May 2016 subject reports

Psychology TZ1

(IB Latin America and IB North America)

To protect the integrity of the examinations, increasing use is being made of time zone variants of examination papers. By using variants of the same examination paper candidates in one part of the world will not always be taking the same examination paper as candidates in other parts of the world. A rigorous process is applied to ensure that the papers are comparable in terms of difficulty and syllabus coverage, and measures are taken to guarantee that the same grading standards are applied to candidates' scripts for the different versions of the examination papers. For the May 2016 examination session the IB has produced time zone variants of Psychology paper one.

Overall grade boundaries

Higher level

Grade:	1	2	3	4	5	6	7
Mark range:	0 - 9	10 - 19	20 - 29	30 - 42	43 - 54	55 - 66	67 - 100

Standard level

Grade:	1	2	3	4	5	6	7
Mark range:	0 - 9	10 - 22	23 - 31	32 - 43	44 - 55	56 - 67	68 - 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, with a clear understanding of the requirements of the internal assessment (IA) shown. The majority of candidates were aware of ethical issues and included a copy of the informed consent, briefing and debriefing instructions in the appendices. Candidates were also aware of the requirements of an experimental design, although this session there was a notable increase in the number of quasi/natural-experiments, which are not acceptable as there is no true manipulation of an independent variable.

Most reports were based on studies from cognitive psychology and this seems to provide good reports at this level of study. Favourite experiments were, as usual, reconstructive memory and experiments related to schema theory, studies related to the duration of the short-term memory, and imagery versus rehearsal.

Some issues with IAs still persist such as:

- Three or four variables were manipulated although a simple experiment with only two conditions is recommended in the psychology guide.
- The prediction made in the hypothesis was not adequately justified or supported by background research.
- For the descriptive statistics, the use of descriptive statistics was not explained.
- For the inferential statistics, tests were identified but not justified. Also, many candidates failed to include the raw data or calculations of the inferential test chosen.
- Discussions were superficial with limited discussion of the IA results in the light of background research and/or no reference to statistics. Identification of limitations of the procedure was not linked to suggestions for modification.

It should be noted that it is not required to make an exact replication of an experiment. A partial replication is adequate but the candidate's experiment should be closely linked to an actual experiment.

Candidate performance against each criterion

Criterion A: introduction

In some IAs, the research presented was not *explicitly* linked to the hypotheses. Candidates should always clearly describe the research (theories and/or studies) and state how they link to the hypotheses. Contradictory research should not be presented.

It is recommended that three studies and/or theories are presented. It is important that the background research presented logically leads towards the research hypotheses and the background research is explained and analysed in sufficient depth. This allows for the formulation of a clear research hypothesis that in turn will stimulate discussion of results in the light of the background research in the discussion section.

Candidates seemed to have difficulty clearly writing the hypotheses. The variables should be operationalized, that is, made measureable. For example, when measuring memory, 'number of words recalled' should be written rather than, 'better memory'. Also, the wording of the hypothesis should be clear as to what the expected outcome will be.

Criterion B: design

Most candidates stated an appropriate design (repeated measures or independent design) but the choice of the design was not always properly justified (that is, why that particular design was chosen over another). This must be clearly explained.

A number of candidates had problems with operationalization of the IV and the DV (that is, clearly making them measurable).

Criterion C: participants

The target population, that is, the population from which the sample was drawn, was not always appropriately identified. Often candidates confuse the actual sample with the target population.

Overall, most candidates included the relevant characteristics of the participants, such as age, gender, colour-blindness (in the case of the Stroop Effect) and/or English level proficiency. There is no need to include irrelevant details such as socio-economic status.

The sampling technique was largely correctly identified, but the use of the technique was often not explained.

Criterion D: procedure

Candidates should make sure to make reference to all ethical guidelines that were followed. It is also necessary that all materials are referenced in the appendices. Without proper referencing, it would not be possible to properly replicate the experiment.

Candidates should also make clear how the groups are allocated (if independent measures is used). Counterbalancing is also recommended when repeated measures design is employed.

Criterion E: results – descriptive

Most candidates included a graph and a table, with sufficient labelling as well as a description of the results. Only one measure of central tendency and one measure of dispersion is required, however. As in previous sessions, only the strongest IAs explained the use of descriptive statistics, that is, why the particular measures of central tendency and dispersion were chosen.

Criterion F: results – inferential

Most candidates did choose an appropriate test and did justify the use of the test (based on the level of data and the design). At times t-tests were chosen (which is acceptable) but often it was not the most appropriate test based on the particular aspects of the experiment. There seemed to be an increase this session in the number of candidates who did include the raw data or the inferential test calculations. It is important that raw data and all calculations of the inferential test are included in the appendices. If the calculation is performed online, a screen shot of the calculation could be included in the appendices as documentation.

A number of candidates did not make a statement of statistical significance and/or the null hypothesis was not accepted or rejected, which is required for full marks.

Criterion G: discussion

As with previous sessions, this section in the report seemed to present the most difficulty for the candidates, as it often lacked development and analysis. The results of the IA are often only referred to but a discussion is required. Candidates should always refer back to *all* research presented in the introduction and discuss these in reference to their own findings.

Almost all candidates presented limitations, but often in a superficial manner, without rigorous analysis. Limitations should be presented that are relevant to this particular investigation, not limitations of a general experimental nature. There is no need to include the strengths of the design and procedure.

It is also necessary that a conclusion is included.

Criterion H: citation of sources

Candidates often did not include references for research mentioned in the introduction. Additionally, candidates did not use a standard citation method, such as APA, or referencing was not complete.

Criterion I: report format

Generally the report formats were well done. Appendices were well organized and labelled.

The abstract must include a summary of the study as well as the results of the study.

Recommendations for the teaching of future candidates

- It is recommended that teachers help candidates find appropriate background research, that is, a theoretical framework and appropriate studies. Finding relatively simple experiments to replicate is recommended.
- The background research in the introduction should be analysed in sufficient depth so that the aim of the candidate's own research is clearly justified and the experimental hypothesis should be clearly linked to background research.
- For the sample, the number of participants in the experiment does not need to exceed 20 (independent measures design) or 10 (repeated measures design).
- It would be helpful if candidates were given past experiments to read in order to familiarize themselves with the aspects of experimental research. Some candidates would benefit from doing a 'pilot IA' in order to familiarize themselves with the format and procedure of an experimental design.
- It is strongly encouraged that both the teacher and candidates read through the psychology guide to ensure that all requirements are met.
- Candidates should be taught how to properly reference research as often the citation of sources was incomplete or inconsistently presented.
- It is generally recommended that candidates are familiar with scientific standards, and the reading of proper background research should 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.

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 work submitted tended to vary in terms of quality and variety. Studies from the cognitive level of analysis were the most popular choice. Most of the works submitted were a replication of studies on cognition and memory topics in psychology. It was pleasing to notice some candidates replicated recent research. In the majority of cases work submitted was suitable for Diploma Programme level candidates studying psychology at standard level and performed with regard to ethical guidelines. However one issue noticed by moderators was that some candidates replicated overly complicated studies and this cost them marks in the results section as they were unable to analyse the results effectively and still meet the assessment criteria.

The majority of candidates selected appropriate studies, described them well and were able to somewhat link their own results to the findings of the original study in the discussion section.

Candidates often managed to score some marks in introduction, design and participants sections but sometimes did not include a clear aim or justification for design or sampling method and therefore could not be awarded full marks. In procedure and results sections some marks were not awarded due to lack of relevant and precise information. In some cases, candidates forgot to provide a verbal account (description) of the descriptive statistics obtained.

In many reports, the discussion section was superficially written and therefore many marks were lost due to lack of depth in discussing the findings and methodological issues of the conducted study.

This May session, surprisingly, there tended to be more examples of reports that did not meet the criteria for experimental work. Candidates tended to replicate studies such as Bartlett's (The War of the Ghosts) study, Brewer and Treyens' study on memory of objects in a room, or conducted a serial position study where they measured how often a certain word was memorized in relation to the position of the word in a list. Although these studies are meaningful and relevant material for studying in relation to some learning outcomes within the core they are not true experiments in which an independent variable is explicitly manipulated by the experimenter (candidate). Teachers should advise all candidates to make sure that their study is a true experiment. The psychology guide provides the following definition of a simple experimental study: in the study there needs to be a manipulation of one independent variable and measurement of one dependent variable, while other variables are kept constant.

Candidate performance against each criterion

There were many excellent samples showing a high level of knowledge of research methodology. At the lower end, it was apparent that some candidates, although they were appropriately instructed, lacked time management skills and therefore failed to put enough time and effort into conducting good research and writing a detailed report.

Criterion A: introduction

In many reports introductions were well written with most candidates clearly identifying and explaining the study for partial replication as well as presenting a clearly stated aim. However, some candidates failed to do what was required – instead they included additional explanation and studies, subsequently failing to clearly identify the study they were replicating. There is no need for providing a review of several studies in the introduction section. Another problem, occasionally noticed by moderators, was that candidates followed HL requirements and wrote hypotheses, thereby reducing available word count for the detail needed to describe the original experiment or making it more challenging to write a good and thorough discussion with a rather limited word count.

In addition to this, in some cases aims were not stated clearly and sometimes they weren't stated at all. Candidates should make sure that they state the aim precisely and clearly. They need to make a link between the two conditions of the independent variable and indicate how they will measure the dependent variable.

Criterion B: design

Although the statement of IV and DV was generally correct, operationalization of both was often a problem. In some reports the definition of the IV and DV needed to be more clearly stated; they were often too vague and imprecise. Another rather common mistake was stating only one condition (usually the experimental condition).

In addition, the proper identification of the design itself was problematic for candidates from many centres. Too many candidates vaguely identified the design as "experimental or laboratory controlled". It seemed that some candidates could not distinguish the design from the method. In addition the use of the design was not always justified effectively. In some cases candidates provided incorrect justifications or no justification for their choice of experimental design.

Overall, there was noticeable improvement in the identification and discussion of ethical considerations (informed consent, debriefing, etc.). In some rare cases candidates failed to provide parental consent forms where participants were younger than 16 years.

Criterion C: participants

In many cases candidates presented a good description of relevant characteristics of participants including appropriate target population characteristics and identifying their sampling technique. However, many candidates did not justify the use of this sampling technique and therefore could not obtain full marks. The term "random" still tends to be a source

of confusion reflected in the description of participant selection and allocation to conditions. Many candidates tend to use it as an explanation of a sampling procedure that does not have a bias. Random sampling is much more than that and means that every participant has an equal chance of being selected. Sometimes candidates strived for a large number of participants. This is not necessary – candidates should have 15 to 20 participants in their study. Fewer than 15 is risky due to a possible problem of outlier results. On the other hand, large sample sizes are strongly discouraged as they are not practical.

Although a large number of candidates were able to give some characteristics of the sample, in some cases they failed to identify relevant characteristics of participants.

Criterion D: procedure

In the majority of cases, procedures were relevant and clearly described, but in some cases materials referred to were not included in appendices (for example, standardized instructions, tests, questionnaires), which affected the replicability of the procedure. Although this section of the report was usually well written there is still some room for improvement. In addition, complete and detailed debriefing was rarely present.

Criterion E: results

The results section tended to be difficult for many candidates and moderators rarely had the opportunity to give maximum marks. Teachers should clearly instruct candidates on what is required and necessary within this section.

There was a huge variation of results sections and this was linked to the nature of the study being replicated. Simple studies allowed candidates to meet the assessment criteria, but complicated ones led to candidates struggling with the presentation of their results. Too often the use of the measures of central tendency and dispersion was not explained/justified. In addition, some candidates did not effectively describe the results in written form. Another common problem was that some candidates included, in this section, raw data or graphs showing each individual participant's score.

In many reports graphs were not labelled clearly enough for conditions to be recognized – the labelling was either inexact or incomplete. Weaker candidates chose the wrong type of graph (histograms or pie charts to show differences between independent groups). In addition, a number of candidates presented their results in an unclear manner - they did not include percentages, measures of central tendency or dispersion. Some candidates provided several graphs in the results section - presenting the data in a variety of ways, but often not reflecting the aim of their study. Also, occasionally there was incorrect application of statistics. For example, when ordinal levels of measurement were used, there were some candidates who found the mean score in spite of the fact that this is not an appropriate measure of central tendency for ordinal data.

In the results section, candidates should ensure they provide table and figure headings and provide sufficient description of what these reflect. It is important that candidates specifically name their measures of central tendency; do these reflect mean, median, mode? Candidates

should describe what these different scores for experimental and control groups reflect; and importantly what the standard deviation or range imply.

Criterion F: discussion

As usual, the quality of the discussion sections tended to vary. In this May session some improvement was present as discussions tended to more clearly follow the criteria for this section. Many more candidates linked the discussion of weaknesses to the type of design chosen. In addition, conclusions tend to be embedded within the discussion section instead of just added up at the very end.

Unfortunately, those candidates who had not clearly described the study that had been replicated in the introduction tended to have difficulty with the discussion section as well.

As in past years many weaker candidates struggled with the discussion section, particularly in relating their results to the replicated study and in evaluating their own design and procedure. Most of the discussions were focused on less relevant issues such as sampling method and participant behaviour/misbehaviour. In addition, some candidates still gave strengths of their study even though this is not necessary. A frequent problem of some reports was that candidates failed to clearly address limitations and suggestions for further research were often omitted or trivially suggested.

Criterion G: presentation

In general, reports were within the word limit (although occasionally candidates hadn't recorded the word count). In the majority of cases reports used the required format and references were provided. However, full publication details of the replicated study were often not given in a consistent or full manner. Candidates should be encouraged to adhere to one standard referencing system. Abstracts varied in quality from those that were clear and concise summaries to those that included very little specific information related to the replication of an original study. At times it seemed that some candidates finalized their reports in a hurry and therefore some items were omitted from appendices (for example, materials used, standardized instructions, consent form, calculations performed). Too many candidates failed to appropriately label appendices and reference them and/or failed to include necessary documents in the appendices. Reports were generally within the word limit.

Recommendations for the teaching of future candidates

- Teachers must be clear on what the basic requirements of the IA are in regard to which topics/experiments are not appropriate for replication due to ethics so that they can guide candidates to make appropriate choices.
- Also, when choosing an experiment it is relevant that candidates can understand all aspects of the study and focus the aim to what is manageable within the assessment criteria.
- More instruction on the advantages and disadvantages of using various experimental designs and sampling techniques will help candidates justify the use of them in their reports.
- An experimental study for SL should not have more than one independent variable and

one dependent variable.

- In the introduction, candidates should be told to only write about the study being replicated and then clearly state their own aim at the end.
- The results section should clearly provide descriptive statistics related to the aim of the study.
- In the results section, candidates should be told to include only one table with the descriptive statistics and one graph, and to justify/explain their choice of descriptive statistics. Candidates also need more specific guidance in how to describe the results in written form.
- Candidates should be encouraged to check all calculations and include them in the appendix
- More emphasis should be put on the importance of a well-balanced discussion that makes explicit connections between the methodology and the results of their study.
- In the discussion section, candidates should refer to their measures of central tendency and dispersion and engage in meaningful comparison of their own findings with the study replicated.
- Candidates need more specific guidance and examples in how to focus their discussions on design and procedure as opposed to sampling and participant behaviour. They also need more instruction of identifying meaningful, relevant limitations and making a connection with the modifications to address those specific limitations.
- More guidance is necessary in relation to the expected format for the internal assessment (for example, knowing where ethical considerations should be commented on, where raw data should be presented, and where do standardized instructions belong).
- Candidates should be encouraged to proofread their reports before handing them in.

Higher level and standard level paper one

Higher level component grade boundaries

Grade:	1	2	3	4	5	6	7
Mark range:	0 - 4	5 - 9	10 - 13	14 - 19	20 - 24	25 - 30	31 - 46

Standard level component grade boundaries

Grade:	1	2	3	4	5	6	7
Mark range:	0 - 4	5 - 9	10 - 13	14 - 19	20 - 24	25 - 30	31 - 46

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

Some candidates had difficulty making clear links between their knowledge and the requirements of the question. Links were often implicit rather than explicit. In addition, marks were lost because of a lack of focus on the command terms. Candidates often had detailed and accurate knowledge and understanding of the syllabus but this was not always used effectively to address the demands of the question.

What remains the biggest concern is the lack of critical thinking that is directly focused on the requirements of the essay questions and is not just a superficial coverage of evaluation of the studies included.

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

Overall, candidates appeared quite well prepared in terms of course content and could describe psychological concepts, theories and research competently.

Some candidates were able to integrate appropriate and developed critical thinking into their essay responses.

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

Section A

Biological level of analysis

Most candidates were able to identify and briefly describe the effects of a relevant hormone in terms of behaviour. What was challenging was addressing the actual function of the hormone and the majority of responses focused on the effects without addressing the function in any meaningful way.

Most candidates were able to provide a relevant and often well described study to support their response but this was often done without an explicit or well developed link to the demands of the question. Most candidates addressed oxytocin, cortisol or melatonin quite competently but a significant number of those who selected adrenaline in terms of the Schachter and Singer study were less well able to link the hormone to a specific function. Therefore, many responses which focused on adrenaline simply identified the hormone, described the Schachter and Singer study and made a very superficial reference to its function in terms of behaviour as the focus of the description was often on the cognitive aspects of the study. Stronger answers were able to explain the role of the hormone in relation to the fight or flight response. There was a significant number of candidates, however, who were awarded no marks for this question as they referred to a neurotransmitter, usually ACh, and not a hormone.

There were quite a few examples of candidates using animal research, almost exclusively referring to Berthold's (1849) study on testosterone in roosters, to address a hormone, without any effective or explicit link to human behaviour.

Cognitive level of analysis

A large number of candidates did not provide an accurately stated principle. One principle that was particularly problematic was, 'mental processes can be scientifically investigated'. Many candidates who selected this principle stated 'the brain can be scientifically investigated' and lost marks as a result as the study and link also focused on biological factors at the expense of cognitive processes. In addition, when this principle was accurately stated there were many candidates that provided a relevant study to demonstrate the principle without sufficiently justifying the study in terms of the scientific aspect of the study. As a result, the response addressed the demands of the question in a superficial way.

Some candidates, however, were able to gain high marks in this question with a well outlined principle and a relevant and accurately described study as well as an explicit link between the two.

Sociocultural level of analysis

This was the question that was most challenging for candidates with a large number of responses that dealt with the role of situational and dispositional factors in a very superficial way. Though most candidates were able to distinguish between both factors and provide

relevant examples of both, it was difficult for many of them to clearly point out the role of these factors in either a theory or study.

Candidates who used attribution theory or examples of research relating to errors in attribution were often more successful in integrating the role of these factors in terms of behaviour in their answers than candidates who used the examples of Zimbardo or Milgram.

Section B

Biological level of analysis

This was a question which proved problematic for many candidates. A very large number of responses indicated that candidates may have misinterpreted 'one interaction between cognition and physiology' as meaning one study was required. There were many essays that addressed just one piece of research to support their response resulting in an underdeveloped essay that did not do justice to the demands of the question. A large number of candidates found it difficult to address the interaction aspect of the question in a sustained way throughout the response, resulting in essays that were focused on description of the research without a clear and developed focus on the interaction between the factors.

Some candidates lost marks here as more than one interaction was addressed so that an otherwise well written essay lost valuable marks for a lack of focus on the demands of the question. Candidates who wrote about the interaction between meditation and physiological processes were often superficially developed in light of the question as the cognitive aspect of meditation was addressed in a very limited way. It was a concern that some candidates continue to use the study of Phineas Gage which is of no direct relevance to the question in any case in terms of cognition. Likewise, those candidates who included the work of Broca on language production found it difficult to develop their essay appropriately in light of the requirements of the question.

Only a small percentage of candidates succeeded in discussing cognition and physiology with a detailed consideration of how both factors interact in terms of behaviour so that many essays were heavily descriptive rather than discursive.

Cognitive level of analysis

Most candidates were able to identify and explain a relevant model or theory of one cognitive process with supporting studies. However, the explanation of the theory itself was often superficial and a large percentage of responses focused more on the studies rather than the actual theory identified at the outset.

A large majority of candidates addressed schema theory and were able to use relevant research to demonstrate the key concepts involved. Similarly, flashbulb memory theory was also a common choice but these responses often tended to briefly explain what a flashbulb memory is considered to be without addressing the theory itself in any detail.

The less successful responses had a focus that was very much on evaluation of the supporting studies and their methodology or ethical issues rather than weighing up the strengths and

limitations of the theory, or did not sufficiently develop evaluation of the theory and simply addressed a few relevant points in the final paragraph.

Sociocultural level of analysis

What characterized many of these essays was a lack of focus on explaining conformity sufficiently at the outset. Explanations of conformity were often vague or inaccurate and lacking development, with many candidates providing little more than a brief definition of the concept and simply identifying the relevant factors that influence it before launching directly into the supporting research.

Most candidates were able to clearly identify relevant factors and link them to a relevant piece of research. Asch's study was the most popular choice for this essay and candidates showed they could describe the study well and link the factors appropriately in most cases. What was often lacking though was any real depth of discussion of how these factors influenced conformity with only the strongest essays showing any detailed and developed analysis. Again, discussion was often limited to evaluation of the methodology of the studies without integrating this into the overall argument. Stronger responses fully developed the discussion of factors in terms of cultural differences in conformity rates, the justification of the use of deception in Asch's study or the prevailing historical and political climate at the time.

A fairly large number of essays did not gain high marks as they focused on Milgram's obedience study. Likewise, very few essays that addressed Zimbardo's study succeeded in clearly addressing conformity to social roles and, again, did not gain high marks.

Recommendations and guidance for the teaching of future candidates

Candidates should be encouraged to choose research studies carefully when approaching exam questions so that their choice of empirical evidence is directly relevant to the demands of the question. In essay responses, teachers should encourage their candidates to select theories and studies that allow them to build sound arguments. Candidates should also be made aware that one supporting study is insufficient evidence to build such an argument (unless the question specifically allows for just one study).

In short answer questions in section A which require candidates to focus on just one example of empirical evidence, teachers should caution their candidates that in the case where more than one example is provided, only the first of these will be credited.

Candidates need ongoing and regular instruction and clear guidance by their teachers in how to approach exam questions and this should be a priority from the outset of the course.

Candidates should receive targeted support in answering exam questions effectively. This could involve a focus on how to actually structure essay responses so that besides the standard inclusion of theory and research, candidates become more confident in the integration of critical thinking that is appropriate both to the question and the relevant command term. They should also be made aware that they are doing themselves a disservice in essay responses by simply focusing on standard and repetitive evaluation of studies and ethical considerations. Evidence

of critical thinking (criterion B) involves looking at the implications of research findings, offering alternative explanations, addressing practical applications of research, and so on.

Further comments

It is clear that many candidates are depending on websites which have prepared stock answers to the learning outcomes. Candidates need to be alerted to the fact that the research that is included on the sites is very often inappropriate for this syllabus. This was made very clear this session with so many candidates relying on the Berthold (1849) study to answer question 1 and which mostly resulted in very low marks as the research could not be successfully linked to the requirements of the question.

Higher level and standard level paper two

Higher level component grade boundaries

Grade:	1	2	3	4	5	6	7
Mark range:	0 - 6	7 - 12	13 - 16	17 - 21	22 - 26	27 - 31	32 - 44

Standard level component grade boundaries

Grade:	1	2	3	4	5	6	7
Mark range:	0 - 2	3 - 5	6 - 6	7 - 9	10 - 12	13 - 15	16 - 22

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

Candidates often had difficulty understanding the command terms which made the task of addressing the questions with an appropriate approach a bit difficult. This was especially a problem for the questions which required the candidates to 'discuss' and 'compare and contrast'. In many cases, candidates provided parallel descriptions of competing treatments, theories, or models rather than true comparisons and contrasts. Evidence of critical thinking for most of the responses was either limited or weak and explanations often lacked depth in most of the scripts. Also, in questions where the command term "evaluate" was used candidates found it challenging to direct their evaluation towards the concept, instead of the evaluation of the empirical study and, in forming a connection between the empirical study and the concept being evaluated.

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

Overall the candidates were able to maintain a good structure of the essay (introduction, main body and conclusion) throughout the response, even if the focus kept shifting. Most responses indicated very good understanding and knowledge of the concepts. For the two most popular options (Abnormal psychology; Psychology of human relationships), the responses were very well organized, with a balanced argument. Almost all responses clearly provided empirical evidence. Questions 1, 7 and 11 were recognized as challenging and therefore most candidates that addressed these questions provided evidence of good knowledge and understanding of relevant psychological studies and/or theories.

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

Abnormal Psychology

Question 1

This was a very popular question and often the responses reflected good knowledge of approaches to treatment. Candidates chose a range of different approaches to treatment. The most popular were medical approaches to treatment and cognitive approaches to treatment. Other choices were: systematic desensitization, group therapy, eclectic therapy, family therapy, mindfulness-based cognitive therapy (MBCT), and others. The best responses provided two different approaches to the treatment of one disorder and then had a number of points that they addressed when comparing and contrasting. Some candidates chose two broad approaches to treatment. This is a perfectly acceptable approach but it often led to general responses reflecting a lot of detailed knowledge but not much evidence of understanding or critical thinking. In the majority of cases the first couple of pages of the response provided long general descriptions and evaluations of the approaches to treatment with only the final concluding paragraph providing a clear and direct focus on the question by actually providing similarities and differences between the approaches to treatments. In addition, the main problem encountered in the responses was related to not addressing the command term explicitly.

Another problem was that some candidates gave a general and vague description and evaluation of two different approaches to treatment without relating it to a specific disorder.

In higher quality responses, evidence of critical thinking was provided by comparing and contrasting the effectiveness of the two approaches, by discussing the appropriateness for different cultural contexts and by addressing ethical considerations.

Question 2

This question was also popular. The quality of the responses varied very much from responses that were very basic and simplistic to responses which thoroughly discussed several ethical considerations in depth. Responses most often included the following studies: Rosenhan's "On being sane in insane places", Cooper's study on cultural differences in diagnosis, Broverman et al.'s study on gender bias. Usually higher quality responses focused their response on only several ethical considerations and then discussed them in depth. The most popular ethical considerations discussed were the following: self-fulfilling prophecy, effects of labelling, possibility of stigmatization, over-diagnosis of certain disorders in relation to gender and culture, and diagnostic bias. Some candidates unfortunately misinterpreted the question and discussed ethical considerations in research studies rather than in diagnosis.

Question 3

The quality of the responses varied – some candidates provided simple and common sense accounts while other candidates gave thorough discussions of gender variations in the prevalence of several disorders. Most candidates decided to focus on more than one disorder. The most popular choices of disorders were: post-traumatic stress disorder (PTSD), eating

disorders, depression and sometimes phobias. Higher quality responses made clear references to relevant studies including studies related to vulnerability models and life stressors (Brown and Harris, 1978) and study on the differences in cognitive styles (Nolen–Hoeksema, 1994).

A common weakness was that the command term 'discuss' was not well focused on therefore answers tended to be overly descriptive.

Developmental psychology

Question 4

This question was the most popular choice in the developmental psychology option. Most candidates chose to compare and contrast Piaget's assimilation/accommodation model of cognitive development and Vygotsky's contextual approach to cognitive development. These responses tended to reflect good knowledge of the main concepts provided by these authors. However, in many cases the work of Piaget was described in great detail while Vygotsky's work was not as well represented. In some cases candidates wrongly chose theories that were relevant for developmental psychology but not for cognitive development. Examples of wrong choices were: Bowlby's theory of attachment, Erikson's life stages, and social learning theory. In the majority of the responses there was evidence of good knowledge and understanding of one theory, and some evidence of knowledge of the second theory. However, many candidates failed to clearly compare and contrast the two theories and instead rather provided some empirical support and evaluated the theories. Unfortunately in these cases the marks for criterion B were placed in the lower mark band as evidence of critical thinking was not relevant for the specific question.

High quality responses made relevant comparing and contrasting points in relation to assumptions of the theories, methodological considerations of empirical research, and productivity of the theories in relation to application and generating psychological research.

Question 5

This was the second most popular question within the option. This question was usually addressed with limited evidence of knowledge and understanding. Most responses referred to the following theories as part of their response: gender schema theory that stresses the key role of cognitive processes in the development of gender roles; social learning theory that highlights the importance of the social environment and emphasizes the potency of observational and modelling processes; and theory of psychosexual differentiation that is based on the assumption that gender roles are related to genetic sex determined by chromosomes. Some less successful responses tried to address the question by referring to the evolutionary theory or the psychodynamic theory.

The most popular studies selected and described by candidates were:

- Martin and Halvorson's study (1983) showing the role of gender schemas on gender roles
- Fagot's study (1978) showing the influence of parents on gender roles
- Mead's study (1935) showing that gender roles depend upon the society

- Money and Ehrhardt's study (1972) claiming that children are gender neutral at birth.

Many candidates had problems clearly providing evidence of critical thinking related to the question. However, when a clear discussion was included candidates referred to:

- Supporting and contradicting evidence
- The role of sociocultural factors such as media, parental influence or stereotypes
- The interaction of nature and nurture
- Methodological and/or ethical considerations of the supporting studies

Question 6

Very few candidates responded to this question. Most responses focused on definitions and discussions of resilience with little focus on the strategies to build it. In the majority of cases it seemed less prepared candidates chose this question. The concept of resilience seemed not to have been understood by many candidates. Often these low quality responses included description of Koluchova's case study of identical Czechoslovakian twins, Genie and even sometimes Money's case study of David Reimer without linking these studies to the topic of resilience. It seems that this question might have encouraged candidates to provide different responses on the effect of trauma and deprivation on later development. Stronger responses referred to:

- Social programmes for youth such as Head Start or the Big Brothers Big Sisters Programme (Tierney et al., 1985)
- Programmes dealing with parental education (Sanders et al., 2002)
- Programmes developing skills to protect and promote well-being (for example, cognitive-behavioural therapy (CBT) and social skills training)
- Stress inoculation training

In most cases evaluation was attempted in a limited manner which most often included reference to methodological, cultural and gender considerations as well as the issues of age and/or maturity of the individual.

Health psychology

Question 7

This was a very popular choice within the option. Many responses reflected good knowledge and understanding of relevant studies and theories related to physiological and psychological aspects of stress. For physiological aspects of stress the following were most often addressed:

- The role of the brain in the development of stress and the mechanisms that exist in the brain that seek to minimize stress (Hegel et al., 1989)
- Adrenal responses to environmental stressors
- The role of cortisol depletion on post-traumatic stress disorder (PTSD)
- The connection between stress and the immune system

Psychological aspects of stress most often addressed were:

- How an individual appraises a situation, that is, cognitive appraisal

- Attributional style, either positive or negative
- Perceived threats to one's "social self"
- Role of personality in managing stress

Research studies/theories were often well presented and included the following choices:

- Canon's fight or flight theory (1914)
- Selye's general adaptation syndrome model (1956)
- Kiecolt–Glaser et al.'s (1984) natural experiment to investigate whether the stress of an important exam had an effect on the body's immune functioning

The best responses tended to consider a small number of physiological and psychological aspects of stress and therefore demonstrated depth of knowledge.

A common weakness in responses was that the command term 'discuss' was not well focused on therefore answers tended to be overly descriptive. In some (rare) cases candidates discussed only physiological or only psychological aspects of stress.

Question 8

This question was not a popular choice and mainly attracted the attention of less prepared candidates. Very often the responses tended to be too general and vague with minimal evidence of critical thinking.

Sociocultural factors most often addressed were socio-economic status, education, cultural norms, sociocultural norms, and the influence of media.

Some candidates also gave reference to biological and/or cognitive factors in order to address the command term "to what extent". This approach was usually successful and these responses usually gained above average marks.

Question 9

The command term "evaluate" seemed to be often ignored by candidates as they only provided thorough descriptions of relevant treatments.

Most often candidates referred to nicotine replacement therapy, drug treatment, mindfulness-based stress reduction (MBSR), eclectic treatment, or group therapies.

In some cases candidates tended to ignore the number "two" provided in the question and instead evaluated one or more than two treatments – in both cases responses failed to receive higher marks since there was a lack of focus to the specific question.

Psychology of human relationships

Question 10

This was a very popular question that most candidates didn't have problems with. However, some candidates launched into a discussion of prosocial behaviour rather than theories of altruism, accompanied by an overly long discussion of the case of Kitty Genovese.

Most responses chose one of the following theories:

- Dawkins' selfish gene theory.
- Kin selection theory.
- Trivers' reciprocal altruism theory.
- Cialdini's negative-state relief model.
- Batson's empathy-altruism model.

Many responses reflected good knowledge and understanding of a theory explaining altruism in humans.

However, a common weakness was that candidates failed to form a connection between the theory/concept/term and the empirical studies they provided. Therefore evaluation provided was not aimed at the concept but rather applied to empirical studies. Therefore the requirements of the question were often not directly addressed.

Question 11

The command term "analyse" was often not clearly addressed. It requires candidates to bring out (emphasize) essential aspects of the role that culture plays in the formation and maintenance of relationships.

Candidates addressed this question by referring to the role of individualism versus collectivism; the difference between continuous versus discontinuous cultures; the idea that in some traditional cultures chastity and homemaking skills are more valued in women; evolutionary theory which suggests there are universal patterns in the formation and maintenance of relationships; and the role of cultural norms in the formation and maintenance of relationships.

Studies often included in the response were:

- Yelsma and Athappilly's (1988) comparative study of arranged marriages and love marriages.
- Buss et al.'s (1990) study of international preferences in selecting mates (a study of 37 cultures).
- Levine et al.'s (1995) study on the role of love in the establishment of marriage
- Buss's (1994) cross-cultural study of relationships.
- Ahmad and Reid's (2008) study of communication styles in arranged marriages.

Evidence of critical thinking was usually provided by including evaluation of relevant research and also by comparing and contrasting cultural similarities and differences in relationships and discussing interaction between biological and cultural factors.

A handful of candidates attempted this question by identifying and describing relevant studies but they unfortunately failed to discuss how the findings of the studies were relevant to the question.

Question 12

This was a very popular choice within the option that tended to attract well prepared candidates.

Effects of exposure to violence were clearly addressed by referring to physiological responses to stress; cortisol depletion leading to chronic fatigue; effects on mental health; lower performance in school; psychosomatic illnesses; the circle of violence; and/or delinquency.

Responses included reference to the following studies:

- Shalev and Freedman's (2005) study on PTSD following terrorist attacks.
- Kumar et al.'s (2005) study on the effect of domestic violence on mental health in Indian women.
- Shuster et al.'s (2001) study of stress responses to exposure to terrorism during 9/11.
- Hyman's (1990) study of long-term exposure and depression.

Although the knowledge component was very clearly provided within the response, candidates tended to have more difficulty in providing clear evidence of relevant critical thinking. Higher quality response addressed issues such as difficulty in obtaining empirical evidence, difficulty in defining terms, for example, what is considered bullying, and gender and cultural differences.

In the majority of cases responses referring to a small number of effects tended to demonstrate depth of knowledge and gain higher marks.

Sport psychology

Question 13

Although very few responses were made to this question, the general level of response was not very good. This appeared to be a difficult and unpopular question for candidates.

Question 14

This was the most popular question within the option and it tended to attract the attention of candidates who clearly had studied the option. Responses most often referred to the inverted-U hypothesis; the catastrophe model; optimal arousal theory; and/or the theory of self-efficacy.

Candidates had more problems including evaluation of the selected theory but some candidates successfully addressed the following issues:

- Methodological considerations
- The accuracy and clarity of the concepts
- Contrary findings or explanations
- The productivity of the theory in generating psychological research
- The applications of the empirical findings.

Candidates who addressed one or a small number of theories tended to provide better responses as they managed to provide knowledge and focused evidence of critical thinking.

Question 15

This was a rather popular choice within the option but it was often chosen by candidates who obviously had little specific knowledge of the topic. Most candidates provided common sense

knowledge of the topic. In several cases candidates addressed only general theories/models of stress without linking them to athlete response to stress.

Discussion was rarely presented in depth but when it was it included cultural, gender and methodological considerations.

Recommendations and guidance for the teaching of future candidates

- In the majority of cases the main problem lies in not being able to interpret the command terms. Therefore, from the very beginning of the course, candidates should be familiarized with the command terms and be exposed to similar kind of questions as those given in the IB papers, so that candidates are well prepared for the final IB exams.
- Candidates need to understand the importance of answering the question being asked. An outpouring of knowledge, although it can be impressive, is of no value if it is not relevant to the question being asked.
- Teachers should try to encourage candidates to form a connection between the theory/concept/term and the 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.
- Teachers should encourage candidates to focus on the main aspects of psychological research including the name of the authors, type of study, where it was conducted, main aspects of the procedure and the findings of the study. This will ensure accuracy of the description.

Higher level paper three

Component grade boundaries

Grade:	1	2	3	4	5	6	7
Mark range:	0 - 1	2 - 3	4 - 6	7 - 9	10 - 13	14 - 16	17 - 30

General comments

A considerable number of candidates seemed to have understood the stimulus material well. Most candidates made some reference to the stimulus material in their responses. However, in some responses there seemed to be confusion about how the field diary was actually used in the study. Some candidates also had problems understanding the role of the researcher in the participant observations and seemed to believe that the researcher was either a patient herself or a real participant. As usual, many candidates used the term 'experiment' as a generic term for 'study' and this is perhaps an area of attention for future consideration in teaching qualitative research methods.

There were some very good responses, indicating that some candidates are well prepared for paper three. Stronger responses demonstrated accurate knowledge of qualitative research methodology and were able to apply it competently to the stimulus material using appropriate concepts and arguments. This was very pleasing as it shows that it is possible for candidates to acquire accurate knowledge of qualitative research methods and apply this competently to the stimulus material.

Overall there was a tendency in weaker responses to not focus on the questions asked but instead to comment on the stimulus material without any reference to knowledge of qualitative research methods. For example, candidates addressed the therapy offered, the setting, or ethical issues involved in carrying out research in a hospice, or attempted to interpret the emotions of the terminally ill patients rather than addressing the fact that the stimulus paper represents a qualitative case study.

In general, weaker responses used limited relevant concepts and knowledge of qualitative research methods. As in previous years, weaker responses showed a tendency to reason based on quantitative research and some candidates actually demonstrated very limited understanding of qualitative research methods and the case study, for example by referring to a lack of quantitative data as a limitation (in question 2).

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

In question 1 there seemed to be some confusion for some candidates about the fact that the focus of the question was on the possible effect(s) of participant expectations on the data collected and not what they as terminally ill patients expected of the therapy as such.

Question 2 generally seemed to offer most problems for candidates as they either did not have sufficient knowledge of a case study and why it was used, or they simply described what was done in the study without really addressing the question.

Although question 3 was overall well-answered by candidates, weaker responses to this question demonstrated a tendency to argue based on the assumption that researchers are by nature so biased that they want their own ideas to be expressed in a study. This demonstrates a general lack of understanding that the purpose of qualitative research is to understand the subjective world of participants, that it is not a problem that a researcher has a personal interest in the area of investigation and that reflexivity is an approach to take biases into consideration in order to ensure that it is actually the participants' experiences that are documented.

It seems that some candidates still have a tendency to comment on the study in the stimulus material instead of explicitly addressing the questions related to methodology. Weaker candidates had problems integrating the stimulus material into their response in a meaningful way. Either they relied on heavy citation of the stimulus material in their answer or they treated the stimulus material as a text to analyse rather than using relevant parts of it as support for their arguments on methodological and ethical considerations. Some candidates had problems with the command terms and lacked an understanding of how to address the command 'explain' or 'discuss'.

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

Most candidates used the stimulus material to some extent and the stronger responses integrated relevant knowledge of qualitative research methodology with relevant parts of the stimulus material. With reference to question 1, most candidates had some knowledge of the effects of participant expectations in research. In question 3 quite a few candidates demonstrated good knowledge of reflexivity and were able to support their explanation with appropriate quotes from the stimulus material, for example referring to the researcher's job and the longitudinal design of the study. This indicates that some candidates have been well prepared in spotting relevant details in the stimulus material and using such details adequately in support of their argument.

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

Question 1

Generally it appeared difficult for some candidates to write an answer with sufficient depth. Although it was obvious that they had some understanding of participant expectations many candidates seemed to lack accurate knowledge.

Stronger responses explained what was understood by participant expectations, for example defining it as “ways in which participants may alter their behaviour because they know they are part of a study" and that this may affect the credibility of the study providing relevant examples. For example, stronger candidates were able to explain that giving informed consent meant that participants probably knew what the study was about and that this could bias the data in a number of ways, referring to participant bias, the Hawthorne effect, reactivity, demand characteristics, conformity or the social desirability effect. Stronger candidates referred to the overt character of the participant observation as well as social processes in a focus group interview as factors that could influence participants' behaviour.

Weaker candidates had difficulties defining participant expectations. Some gave only irrelevant reasons for why the participants should want to sabotage the study. This resulted in superficial description of the study, the participants, or the researcher with no reference to the question asked.

Question 2

The command term "discuss" seemed to be a major challenge in this question and candidates generally seemed to struggle to answer the question. Although many candidates demonstrated at least some knowledge of the case study and/or the way data was collected in the case study, responses tended to be somewhat unfocused. Most candidates referred to the two data collection methods used. Stronger candidates also addressed why these two methods could be appropriate in the context of the case study in the stimulus material, for example, with reference to strengths and limitations of the methods in relation to the area of investigation of the case study. Some addressed characteristics of a case study referring to the aim of the study as 'the case' and developed this further, thus demonstrating accurate knowledge.

Stronger responses defined the case study as a way to gain an insight into a unique phenomenon (case), often of a sensitive nature (as in this study). Such responses referred to triangulation with reference to the data collection methods (overt participant observation and focus group interviews) and were able to discuss the choice of these methods in this particular case study. Issues of generalization from single case studies were often competently discussed, for example with reference to the possibility of theoretical generalization.

Weaker responses failed to address the question asked about the use of a case study and rather wrote about the methods used in a generic way without relating this to the stimulus material. Some also wrote about ethical issues or suggested that the researcher used a different sample or an alternative method. Weaker responses had a tendency to repeat what was said

about participant expectations in question 1 or writing in a generic way about the case study and then repeating it again.

Question 3

Many candidates demonstrated at least some knowledge of reflexivity as a means to discover possible biases in the research process. Stronger responses approached this question with explanations of why reflexivity could be important to apply in the study and referred to appropriate details in the stimulus material. These included the fact that the researcher worked at the hospice and knew many of the patients; that the study was longitudinal, and therefore she could lose objectivity; and that she had used several data collection methods and asked an external researcher to check her analysis of the data. Many candidates actually referred to epistemological reflexivity and triangulation as well as personal reflexivity with reference to the field diary in which she recorded her own thoughts and feelings during the research.

The weakest responses did not explain what reflexivity is and why it was used in this particular study. Some weak responses also stated that the aim of reflexivity is to remove all bias while others obviously had misunderstood reflexivity and suggested ways to avoid reflexivity in order for research not to be biased. Weaker responses also had a tendency to just repeat what was said about participant expectations in question 1 or write in a generic way about reflexivity.

Recommendations and guidance for the teaching of future candidates

Paper three is based on a short description of a qualitative research study (the stimulus material) accompanied by three questions related to the methodology used in that particular research study. Candidates must answer all three questions paying attention to the command term and using their knowledge of qualitative research as well as information from the stimulus material to support their analysis. Candidates should be trained in addressing each question in a straightforward manner and avoid “filling in” with general knowledge that is not directly relevant to the question asked and will therefore not be awarded credit. Unfortunately many candidates write both introductory and concluding paragraphs that take up too much of answers and contribute little in terms of knowledge and understanding. It is therefore important to instruct candidates that a short-answer question does not require an introduction and a conclusion.

Although candidates this May generally referred to the stimulus material in some way there are still issues with this. It was often seen that candidates commented on the content of the stimulus material in a common sense way with limited knowledge of qualitative research methods. Preparation for the exam is best done using past exam papers for training so that candidates will get an opportunity to acquire an understanding of how to use the stimulus paper and combine relevant points in it with accurate knowledge and understanding of qualitative research methods.

Ideally, teaching towards paper three should include exposure to a number of qualitative studies to give candidates more opportunity to understand the philosophy of qualitative research. The optimal strategy is that candidates conduct small research projects on each of the methods in

order to get an insight into the reasoning of a qualitative researcher as this would be very useful in relation to developing the thinking skills necessary for paper three.

It is also recommended that teachers provide opportunities to practise the command terms in relation to paper three. Too many candidates still have problems here so understanding what a specific command term requires in paper three should be part of effective teaching.

Finally, it is recommended to prepare candidates in such a way that they have both (1) a general knowledge of qualitative research methods as outlined in the guide and (2) competence in applying this knowledge in relation to the stimulus material as well as (3) competence in using appropriate terms and concepts from qualitative research methods. It is also recommended that candidates are trained to write balanced evaluations and discussions instead of presenting personal opinions or speculations with limited relevance to the questions asked.