

May 2017 subject reports

Psychology TZ2

(IB Africa, Europe and Middle East and IB Asia-Pacific)

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 2017 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 - 31	32 - 43	44 - 56	57 - 67	68 - 100

Standard level

Grade:	1	2	3	4	5	6	7
Mark range:	0 - 9	10 - 22	23 - 33	34 - 45	46 - 56	57 - 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 good standard, with the majority of centres showing an understanding of the requirements of the internal assessment (IA). The majority of IAs submitted met the criteria for an experimental design, and most IAs also employed two conditions (and not three or more), as recommended. Many of the more successful IAs were based on studies from the cognitive level of analysis, such as effects of interference and memory recall, cue dependent memory, reconstructive memory, experiments related to schema theory and imagery versus rehearsal. Some newer research was presented which was pleasing to see. A growing number of IAs, however, investigated studies from the sociocultural level of analysis, specifically compliance techniques such as foot-in-the-door technique, as well as attribution errors, such as self-serving bias. These should be investigated with caution as the degree of deception was at times unethical. Foot-in-the-door was also done as a survey on a few occasions, and not an experimental design.

In general, the weaker reports shared the following characteristics:

- Background studies and/or theories were not clearly explained and/or made relevant to the hypotheses.
- More than two conditions were manipulated.
- The hypotheses were not operationalized, that is, made measurable.
- The IV and DV were not clearly operationalized and made measurable.
- The target population was not clearly identified.
- The use of the descriptive statistics was not explained.
- Discussions were superficial with little or no consideration of their results in the light of background research and/or no reference to statistics.
- Referencing was not a standard format or not complete.

It should be noted that candidates are not required to make an exact replication of an experiment. A partial replication is adequate. However, the candidate's experiment should be closely linked to a published experiment. A few candidates also 'combined' studies and replicated various aspects of each; this is strongly discouraged.

Candidate performance against each criterion

Criterion A: introduction

In many IAs, the research presented was relevant and adequate in number, but could have been explained in more detail. Many candidates also found it challenging to explicitly state how the research was highly relevant to the hypotheses. Candidates should also be aware that the background research

should always logically lead towards the candidate's own research hypotheses, which in turn allows for the formulation of a clear research hypothesis.

It is also critical that the variables stated in the hypotheses are operationalized (that is, made measurable) and a statement of significance made.

Criterion B: design

The design (repeated measures or independent groups) was not always properly justified. When identifying the IV and DV, candidates often did not operationalize them (that is, clearly make them measurable).

Overall, candidates had a good understanding of the ethical guidelines. It is permissible to use participants under the age of 16 if parental consent is given. This should be stated in this section when discussing ethical considerations.

Criterion C: participants

Candidates neglected to state the target population, that is, the population from which the sample was drawn. Candidates also often confuse the actual sample with the target population.

Most candidates did identify the appropriate sampling technique, but some didn't explain the use of this method, that is, how this method was used.

Criterion D: procedure

It is necessary that all materials (for example, informed consent form, standardized instructions, etc) are referenced in the appendices. Without proper referencing, it would not be possible to properly replicate the experiment. Candidates must also make clear how the control and experimental groups differed.

Criterion E: results – descriptive

Only the strongest candidates explained the use of descriptive statistics, that is, why the mean was chosen as the measure of central tendency and/or the standard deviation as the measure of dispersion. Most candidates included a graph and a table, but proper labelling of the graph was an issue with incorrect labelling or no labelling at all. Candidates must also present the results in narrative form as well as in a table and graph. Only one measure of central tendency and one measure of dispersion should be presented.

Criterion F: results – inferential

Many candidates chose an appropriate test and justified 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 (sample and/or variance of data). The statement of significance should also always be appropriate and clearly stated.

It is important that all raw data and calculations of the inferential test are included in the appendices. If the calculation is performed online, a screen shot of the calculation should be included in the appendices as documentation.

Criterion G: discussion

Candidates should always refer back to all research and/or theories presented in the introduction and discuss these in reference to their own findings. Candidates who included research and/or theories in the introduction that were not highly relevant often struggled with this aspect of the discussion.

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 experiments in general. Also, certain aspects of the study that were presented as limitations could have been easily avoided with the use of a pilot study, thus indicating a limited amount of planning which should be avoided. Strengths are not relevant and only take up unnecessary word count.

Modifications should be based on the relevant limitations of the study. It is also necessary that a conclusion is included for all IAs.

Criterion H: citation of sources

Referencing continues to be an area of weakness. Candidates often did not cite research within the text or use a standard citation method (such as APA) in the reference section.

Criterion I: report format

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

Recommendations for the teaching of future candidates

- Assist candidates in selecting an appropriate experiment to replicate with an appropriate theoretical framework and background research. Finding relatively simple experiments to replicate is recommended. Again, it is advised that candidates do a partial replication of studies rather than try to 'create' their own study. Only two variables/conditions should be manipulated.
- For the sample, 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 approximately this number.
- It would be helpful if candidates were given published research articles from journals to read in order to familiarize themselves with the aspects of experimental research.
- Candidates and teachers should be fully aware of the assessment criteria to ensure that all guidelines are met.

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

In general, the studies used were varied and mostly easy to replicate. There was a varied and interesting range of work submitted. Experiments from cognitive psychology were the most popular choice. Most of the works submitted were a replication of studies in the range of perception, cognition and memory topics in psychology. The most popular two studies replicated were on Stroop effect and the Loftus & Palmer study on effect of leading question on speed estimation. Also popular were replications of the following studies: Roediger and McDermott, Glanzer & Cunitz, Tversky & Kahneman and Craik & Tulving. However, there were some studies that were replicated in the realm of social psychology that were concerning, especially when performed on young children.

In the majority of cases, work submitted was suitable for Diploma Programme level candidates studying psychology at standard level and conducted with regard to ethical guidelines.

Most candidates described the original studies rather well and were able to somewhat link their own results to the original study in the discussion section.

Candidates often scored full marks in participant and procedure sections but failed to include justifications of the design or operational definitions of the independent and dependent variable and therefore could not be awarded full marks for the design section.

In the results section, candidates often failed to gain maximum marks for lacking details and the discussion section was where many marks were lost due to lack of depth in discussion. There were some examples of reports that did not meet the criterion for experimental work, but they were few.

Candidate performance against each criterion

There were some very solid samples showing a high level of ability. At the lower end, it was apparent that some candidates were not appropriately instructed and therefore failed to put an appropriate internal assessment together.

Marks tended to be rather frequently missed on criteria B, C, and D when candidates were not precise in their explanations, justifications, and descriptions. An additional general weakness noticed by some examiners was that candidates were reluctant to go beyond a descriptive approach in the discussion section. References were not presented fully in a few samples. Candidates should be instructed that referencing should occur whenever a study/theory is described in the introduction.

Criterion A: introduction

In higher quality reports the introductions were well written, reflecting that candidates were able to clearly identify and explain the study for partial replication as well as present a clearly stated aim.

However, in many reports the aim was stated in non-specific terms or not at all. The study to be replicated was usually correctly identified but often not described in sufficient detail. A number of reports focused on theory rather than the study itself.

In many samples, where full marks were not awarded in criterion A, it was due to the lack of results included in the description of the study to be replicated.

Criterion B: design

Although identification of IV and DV was generally correct, operationalization of both variables was often a problem. In some cases only one condition (the experimental condition) was stated. In addition, the proper identification of the design itself was still problematic for candidates from many centres. Too many candidates vaguely identified the design as just “experimental”. Also, some candidates wrote inaccurate or incorrect justifications or no justification for their choice of experimental design. In relation to providing evidence of respecting ethical considerations, some candidates failed to include a copy of the parental consent form in the appendices when using minors as participants.

Criterion C: participants

Often candidates presented a detailed description of appropriate target population characteristics and correctly identified 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” is still often misused and tends to be a source of confusion. Candidates from some centres tended to incorrectly identify their sampling method as random or wrongly indicated that two sampling methods were used– random and opportunity – these terms seem to be often misunderstood as they are simultaneously applied to both participant selection and allocation to conditions.

In some cases reports lacked sufficient participant detail. For example, for those replicating Stroop, attention to the interference of possible colour blindness in the participants was often overlooked as was driver experience in Loftus & Palmer.

Criterion D: procedure

In the majority of cases, procedures were relevant and clearly described, but in some cases materials referred to were not fully included in the appendices (for example, standardized instructions, tests, questionnaires) or were presented in an incomplete manner – for example, providing materials for the experimental condition but not for the control condition. All of this affected the replicability of the procedure and therefore sometimes full marks could not be awarded. Overall, although this section of the report was usually well done there is still some room for improvement.

Criterion E: results

The majority of reports applied some descriptive statistics to data and some candidates were able to clearly explain why these particular measures had been chosen to illustrate the data. Overall, very often the results section seemed to be a weak point in many reports.

Although the majority of reports included a graph often it was not correct or clearly labelled. In many instances when the graphs were not labelled the conditions could not be recognized and therefore graphical presentation could not be awarded marks. In addition, weak reports reflected that candidates chose the wrong type of graph or presented only raw data in the graphs. A few 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 descriptive statistics.

Another common problem was that candidates did not fully interpret their descriptive statistics. Also, calculations (for example, of mean and/or standard deviation) were sometimes inaccurate or missing.

Criterion F: discussion

As usual the quality of the discussion section tended to vary. High quality responses tended to have discussions in which weaknesses of the study are related to the type of design or procedural issues. In high quality discussions, the conclusions tended to be embedded within the discussion instead of just added up at the very end. Most of these reports correctly identified at least one design or procedural limitation with a modification suggested.

On the other hand, the major weaknesses recognized by examiners related to this criterion were: failure to sufficiently compare the obtained results to those of the original study, lack of development in the discussion and lack of rigor in the analysis. In addition, a number of reports focused on the strengths of their replication which restricted the number of words that could be used for other areas in this criterion. Too often, required elements of the results and discussion were missing, suggesting that candidates may need to be more careful, and put in more time and effort, when working on these sections. Relatively few candidates made a serious attempt to analyse their own results and develop discussion of their statistical results. Often when more than one measure of central tendency and dispersion were used the discussion tended to disregard this and focused on just one value. A large number of candidates failed to discuss the obtained measure of dispersion in any meaningful way or compare it to the results of the original study.

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. Full publication details of the replicated study were often not given. Candidates should be encouraged to adhere to one standard referencing system. Abstracts varied in quality from those that were clear and complete with all information required in an abstract to those that included very little specific information. Overall, the majority of abstracts were often written as a general overview and lacked specifics – several examiners reported that candidates tended to omit actual results obtained in their study.

Recommendations for the teaching of future candidates

- Candidates need to be guided to provide specific information in a clear and concise manner.
- Teaching should include clear explanation of aim and procedure for carrying out the aim.
- Teachers should make sure that candidates are replicating a study where there are only two levels of IV and not three (as in the case this year with some Craik & Lockhart replications) or multiple IVs (as in the case of some serial word list studies)
- Criterion E was generally followed but many candidates were unable to connect the central

tendency and dispersion with the aim of the study. Teachers could consider organizing activities for practising examples of how to present data in graph form, table form and description of results.

- For criterion F, candidates were often superficial in analysing strengths and limitations of the assessment. Also, limitations and improvements need to be related to the study specifically conducted by the candidates and improvements need to be realistic and explained. More attention to the discussion is needed, perhaps including some instruction on how to interpret a set of descriptive statistics, and how to focus on elements of design and procedure when analysing the limitations of a study.
- A well-developed abstract has the following: candidate refers to results from measures of central tendency, compares these with the original, highlighting any difference observed, makes explicit connections between the methodology and the results of their study and suggests relevant modifications beyond the sampling and number of participants.
- More guidance is necessary in relation to the expected format for the internal assessment (for example, knowing where ethical considerations should be addressed, raw data presented, standardized instructions belong).
- More attention and guidance should be given to candidates about standard methods of referencing.
- Candidates should be encouraged to proofread their reports before handing them in.
- Candidates often include elements of the HL IA such as a hypothesis in the introduction and discuss significance in their results/discussion sections. Teachers should clearly inform candidates that these additions are unnecessary and often cause a potentially harmful increase in word count.

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 - 14	15 - 19	20 - 25	26 - 30	31 - 46

Standard level component grade boundaries

Grade:	1	2	3	4	5	6	7
Mark range:	0 - 4	5 - 9	10 - 14	15 - 19	20 - 25	26 - 30	31 - 46

General comments

Candidates generally demonstrated good content knowledge but many demonstrated limited evidence of critical thinking. Conceptual understandings were generally strong – for example, social identity theory, schema theory, neurotransmission. Descriptions of research, however, were not always systematic, with superficially stated aims, unclear procedures and vaguely stated results. Very few candidates addressed the command term in the essays; many responses had a systematic approach to critical thinking that often led to arguments that were only of marginal relevance to the question. A more holistic approach to the command term was seen only in the top candidates.

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

For the essays, many candidates wrote about ethics, generalization of the sample and ecological validity, even when it was not highly relevant. There was also a superficial approach to these evaluation strategies as will be addressed with regard to the specific questions below.

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

Overall knowledge of the three levels of analysis was satisfactory with the majority of candidates able to identify appropriate concepts and theoretical explanations and provide relevant research examples.

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

Section A

Biological level of analysis

There were many good descriptions of the process of neurotransmission, but this was not required to earn high marks. Strong answers focused on the role of serotonin in depression, acetylcholine in learning and dopamine in addiction. Stronger responses demonstrated clear understanding of the role of the neurotransmitter and its effect on a behaviour. There were some candidates who wrote several effects of one neurotransmitter, rather than focusing on the demands of the question.

Many students failed to make more than a superficial link between animal research and human behaviour. Stronger responses made explicit links between animal research and examples of how the results can be used to explain human behaviour.

Cognitive level of analysis

There were many strong responses to this question, with many focusing on Bartlett, Loftus & Palmer and Yuille & Cutshall's research. Weaker responses did not focus on the description of the study, omitting the aim, important parts of the procedure or the results.

Sociocultural level of analysis

Many candidates demonstrated a limited understanding of the "etic" concept. Many candidates simply stated that "the etic approach assumes that behaviours are universal", and then gave an example of differences between cultures, only to draw the conclusion that the behaviour was etic. It was clear that many candidates did not understand the link between "universal behaviour" and cross-cultural research.

Stronger responses focused on the deductive approach of etic research and the use of standardized tests, procedures or questionnaires from one culture to another to study a behaviour.

A significant number of candidates did not distinguish between the term "etic" and "ethics."

Section B

Biological level of analysis

There were many very strong responses to this question, focusing on effects such as neuroplasticity, lowered immune response, jet lag, seasonal affective disorder and heart disease. Environmental influences included environmental stressors, social hierarchies, stimulation and poverty. Many candidates focused on a single effect – most commonly, neuroplasticity – rather than discussing two distinctly different effects. Some candidates failed to make a clear link to an environmental factor.

Stronger candidates defined terms such as neuroplasticity and directly addressed the physiological effects. Many candidates did not demonstrate clear understanding of how the environment affected physiology – for example, how stressors lead to activation of the HPA Axis, resulting in suppressed immune system or high levels of cortisol which may lead to hippocampal cell loss.

Few candidates truly discussed the question of the interaction between the environment and physiology. Instead, critical thinking was often of marginal relevance to the question, often focusing solely on ethical considerations. There was also the misunderstanding that using MRI scans to examine brain structure is not ecologically valid. This was the result of a formulaic approach that argued that all research that is done in a laboratory is not ecologically valid, which is not correct.

Cognitive level of analysis

The majority of the responses focused on memory research, addressing the role of MRI, fMRI, PET and CAT scans. Many responses demonstrated a limited understanding of the technology and focused instead on methodological considerations such as sample size, generalizability or ethics. In addition, some candidates chose their examples poorly, describing both the case studies of HM and Clive Wearing which have no clear difference in the use of the technology.

Critical thinking for this question was often weak. Candidates often incorrectly argued that MRIs have low ecological validity, do not have cross-cultural validity, or are open to demand characteristics. In addition, evaluative points were often stated, but not developed. For example, the cost of technology was often seen as a limitation, but this was not explained with regard to the resulting sample size and hence the effect on reliability.

Stronger candidates discussed the use of technology more holistically – e.g. the problems of artefacts, researcher bias and the composite images that are created. They also discussed the strengths which included a more ethical approach to brain research, the possibility for collecting large amounts of data and researcher triangulation.

Sociocultural level of analysis

Stronger responses clearly outlined the theory in some detail, explained relevant research and then evaluated the theory. Several candidates, however, did not evaluate the theory, but only focused on the studies. In addition, many responses only looked at the limitations and did not address the strengths of the research.

Recommendations and guidance for the teaching of future candidates

Evaluation is an area where candidates need to be supported by teachers as it was rarely well integrated throughout the essay. Evaluation of studies alone is not enough for the highest marks in criterion B. Formulaic approaches are not a strategy for success. It is important that candidates develop a “toolbox” of critical thinking that addresses a range of considerations. In addition, to meet the command term “discuss”, a more holistic approach needs to be taken to discuss the question.

Clarity of expression is another issue that teachers should advise candidates on. Candidates should define terms and be sure that theories and studies are clearly explained, not assuming too much knowledge on the behalf of the reader.

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 - 17	18 - 22	23 - 28	29 - 33	34 - 44

Standard level component grade boundaries

Grade:	1	2	3	4	5	6	7
Mark range:	0 - 2	3 - 5	6 - 7	8 - 10	11 - 12	13 - 15	16 - 22

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

The majority of candidates understood the general requirements of most of the questions set for paper two. However, many candidates failed to address the specific requirements of the questions. This concern was often expressed by examiners in relation to responses to questions which required the candidates to 'contrast' (in question 1), 'examine' (in questions 4 and 5) and, to 'to what extent' (in questions 2 and 13). Overall evidence of critical thinking for most responses was rather weak and explanation lacked depth in most of the scripts. Also, candidates found it challenging to direct their critical thinking towards the main issue in the question. Most provided routine evaluation of empirical studies related to the topic. Forming a connection between the empirical study/theory and the specific topic of the question was often a challenge. In addition, some candidates did not maintain focus on the command term and/or the requirements of the question throughout the response, resulting in a weaker response.

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

Overall the candidates were able to maintain the structure of the essay (introduction, main body and conclusion) throughout the response, even if the focus kept shifting. Most responses indicated rather good understanding and knowledge of the concepts. For the most popular options, (Abnormal Psychology; Psychology of Human Relationships), the responses were very well organized and tended to reflect some knowledge of psychological research.

In centres that were well prepared for the examination the questions were clearly addressed, often beginning with defining key terminology that helped in structuring the response. In these cases research studies/theories were included in supporting the key points of the response.

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

Abnormal psychology

Question 1

Although the command term “contrast” required candidates to give an account of differences between one individual approach and one group approach to the treatment of one disorder, the majority of candidates did not write their responses in this manner.

Problems encountered included:

- The command term “contrast” was ignored and instead candidates evaluated one individual approach and one group approach to treatment.
- Candidates provided a general response in which they didn’t indicate the specific disorder but just presented general information about an individual approach and a group approach to treatment.
- Candidates provided a medical treatment and compared it to either an individual approach to treatment or a group approach to treatment of one disorder.

Responses of higher quality tended to refer to cognitive behavioural therapy (CBT) or systematic desensitization for individual approach to treatment and to group cognitive therapy or group mindfulness-based cognitive therapy (MBCT) for group approach to treatment.

Question 2

In the majority of cases the command term “to what extent” was recognized and candidates made clear reference to not only biological but also other factors influencing abnormal behaviour. However, in weaker responses, candidates simply made a summative discussion of biological factors without regard for the command term.

Responses of higher quality tended to address the question by focusing on one specific disorder (usually depression or eating disorders) and then addressing several biological factors including genes, hormones, neurotransmitters and less often the evolutionary explanation. Most candidates chose to address a larger number of biological factors in order to demonstrate breadth of knowledge.

Empirical support tended to be provided in the majority of cases; however, in most cases there was only limited evidence of critical thinking.

Question 3

In the majority of cases, candidates choosing this question tended to provide rather good knowledge of relevant material but often failed to provide evidence of critical thinking specifically related to the question. Response usually referred to the following discussion points:

- how different cultures define abnormality
- different classification systems exist in different parts of the world
- difference in prevalence rates across cultures
- changes in culture over time
- symptoms may be culturally determined

- culture-bound syndromes
- there is a lack of understanding of different cultural norms.

Usually candidates discussed a larger number of cultural considerations in order to demonstrate breadth of knowledge.

Relevant knowledge of empirical studies was often included but critical thinking was often not linked to the requirements of the question – candidates gave evaluation of studies but didn't know how to link this to considerations in diagnosis.

This question occasionally attracted completely unprepared candidates who provided general ("common sense") information, or provided a discussion of Rosenhan (1973) to illustrate validity of diagnosis.

Developmental psychology

Question 4

This question tended to be rather unpopular. Also, when chosen by candidates the responses were not well presented and failed to provide relevant information. In the majority of cases, less prepared candidates addressed this question and provided responses making reference to Rutter's study, case study of Genie or Czech Twins Case Study – Koluchová (1976). The major problem with these responses was that the response was not addressing the question but just providing a description of these studies and possibly evaluating the methodology used. Some weaker candidates launched a description of Harlow's or Ainsworth's work without discussing how this was related to sociocultural factors.

In some cases, candidates provided general knowledge on stereotypes and how they influence individuals' perception of people living in poverty, or provided general knowledge of social learning theory and did not link it to human development but rather to general behaviour of children.

Question 5

There was a general lack of quality in responses to this question. This question was often chosen by candidates who didn't provide much evidence of psychological research. In some cases, candidates only focused on the second part of the question (development of identity during adolescence) and provided a long but inadequate account of Erikson's theory suggesting that body image was an additional factor influencing identity.

Better responses tended to refer to the following psychological research:

- Simmons and Blyth (1987) – the cultural ideal hypothesis
- Ferron (1997) – cultural differences in the way adolescents view bodily changes
- Mead's cross-cultural theory

There was very limited evidence of critical thinking relevant to the question but when it was present it tended to focus on the following factors:

- The difficulty of generalizing the psychological effects of physical changes.
- The development of identity is influenced by the interaction of biological, cognitive and social factors and is not dominated by biology.

- Culture is also a strong determinant in self-perception and body shape perception.

Question 6

Definitely the most popular question for the option and overall a very popular question.

The quality of responses tended to vary. There were some excellent responses reflecting very good knowledge of Piaget's theory, but also responses where candidates provided a rather inaccurate overview of Piaget's theory. Usually this was a sketchy account of stages including some knowledge of relevant studies. Although Vygotsky's theory was discussed in a few cases, these were a definite minority.

Evidence of critical thinking was usually provided in terms of methodological considerations, applications of the theory and providing alternative explanations (Vygotsky's theory).

Health psychology

Question 7

This question was by far the most popular choice in the option. Concerning quality there was a wide range of responses present.

Lower quality responses provided general responses about stress, including everything they have learned about the topic.

Some high quality responses presented a focused approach to the question, including explicit reference to strengths and limitations of strategies for coping with stress. Popular strategies evaluated were:

- problem-focused and emotion-focused strategies (Lazarus and Folkman, 1975, 1988)
- forms of cognitive behavioural therapy such as stress inoculation training (Meichenbaum, 1985)
- social support groups/networks (Brown and Harris, 1978)
- mindfulness-based stress reduction strategies (Kabat-Zinn, 1979)

In some rare cases, candidates addressed ineffective or unhealthy coping strategies, such as drug taking, alcohol abuse, or overeating but these responses lacked empirical support.

Evaluation of the strategies was usually presented by way of research supporting or refuting the effectiveness of these strategies and/or by comparing/contrasting strategies.

Question 8

This was not a popular choice and most responses lacked specific knowledge relevant to the question but rather provided information from TV, news or general knowledge.

In addition, the majority of responses failed to discuss factors but instead provided a description of how certain factors relate to overeating and development of obesity.

Most candidates included biological factors (genetic predisposition, the role of dopamine) and sociocultural factors (sedentary lifestyle and high-fat diet).

When the response included relevant research it tended to relate to biological factors, for example Stunkard et al.'s (1990) study of identical twins reared apart; Volkow et al.'s (2002) fMRI (functional magnetic resonance imaging) study.

When discussion was provided it tended to be presented in the form of addressing cultural and ethical considerations or providing empirical evidence and outlining related methodological factors.

Question 9

This was not a popular choice and usually the responses provided tended to lack focus and specific knowledge. Candidates usually addressed the following factors:

- biological factors such as genetic predisposition to addiction
- cognitive factors such as expectancies regarding the effects of substances
- parental influence
- peer pressure
- the role of advertising and marketing.

In the majority of cases, candidates failed to provide explicit evidence of critical thinking. If it was present then it usually tended to be in the form of evaluating relevant research studies.

Some responses made an attempt to address the question by referring to social learning theory but the response was not really addressing substance abuse or addictive behaviour specifically.

Psychology of human relationships

Question 10

In the majority of cases, candidates tended to provide specific knowledge relevant to the question and seemed to be well prepared for this topic.

Cross-cultural differences most frequently addressed were:

- cultural and societal norms
- different socialization processes in an individual's upbringing
- cultural dimensions (for example, individualism versus collectivism).

Relevant studies included in the response were:

- Levine et al.'s (2001) studies on cultural differences in helping behaviour
- Whiting and Whiting's (1975) research into altruism levels in children from industrialized and non-industrialized countries
- Miller et al.'s (1990) study examining cultural norms and moral values on social responsibility.

Higher quality responses tended to provide explicit evidence of critical thinking in the form of:

- analysis of the methodology
- alternative explanations of prosocial behaviour
- addressing the issue of universality (for example, kin selection theory) versus cultural differences.

Candidates providing a small number of cross-cultural differences in order to demonstrate depth of knowledge usually scored higher marks. In some cases, candidates engaged in a discussion of altruism and the theories involved therein.

Question 11

Candidates seemed to be well prepared for this question although at times they were struggling to provide clear evidence of critical thinking relevant to the question.

Relevant studies and/or theories most frequently addressed by candidates were:

- the importance of self-disclosure (Altman and Taylor's social penetration theory, 1973)
- the role of micro-expressions (Gottman and Levinson, 1986)
- attributional styles (Bradbury and Fincham, 1990)
- gender-based communication styles (Tannen, 1990).

When discussion was attempted it tended to be provided in the form of:

- methodological considerations
- gender differences in communication
- difficulties of carrying out research on communication styles
- ethical concerns when conducting research.

Question 12

This question tended to attract the least prepared candidates taking this option.

Candidates tended to present the following strategies:

- a community based strategy (for example, MACS (Metropolitan Area Child Study), 2002; Olweus, 1993)
- group treatment programs, such as the Duluth model (for example, Robertson, 1999)
- zero tolerance anti-bullying programmes (usually candidates failed to provide empirical evidence for these)
- research into jigsaw classrooms against bullying (for example, Aronson, 1979).

Discussion of the effectiveness of the strategies tended to focus on:

- cultural, gender and ethical issues
- methodological issues
- long-term versus short-term effectiveness
- the difficulties in assessing the effectiveness of a strategy
- contrary and/or supporting findings or explanations.

Some weaker candidates provided a general overview of violence rather than focusing on the effectiveness of strategies for reducing violence.

Sports Psychology

Question 13

There were very few responses to this question. Those candidates who answered this question focused on either social facilitation or social loafing. Those candidates tended to either neglect the empirical evidence, or provide anecdotal evidence of these phenomena.

Question 14

When the question was addressed it reflected that candidates were prepared for the option.

Evidence of critical thinking was present but not always linked to explaining the relationship between team cohesion and performance. Well prepared candidates addressed:

- analysis of negative and/or positive effects
- bi-directionality
- factors other than team cohesion that influence performance
- evaluation of relevant research.

Question 15

This was the most popular choice within the option but it usually tended to attract candidates who were not prepared for this option. Candidates rarely provided evidence of knowledge of psychological research. Most candidates focused on improvement of performance, peer pressure, more rapid recovery from injury and stress reduction as reasons for using drugs in sport.

When empirical evidence was provided it tended to be related to application of game theory (for example, prisoner's dilemma) to drug usage in sport. Some candidates provided case studies taken from the media to use as evidence of athletes using drugs in sports.

Recommendations and guidance for the teaching of future candidates

It seems that the main problem lies in not being able to interpret the command terms. Therefore, from the very beginning of the course, the candidates should be familiarized with the command terms and be exposed to similar kind of questions as those given in the examination papers, so that candidates are well prepared for the final examinations.

Most importantly, teachers should ensure that candidates 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.

Specific psychological terminology and concepts must be taught.

Candidates should not be encouraged to present large numbers of studies that are misremembered, and not made relevant. Instead focus should be made on one or two relevant studies and one or two updated examples, and on applying these to answering the question effectively.

Candidates should be instructed as to the demands of each paper for which they are being examined. In particular, for paper two, more candidates than in the past responded to two or three questions within an option rather than one, or responded to questions from more than one option in SL and more than two options in HL. This is highly inefficient in terms of time management and generally produces shallow responses.

Higher level paper three

Component grade boundaries

Grade:	1	2	3	4	5	6	7
Mark range:	0 - 1	2 - 3	4 - 7	8 - 10	11 - 13	14 - 16	17 - 30

General comments

Overall, candidate performance on paper 3 was slightly lower than last May. Quite a number of candidates wrote introductions to the short answer questions in paper 3. This is not required and it takes focus from the actual response. In some cases, a whole page was used to introduce the answer and what is written in the introduction is often repeated in the actual response.

Many candidates seemed to have a good understanding of the stimulus material and they made some reference to the stimulus material in their responses. However, in some papers there was a tendency to comment on the content of the stimulus material instead of using it to demonstrate knowledge of qualitative research related to the study described in the stimulus material. It also seemed that some candidates did not fully understand that the study investigated what participants thought could be the motivation for cyberbullying and not really their own role in it. For example, some candidates argued that it was very unethical to categorize participants as “bullies” and “victims”. Most candidates had a generic knowledge related to the questions asked but some had difficulties applying it to the stimulus material and some did not even try to do it. As usual, many candidates used the term “experiment” as a generic term for “study” and this is therefore an important area of attention for teaching research methods. There was also a general tendency in question 3 to only address generalization based on probability indicating no knowledge of generalization in qualitative research.

As in previous years, there seemed to be issues with responding to the command terms. Candidates tended to “explain” or “describe” regardless of the command term. This was especially a problem in questions 1 and 3.

Overall there was a tendency in weaker responses not to focus on the question asked but rather to write comments on the stimulus material without any reference to knowledge of qualitative research methods. For example, candidates addressed cyberbullying as a problem in the modern world and offered personal opinions instead of focusing on the actual study.

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

This year, the main challenge for candidates seemed to be the command terms. A frequent problem in many answers was that candidates had generic knowledge related to the questions but found it difficult to apply it to the stimulus material and the questions asked.

In question 1 there was some generic knowledge of semi-structured interviews in most answers but many candidates struggled to discuss that in relation to the study in the stimulus material. First of all,

candidates tended to merely explain what a semi-structured interview is with some detail and offered some strengths and limitations of the method but there was no discussion related to the researchers' choice of this method in this particular study.

Question 2 was overall well answered but some candidates had problems identifying relevant ethical considerations based on the stimulus material. Instead, these candidates offered advice to the researcher on how to treat the victims of cyberbullying or said that it was unethical to conduct such a study in the first place because it could cause further trauma to the victims, or that it was not ethical to classify participants as bullies.

Question 3 seemed to pose the most difficulties and was frequently answered with rather generic or simplistic responses with the majority simply applying their knowledge from quantitative research regarding the size of the sample or sampling method. Candidates also generally struggled with the command term "to what extent".

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

The majority of candidates were well prepared in terms of ethical considerations in qualitative research and many candidates had a good understanding of the stimulus material and used it well. Stronger responses integrated relevant knowledge of qualitative research methodology with relevant parts of the stimulus material. With reference to question 1, many candidates had some knowledge of semi-structured interviews and could apply that to the question asked. In question 2 quite a few candidates demonstrated good knowledge of relevant ethical considerations and were able to support their explanation with appropriate quotes from the stimulus material, for example referring to informed consent and why this was important in general and in this particular study. Many answers also explained issues of anonymity related to the study in the stimulus material. 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. In question 3, stronger responses demonstrated some knowledge of generalization in qualitative research and relevant terminology and were able to identify the possibility of theoretical generalization.

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

Question 1

The command term "discuss" seemed to be a major challenge in this question and candidates generally seemed to struggle to answer the question. This was particularly so in weaker responses.

Stronger responses demonstrated accurate knowledge of semi-structured interviews and were able to apply this adequately to the stimulus material, discussing possible reasons for choosing this method over others with some reference to strengths and limitations of the semi-structured interview in relation to this specific study. Such answers often included reference to alternative methods such as focus group interview and argued that anonymity and group effects would be a problem as a reason for choosing semi-structured interviews in this case.

Weaker responses could not apply knowledge of semi-structured interviews to the study in the stimulus paper but rather offered some generic – and often quite limited – knowledge of semi-structured interviews. Such answers simply referred to strengths of this type of interview but did not connect it to the specific study. Or they listed a range of strengths and weaknesses of the method but did not link it to the stimulus material.

Question 2

This was the question that most candidates seemed comfortable with and many addressed the command term well. Many candidates had a good grasp of relevance of particular ethical considerations in the study as well as why it should be applied in this study. Many candidates referred to a larger amount of ethical considerations, sometimes resulting in a rather generic approach instead of focusing on relevant considerations in the actual study.

Stronger responses approached this question with reference to the ethical considerations mentioned explicitly in the stimulus paper (for example, informed consent, protection of participants, and anonymity) and provided good explanations for their relevance in this particular study.

Weaker candidates seemed to struggle with ethical considerations that were not explicitly stated in the stimulus material, for example, arguing that the researcher did not debrief the participants, or that participants suffered mental problems or even trauma because of the stress they experienced during the interview, thus forgetting that participants could decline to participate in the study. Weaker candidates did not address the command term “explain” effectively as most of them merely described what was done in the study instead of explaining why specific ethical considerations were appropriate in this study.

Question 3

Many candidates seemed to struggle with the command term “to what extent” and defaulted to explanation or description instead.

Stronger responses demonstrated knowledge and understanding of the specific conditions for generalization in qualitative research as well as an understanding of how this differs from generalization in quantitative research. Strong responses were able to demonstrate sound knowledge including specific terminology for generalization in qualitative research, for example, referring to transferability to other settings with similar problems (or inferential generalization). Such responses could apply this knowledge to the stimulus material, for example referring to a small sample and sampling bias. Stronger responses argued that context and sample in the study provided limited possibility for generalization outside similar contexts unless other studies could confirm the findings. The best responses also observed that this study seemed to be the first of its kind and therefore could perhaps create a platform for theoretical generalization if further research should be conducted, and also made reference to the stimulus material in that regard.

Weaker responses demonstrated superficial or no specific knowledge of generalization in qualitative research. Some candidates seemed confused about the difference between representational and inferential generalization and were not able to make a distinction between the two. Many answers did not expand past looking at the demographic components of the sample to answer this question and some candidates confused generalization with validity. Although most candidates had some knowledge of generalization in quantitative research based on probability and statistics and although it is

acceptable to make brief reference to generalization from populations in a question like this, the focus in paper 3 must be on generalization in qualitative research.

Recommendations and guidance for the teaching of future candidates

Paper 3 is based on a short description of a qualitative research study (the stimulus material) accompanied by three short-answer 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 any 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.

It is recommended that teachers provide a number of opportunities to practise the command terms in relation to paper 3. Too many candidates still have problems here so understanding what a specific command term requires in paper 3 should be part of effective teaching. This could be done by using past papers or simply using specific questions from past papers and change the command term to practice answers and have candidates see the difference in answers when the command term changes.

Although candidates this May generally referred to the stimulus material in some way there were still issues here. 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 paper 3 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 3.

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