose only one measure of central tendency (the one which is appropriate for the level of measurement used when obtaining the results). In addition, an appropriate measure of dispersion for the measure of central tendency should be included. Justification of descriptive statistics was often inappropriate or missing. Often calculations regarding the standard deviation were not seen in the appendices.

Criterion F: discussion

The quality of the discussion sections varied greatly. In general the majority of candidates provided some discussion of their results in relation to the original research in the introduction section but this was often done in a vague and general manner. Usually an identification of the main weaknesses of the study was provided and several suggestions for improvement were stated. However, some candidates overlooked obvious confounding variables that might have affected their study. Some candidates spent a lot of time and words writing about the strengths of the study although it is not required according to criterion F. Some candidates forgot to include a brief conclusion.

Criterion G: presentation

Presentation in many reports was excellent and fully in accordance with the requirements of criterion G. Unfortunately, there were still reports in which the maximum number of words permitted was exceeded or in which the reference section was missing. In some cases candidates used a standard citation method but there was a lack of in-text citations.

In weaker reports a common mistake was failure to label appendices and reference them in the body of the report. Also, tables with raw data were sometimes incorrectly included in the results section.

Recommendations and guidance for the teaching of future candidates

- Teachers should encourage candidates to read the original article of the study that they plan to replicate.
- When doing research inspired by Bartlett it is perhaps worthwhile to consider how the original findings could be modified to make a true experiment instead of replicating Bartlett since that study was not a true experiment either.
- Teachers should try to provide exercises that might help candidates understand how they should clearly and properly operationalize the independent and dependent variables.
- The sampling method needs to be identified and justified. The size of the sample of participants should be around twenty; using too many participants in the sample should not be encouraged. If candidates include participants under the age of sixteen, they should obtain parental consent (or if the research is conducted in a boarding school the consent of the headteacher or principal is necessary).
- Candidates should also be mindful of the information given related to criterion E (results). Only the most pertinent central tendencies should be reported. Candidates should provide a justification or explanation of the use of descriptive statistics by making reference to the aim of the study.
- Before writing the discussion section candidates should carefully examine the procedure conducted in their study, and analyse ethical considerations or sampling biases that might have had an impact on the results obtained.
- Candidates should be encouraged to place all of their statistical workings in the appendices.
- Finally, teachers can encourage candidates to mark their own draft report using the criteria to improve their mark in the internal assessment.

Higher and standard level paper one

Component grade boundaries

Higher level							
Grade:	1	2	3	4	5	6	7
Mark range:	0 - 5	6 - 11	12 - 14	15 - 20	21 - 27	28 - 33	34 - 46

Standard level **Grade:**	1	2	3	4	5	6	7
Mark range:	0 - 5	6 - 11	12 - 14	15 - 20	21 - 27	28 - 33	34 - 46

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

Some candidates gave very long answers for section A questions leaving less time to adequately answer section B questions that should be more developed. Some candidates used the same study to answer two or even more questions showing a lack of breadth of knowledge. In addition, errors made in the description of a study which is used repeatedly end up having a significant impact on the candidate's overall performance. Moreover, the critical thinking elements in the extended responses were often limited and rarely well integrated to the answer.

Many candidates struggled with the demands of the command terms. It appears that many have been taught a systematic approach to answering the questions – which includes naming all principles and giving a history of the level of analysis before getting to the question. This is an ineffective approach to the questions and should not be encouraged. It is important for candidates to remember that when the question asks for *one* study or principle, if the candidate writes several examples, only the first one is assessed.

There were many candidates who used a significant amount of research from the 19th century to answer the questions on this examination. Historical research should be avoided. It is not in the spirit of the current curriculum that attempts to introduce candidates to modern psychology.

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

Candidates were generally familiar with syllabus requirements, particularly the content. There were examples of outstanding levels of organization and structure.

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

Question 1

Most candidates identified a relevant principle and used an appropriate study as an illustration of the selected principle. However, few of them clearly explained how this principle was demonstrated in the particular study. Too often, the focus of the answer was on the description of the study without explicitly addressing the question. Some candidates confused principles with theories.

Question 2

Some excellent responses looked at the role of cortisol, acetylcholine or beta-amyloid proteins on memory. However, many candidates identified "brain damage", "localization of function" or "Alzheimer's" as a biological factor, often failing to actually address the actual biological factor and how it affects the cognitive process. Often the focus of the response was on a particular study, rather than on the influence of a biological factor on a cognitive process; the explanation was rarely fully developed. Many candidates attempted to write about how biology and cognition interact in emotion. This did not meet the demands of this question. Sadly, many candidates argued that Schachter & Singer was a study of the effects of adrenaline. This is a fundamental misunderstanding of the aim of the study.

Question 3

Candidates seemed to have a superficial knowledge and understanding of these concepts. Many candidates believed that etic approaches study "universal" behaviours and emic approaches study "culturally specific behaviours". Many candidates did not understand that emic and etic are approaches to how cultural research is done; it is not an assessment of the outcome of a study.

Often candidates used examples that were irrelevant such as Festinger's study or Bartlett's War of the Ghosts study. When they used appropriate studies, they rarely clearly linked them to the concepts. Many candidates used examples from anthropology that outlined rituals of tribal communities, rather than focusing on psychological research.

Question 4

Most candidates addressed the nature/nurture debate, using intelligence, schizophrenia, depression and criminal behaviour as examples of behaviour. Some answers were well developed with studies supporting and inferring the genetic inheritance of those behaviours. However, many others showed knowledge of the issue but little critical thinking: they only evaluated the studies without integrating explicit arguments for and against genetic inheritance. Several candidates examined the ethical concerns of research on genetics. This was not relevant to the demands of the question. Many candidates did not understand the methodology used to study

genetic arguments of behaviour, citing "experiments" instead of "correlational research". Finally, there were several candidates who made use of research from the 1940s – for example Tryon. This was eugenics research that should be avoided. Candidates should have a more current understanding of the role of genetics and not be reliant on research from before the Second World War.

Question 5

Many candidates chose this question but too often candidates evaluated the selected methods rather than discussing how and why the methods are used at the cognitive level of analysis. In addition, candidates often examined biological rather than cognitive research. Candidates that used research that exemplified the method tended to achieve higher marks. Finally, there were several poor choices: Bartlett's work is difficult to classify as a true experiment and Ebbinghaus's "auto-experiment" does not meet the modern definition of the term.

Question 6

Most candidates described ethical issues related to classical studies such as Milgram's, Festinger's and Zimbardo's studies. Rarely did they discuss, for example, why those ethical guidelines were broken or how experimenters could counterbalance for breaking those rules. Nor did they look at the ethical considerations relevant to the sociocultural level of analysis – for example, a focus on naturalistic observations or studying groups rather than individuals. Candidates focused too much on ethical violations rather than the demands of the question.

Recommendations and guidance for the teaching of future candidates

- Teachers need to focus on developing essay-writing skills. It is important that candidates realize that there is no single strategy that will allow them to answer all of the questions. Teachers should give feedback to candidates on their work that includes helping them to understand what was unnecessary or irrelevant in their responses. The command terms should be clearly reviewed by teachers in order for candidates to correctly focus on the requirements of the questions.

- Candidates should be advised to use different studies in support of their answers; using the same one shows lack of breadth of knowledge of the content areas.

- Candidates should know how studies link to the principles. Candidates are often confusing biological and cognitive research. Though the course looks at how these levels of analysis interact, it is also important for candidates to realize that these are also distinct areas of research.

- It is important that the research that is taught be carefully chosen. It is the goal of the programme that candidates have a good understanding of modern psychology, not of the history of psychology. Therefore, it is strongly recommended that research from the 1800s not be used to address the learning objectives.

- For section B responses, candidates should be advised that critical thinking is necessary for the highest marks and it has to be well integrated to the requirements of the question.

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 - 15	16 - 21	22 - 26	27 - 32	33 - 44

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 which appeared difficult for the candidates

No single option appeared to be more difficult than the others. As in previous examinations, candidates demonstrated better understanding and knowledge of topics than the ability to apply critical thinking skills to the material covered in the options. However, many candidates were able to provide basic evaluation of studies by focusing on issues such as lack of generalizability, insufficient sample size, lack of ecological validity, and various ethical concerns thereby earning marks in the middle range for criterion B.

Questions two and four proved difficult for some candidates who were unclear on the meaning of the command term “to what extent”. Understanding of this command term was relevant for providing focused evidence of critical thinking. Although both of these questions were rather popular many responses didn't receive the highest marks since they only focused on how certain factors affect development (disorder) but not “to what extent” they influence development (disorder).

The great majority of teachers considered the test to have been set at the appropriate level of difficulty and to have been of a similar standard or slightly more difficult than last year’s examination.

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

The majority of candidates were able to demonstrate a basic understanding of major concepts within the options and to use appropriate research in support of their responses. The responses of a significant number of candidates revealed that they had developed an

understanding that a comprehensive explanation of human psychology requires attention be paid to all three levels of analysis: biological, cognitive, and sociocultural.

As in previous examinations, it was clear that some candidates were well prepared in how to write a response. A strong response usually defined the command term in the first paragraph. For example, in a question in which the command term "evaluate" was used, if the first paragraph had a sentence along the lines of, "In this answer I will analyse the strengths and limitations of…", the response tended to be a good one.

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

Abnormal psychology

Question 1

Overall, largely descriptive responses were offered that provided descriptions of findings from empirical studies but failed to discuss how cultural variations affect the prevalence of disorders. Many candidates could not provide a clear and relevant conclusion.

Higher quality responses provided an in-depth discussion of two or three cultural variations. The best responses offered good knowledge about cultural variations in the prevalence of either depression or anorexia and supported their knowledge with relevant empirical evidence. The responses that best reflected that the candidates clearly understood the question provided only the material that was relevant to the question.

Question 2

Most candidates had no problems with answering the first part of the question. Lower quality responses provided a general, vague description of some symptoms while higher quality responses gave clear and relevant outlines (at times too long) in which they provided examples of behavioural, cognitive, emotional and physiological symptoms.

Responses to the second part of the question reflected rather good knowledge of cognitive factors that influence the etiology of a disorder – the most popular choices of disorders were depression and anorexia. Better responses did directly address the command term by consideration of how biological and sociocultural factors could have a role in the etiology of the disorder that was outlined.

Weaker responses gave a basic description of the negative cognitive triad with no or minimal attempt to provide confirming psychological studies. In addition, there was minimal evidence of critical thinking in these responses. There were some attempts to provide evidence of critical thinking but these weren't related to the requirements of the question but rather simplistic evaluations of studies. These responses usually ended with a conclusive sentence indicating that all factors need to be included in

order to explain the etiology of a specific disorder. At times, when responses were of marginal relevance to the question they included a general overview of all factors that might cause a specific disorder without focusing on cognitive factors.

Question 3

Most candidates chose to write about the use of SSRIs to treat depression. Usually good to excellent responses were provided. Most candidates provided good knowledge and understanding of biomedical approaches to treatment and used appropriate and relevant psychological research in support of the response – although there tended to be some gaps and inaccuracies. Candidates had more problems when evaluating biomedical approaches to treatment – strengths and limitations were outlined but not thoroughly addressed. Also, some candidates evaluated the studies rather than the treatment.

Higher quality responses were usually well structured and organized but at times failed to provide a good conclusion. The best responses evaluated more than one biomedical treatment.

Developmental psychology

Question 4

Most responses provided a general and vague account of some biological factors, for example, brain development, or made reference to biological factors that predetermine the stages of development in Piaget's theory of cognitive development. Some candidates simply described Piaget's theory with no reference to biological factors. A weak attempt to provide an opposing view was attempted in some cases – for example, Vygotsky's sociocultural approach to cognitive development. Responses usually provided basic knowledge with many inaccuracies and gaps.

Question 5

Most responses tended to be overly descriptive and failed to provide a link between physical change and development of identity. Both of these were usually addressed in an overly general and superficial manner. In a few weak responses there was an overemphasis on physical change. Erikson was used very poorly in a number of responses.

Question 6

Candidates usually chose social learning theory, gender schema theory and biosocial theory of gender role development. There were many good responses to this question – candidates provided detailed descriptions of theories or studies and gave a detailed balanced review of these theories usually by discussing methodological and cultural considerations. Mead's studies and the David Reimer case were often cited.

Some candidates failed to address the question as set and wrote a response on different factors influencing the formation and development of gender roles – these responses tended to provide rather good knowledge but failed to provide evidence of critical thinking relevant to the question and did not get many marks for criterion C.

Health psychology

Question 7

This was a very popular question within the option – usually it was addressed with solid knowledge and understanding of both physiological and psychological aspects of stress. Candidates often chose to write about the fight or flight theory and GAS as physiological aspects of stress. Cognitive appraisal was usually described as a psychological aspect of stress. Usually more emphasis was given to physiological aspects of stress. Most candidates provided evidence of critical thinking but it was rather limited.

Question 8

The best responses chose a specific formal health campaign conducted in a certain country and provided a clear and detailed description of the health promotion strategy in which the aim and methods of health promotion were included. The effectiveness of the health promotion strategy was assessed by providing relevant theoretical background or evidence of empirical studies and discussing methdological issues of the studies or cultural issues involved.

At times candidates provided general responses without referring to a specific health promotion strategy. These responses did not attract many marks.

Question 9

Most responses were of lower quality and tended to focus more on factors causing obesity or ways to prevent obesity than on treatment of obesity. Overly descriptive and general responses were provided. Ocassionally, candidates provided only strengths or only limitations of the treatment for obesity.

Psychology of human relationships

Question 10

Responses more often provided description and evaluation of research studies than theories. Many candidates chose the Whiting and Whiting study (1975) – in most cases good descriptions of this study were provided but candidates gave a limited evaluation. Many candidates also used Levine's studies.

In some cases candidates did not address the question as set but rather chose to write an essay discussing cross-cultural differences in prosocial behaviour.

Question 11

Many responses reflected good knowledge and understanding of the topic. The most popular choice was Olweus's strategy for reducing violence.

Question 12

High quality responses tended to focus on cognitive factors influencing attraction and/or helping behaviour. Some candidates used attribution theory to good effect. These responses provided thorough discussions supported with relevant empirical evidence. Other responses provided superficial, overly descriptive accounts of two cognitive factors.

Sport psychology

Questions 13, 14 and 15

There were very few responses to questions in this option, which provided superficial and general knowledge about the topic. Responses lacked the specific information which would be expected if candidates truly studied the option.

Recommendations and guidance for the teaching of future candidates

- As in every previous examination, a clear understanding of the command terms is a prerequisite for a strong answer. The different demands of the command terms should be explicitly taught. In particular, the command terms “discuss” and “evaluate” should receive special attention, and candidates should also understand that a strong response to a question using the command term, “to what extent” requires paying attention to factors not specified in the question. For example, a question such as, “To what extent do biological factors influence human relationships?” requires some discussion of cognitive and sociocultural factors in order for the response to be awarded marks in the upper range. Time spent on instructing candidates how to structure and write an effective response to the type of questions in the examination is time well spent.

- Writing skills should be systematically developed. Often responses are written as long paragraphs providing general knowledge that is not specifically linked to the requirements of the question. In order to structure their answer candidates could be trained to provide definitions of key terms stated in the question and to support their arguments with theories and/or studies that are highly relevant for their statements.

- If the first part of a question uses a command term from assessment objective 1, this command term indicates that this part of the question has a low cognitive demand and therefore candidates should not devote a lot of time answering this part of the question.

- Candidates should be given past paper questions to do and once they are done, the teacher should explain the assessment criteria and markscheme to the candidates. After this review has occurred, candidates should be asked to assess their own work.

In this way, candidates can raise their awareness of best practice for answering a question and can take responsibility for their learning.

- Most importantly, teachers should ensure that candidates form a connection between the theory/concept/term and empirical studies and in doing so ensure that they are evaluating the concept not just the empirical studies, according to the requirements of the question.

- Candidates should be reminded that when *one* study is requested, adding several studies lowers the quality of the response since only the first one will be considered. If a question asks for a discussion or evaluation of one study or theory, additional studies or theories may be brought in as part of the discussion and earn marks only if this research is clearly and explicitly used to evaluate the first research discussed.

- Teachers should not encourage candidates to provide huge numbers of studies since they are often misremembered, and not made relevant to the question. Instead, focus should be made on one classic study and one or two updated examples, and then on applying this research to answering the question effectively.

- In centres that are new to the IB programme it is imperative that teachers are aware of the demands of the psychology programme. Generalized, anecdotal discussions are not acceptable responses to examination questions. Specific psychological studies, terminology and concepts must be taught.

Higher level paper three

Component grade boundaries

Grade:	1	2	3	4	5	6	7
Mark range:	0 - 2	3 - 4	5 - 8	9 - 11	12 - 15	16 - 18	19 - 30

General comments

Many candidates in this examination session seemed to have a good understanding of the stimulus material. However, the fact that the stimulus material covered a study on social networking profiles made some candidates go off track and talk about their own experiences with social networking and thus they forgot that they had to answer a question related to a specific research study on the topic. For example, candidates mentioned that there might be very personal material on their profile that they did not want an adult researcher to see, thus indicating that they had not read the stimulus material properly. As in previous years, the weaker responses were quite speculative and arguments were based on beliefs and anecdotes rather than knowledge of qualitative research methodology. There was also a tendency among the weaker candidates to argue based on knowledge of quantitative research although this was less obvious than in previous sessions. This could indicate that more training in application of qualitative research methodology to the stimulus material is needed.

As usual, some candidates had problems integrating the stimulus material with knowledge and understanding of qualitative research methodology but overall there were few examples of candidates who did not do this at all. As in previous sessions, quite a few candidates used the term “experiment” as a generic term for “research study” but it did not always affect their reasoning with regard to qualitative research methods.

Generally, it seemed to be challenging for some candidates to use the information in the stimulus material appropriately. For example, some candidates failed to understand that they should write about how the researcher investigated social networking sites and not what social networking sites are. Weaker candidates were often quite generic in their approach to explaining methodology and had difficulties integrating the stimulus material with the questions asked.

There was some spread in the marks awarded and candidates scored all along the mark range with some in the low range, most in the middle and some in the higher range. This is an indication that many candidates were rather well prepared to answer paper 3 questions.

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

The most difficult question for candidates appeared to be question two that involved knowledge and understanding of how reflexivity could be used in the context of the study. Many candidates seemed to have some knowledge of reflexivity as it is used in qualitative research as a way to ensure credibility but some demonstrated limited or no knowledge. The command term “explain”' was often not effectively addressed and use of the stimulus material not always effective to demonstrate the candidate's points.

There was a tendency in some papers to base analysis on speculation rather than 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 considerations before and after the interview and quite a few candidates were able to answer question three on the narrative interview well.

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

Question 1

For question one that asked for considerations that could have been taken both before and after the interview, there were many good answers. Overall, strong responses demonstrated that candidates could apply their knowledge on qualitative research methodology to the stimulus material. Some candidates only referred to ethical considerations and often in a generic way. Some candidates had difficulties integrating the stimulus material. The weaker responses were quite generic and had limited focus on the demands of the question. Considerations “before” referred to sampling, choice of interview method and data recording techniques and ethical issues involved. Considerations “after” often focused on ethical issues alone.

The major problem for weaker candidates seemed to be how to apply their knowledge of considerations before/after conducting the interviews to the study in the stimulus material.

Question 2

This question seemed particularly difficult to a number of candidates and some scored zero or very low marks in this question. Responses at the lower end of the markband typically referred to reflexivity as “considerations” in general thus demonstrating very limited knowledge of what this term means in relation to qualitative research methodology. Stronger responses made reference to both personal and epistemological reflexivity and gave suggestions as to how reflexivity could be applied within the study in the stimulus material. Although knowledge of these terms was not needed to access the higher markbands it was obvious that those who used them were more knowledgeable on application of reflexivity.

Candidates often referred to potential researcher bias in relation to the use of social networking, for example if the researcher was older than the participants. With regard to epistemological reflexivity, most candidates referred to considerations on sampling and whether narrative interview was the most appropriate method.

The weakest responses mostly described what social networking is and argued that an older researcher would not understand how important it is to young people and therefore could not do research on it.

Question 3

It seemed that quite a few candidates were not very familiar with the use of the narrative interview. However, most candidates were able to score marks here. To obtain the higher markbands both strengths and limitations of the use of narrative interviews should be addressed.

Stronger candidates were able to explain that narratives are individual interpretations of the world as it is seen from the perspective of those individuals and that such an approach to investigating personal profiles on a social networking site seemed reasonable. Candidates argued that strengths of this approach were that participants could present their own reality without much interruption and that they could communicate more freely using their everyday language. As for limitations, the stronger responses referred to the large amounts of data generated from narrative interviews or that the researcher may have problems interfering in the interview process without biasing the participants' narratives. Weaker candidates often mentioned that participants risked going off track or the risk of useless data because the researcher did not intervene at all. This shows somewhat limited knowledge of the narrative interview. Some candidates compared the narrative interview to the semi-structured or focus group and used this to justify the researcher's choice of narrative interview. Others said the researcher should have chosen the semi-structured interview or even a questionnaire because such methods gave more reliable data. Such responses scored low marks as they did not focus on the demands of the question.

Recommendations and guidance for the teaching of future candidates

- The main challenge in paper 3 seems to be that candidates must learn to apply their knowledge of qualitative research methods to the stimulus material. A second challenge is that candidates must learn to distinguish between quantitative and qualitative research methods.

- The questions in the examination paper are based on an introduction with an outline of a study or scenario (the stimulus material). This is supposed to 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. 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 and teaching this paper should involve showing candidates how to find relevant parts of the stimulus material that could support explanation of research methodology.

- Paper 3 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 specific study. This is a demanding task as candidates need to be creative in combining knowledge of qualitative research methodology with the scenario in the stimulus material. Therefore candidates should have the opportunity to practise "what it is like to be a qualitative researcher" during their course so that they are able to apply relevant methodological considerations to the stimulus material. This means that teaching should include exposure to a number of qualitative studies and preferably also give the opportunity to conduct some minor research projects in order to get an insight into the reasoning of a qualitative researcher. It is also 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.

- 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 terms in relation to paper 3, 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.