PSYCHOLOGY TZ1

(IBNA / IBLA)

Overall grade boundaries

Higher level

Grade:	1	2	3	4	5	6	7
Mark range:	0 - 8	9 - 19	20 - 29	30 - 40	41 - 53	54 - 65	66 - 100

Standard level

Grade:	1	2	3	4	5	6	7
Mark range:	0 - 9	10 - 22	23 - 31	32 - 42	43 - 55	56 - 66	67 - 100

Time zone variants of examination papers

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 2014 examination session the IB has produced time zone variants of psychology paper one. Grade boundaries for the different time zoned papers are set separately, and careful judgments are made that are based on criteria for performance level, to account for differences in the papers.

Higher level internal assessment

Component grade boundaries

Higher level

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. As in last year's May session, there was a notable reduction in internal assessment (IA) reports that did not meet the criteria for an experimental study. The majority of candidates were aware of ethical issues and included a copy of the informed consent, briefing and debriefing instructions in the appendices. There were noticeably good attempts at the explaining and describing the background research and making explicit links to the hypotheses.

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

In general, some weaker IAs had the following issues:

- There were several variables being manipulated although a simple experiment with ***only two conditions*** is recommended in the guide.
- For the descriptive statistics, the results were not stated in words (as well as in tabular form) and/or the use of descriptive statistics was not explained.
- For the inferential statistics, tests were absent or not justified. Also, many candidates failed to include the 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 own procedure was not linked to suggestions for modification.
- The referencing was poorly done.

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 sufficiently explained and explicitly linked to the hypotheses. Candidates should also always clearly describe the research presented and *explicitly* state how the theories and/or studies presented in the introduction link to the hypotheses.

It is important that the introduction is clear and focused on relevant background research so that it logically leads towards the candidate's own research hypotheses and the background research is explained and analysed in sufficient depth (for example, aim, procedure, findings). This allows for formulation of a clear research hypothesis in the introduction and to stimulate discussion of own results in the light of the background research in the discussion section.

Candidates seem to have difficulty clearly writing the hypotheses. The variables should be operationalized and the wording should be clear as to what the expected outcome will be.

Criterion B: design

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

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

Criterion C: participants

Overall, most candidates included the relevant characteristics of the participants, although at times irrelevant characteristics were included, for example, socio-economic status.

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

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

Criterion D: procedure

It is necessary that all materials are referenced in the appendices. Without proper referencing, it would not be possible to properly replicate the experiment.

Criterion E: results – descriptive

Often candidates presented the results only in tabular form without describing the results in a narrative form as well. Most candidates included a graph and a table, but often they were not labelled appropriately with sufficient detail. Raw data should only be included in the appendices.

Only one measure of central tendency and one measure of dispersion is required. As in previous sessions, only the strongest IAs explained the use of descriptive statistics, that is, why was the particular measure of central tendency or dispersion was 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 not apply the inferential tests properly.

A number of candidates did not make a statement of statistical significance and/or the null hypothesis was not accepted or rejected.

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.

Criterion G: discussion

As with previous sessions, this section in the report seemed to present the most difficulty for candidates, as it often lacked development and analysis. Candidates should always refer back to *all* research presented in the introduction and discuss these in reference to their own findings. Candidates who included research in the introduction that was 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 a general experimental nature.

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 and guidance 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. Again, it is recommended that candidates do a partial replication of studies rather than try to "create" their own study.

- 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 design) or 10 (repeated measures design), and it is recommended to observe this.

- 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.

- 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

Standard level

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

The range and suitability of the work submitted

The quality of internal assessment (IA) reports submitted tends to suggest that the level of guidance provided by the teachers was generally very good, with a large number of centres submitting reports on appropriate topics. Most work was suitable since they were replications of simple classic experiments. In the majority of cases studies involved manipulation of a clearly identified independent variable to determine its effect on the dependent variable. The most popular topics came from cognitive psychology such as replications of Stroop and Loftus and Palmer. Some interesting works were presented on the halo effect, heuristics, weapon focus and decision-making. The majority of reports reflected that candidates followed the ethical guidelines provided.

The majority of reports had adopted the required format, information was clearly provided and reports were divided into sections according to IB guidelines.

Unfortunately, there were also some inappropriate works such as one-condition studies that do not meet the criteria for a simple experiment as required by the IB.

Candidate performance against each criterion

Criterion A: introduction

Most introductions provided relevant information. Many candidates presented the key study to be replicated clearly and related this to the aim of the study.

However, there were some problems that appeared a number of times:

- Many candidates wrote about three or four studies. A few candidates provided a general overview of a certain topic in cognitive psychology but failed to clearly identify the study they were replicating.
- The candidate's aim cannot just be "to replicate X's study": their aim should be expressed in terms of their desire to investigate the effect of their IV on their DV.
- Many candidates included a null and research hypothesis – this is not necessary for standard level IA reports.

Criterion B: design

Most candidates clearly identified and justified their research design although some candidates gave a definition of the design rather than a justification. Justification of

design was often presented in a very vague manner – by describing the way the experiment was conducted or by claiming that the candidate decided to choose the same design as was used in the original study. Candidates accurately identified the IV and DV in their study. However, few candidates appropriately identified the levels of the IV. In most cases ethical considerations were clearly addressed but at times debriefing letters were not included in the appendix although candidates stated that participants were debriefed after conducting the experiment.

Criterion C: participants

Most candidates were able to present some relevant characteristics of the sample and were able to identify and explain/justify their sampling method accurately. However, there were still candidates who claimed to have used a random sample, when in fact they had used an opportunity or volunteer sample: the fact that participants are randomly allocated to conditions does not mean that they are then a random sample. Moreover, some candidates who were investigating the Stroop effect failed to address the possible relevance of colour blindness.

In weaker reports some candidates did not include the most relevant characteristic of the sample – sample size.

Criterion D: procedure

Most procedure sections were well written. However, a few candidates did not include their standardized instructions and debriefing notes in the appendices. In some cases, the materials list was missing, and/or materials were not referenced to the appendices.

Unfortunately, it was clear from some of the descriptions of the procedure that a lot of testing is taking place in corridors at break or in the corner of the canteen. This should be avoided whenever possible and appropriate conditions should be ensured within the school.

Criterion E: results

This was the weakest part of the report for many candidates. Candidates often applied all descriptive statistics, even if they were inappropriate for the level of data. Few candidates gave reasons for their choice of descriptive statistics.

Some candidates included raw data in this section or graphs showing each individual participant's score. Graphs and tables were rarely clearly presented, often reflecting inexact or incomplete labelling.

Some candidates tried to use inferential statistics which only served to further confuse the results of these candidates.

Criterion F: discussion

In general the majority of candidates provided some discussion of their results in

relation to the original research provided in the introduction section. However, this was often done in a vague and general manner. Usually an identification of the main weaknesses of the study was provided and several suggestions for improvement were stated. However, some candidates overlooked obvious confounding variables that might have affected their study.

The following issues were quite common:

- Moderate to weaker candidates often struggled with analysis, citing simplistic limitations/modifications.
- Many neglected comparing their results to those of the original study or did it in a superficial manner.
- Quite a few failed to include a conclusion.
- Some candidates wrote about the strengths of the study which used their word count although it is not required according to criterion F.

Criterion G: presentation

Candidates still have difficulties in properly citing internet and secondary sources. Some candidates failed to include a reference to the study that had been replicated. Abstracts were variable in quality from those that were clear and concise summaries to those that included very little specific information related to their replication of an original study. Many candidates forgot to provide any calculations in their appendices. Some 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

- The precise nature of what is necessary for a study to be considered an experiment should be examined by teachers and its correct meaning clearly explained to all candidates. It is recommended that candidates replicate experiments in which **one** independent variable is manipulated and the effect of this is measured on **one** dependent variable. When multiple variables are used (manipulated or measured) this almost inevitably leads to confusion and the use of longer/less clear explanations.
- Only one measure of central tendency and one measure of dispersion should be included for criterion E. Candidates need to practise identification of optimal measures depending on the type of data gathered. Teachers should encourage candidates to check their calculations, clearly write them and include them in the appendices. For full marks for criterion G, calculations should be included in the appendices.
- Candidates should develop skills of interpreting results and analysing the data that is collected. Not only does this include how to calculate descriptive statistics but also

how to analyse them. For example, candidates should be able to discuss what it means if the calculated means of the two conditions are different, yet the standard deviation of each condition is different.

- Candidates should be trained in writing a thorough and clear discussion section (perhaps by reading a couple of research articles to become familiar with the idea and style) and they should consult the checklist to be sure that all the IB requirements are met, *e.g.* in terms of discussing statistics from the results sections.

- Understanding of relevant limitations of own research and suggestions for modification should be tied together and it is not enough to say that a particular study should use random sampling and more participants to be valid. The limitations should be stated clearly and they should be linked to the candidate's own experiment.

- Teachers should provide support for weaker candidates, particularly in the organization and formatting of their reports, as well as in teaching them how to construct and label a table and graph.

- Candidates would also benefit from reading the descriptors concerning the internal assessment. They should have a checklist of requirements for each criterion to assist them in organizing their report and ensure they address and include all required elements.

Higher and standard level paper one

Component grade boundaries

Higher level

Grade:	1	2	3	4	5	6	7
Mark range:	0 - 4	5 - 9	10 - 12	13 - 17	18 - 23	24 - 28	29 - 46

Standard level

Grade:	1	2	3	4	5	6	7
Mark range:	0 - 4	5 - 9	10 - 12	13 - 17	18 - 23	24 - 28	29 - 46

General comments

Section A

Responses tended to describe studies in a relatively formulaic fashion, lacking development and links to the specific requirements of the questions. Learning the research itself appears to be an area of strength, the problem lies in how to use the research to answer the question effectively. Choosing appropriate evidence for the short answer questions was also an issue with some irrelevant material selected, particularly in question one and question two. Many candidates used the wrong research for the wrong questions. Moreover, many candidates included evaluation of research studies in their section A responses, which is not necessary and can result in a loss of focus on the requirements of the question. Studies like Little Albert or Pavlov are still inappropriately used to answer questions.

Section B

Many candidates went straight into examples of studies rather than set the question into context (evolutionary explanation, schema theory, attribution theory). In addition, critical thinking was rarely highly relevant to the questions. There were very few candidates who were able to explicitly integrate their critical thinking in their response to the question. For example, candidates rarely provided focused, developed and balanced evaluation of schema theory that went beyond the simple strengths and limitations of the studies given. Often, critical thinking was either a list of advantages and disadvantages as a separate entity after a descriptive account or the argument was repetitive and superficial.

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

What candidates appeared to find difficult was answering the command term for each question. For example, candidates did not adequately outline a principle in question three but only stated it; they did not explain how a biological factor affects a cognitive process in

question two but described a study. Moreover, it appeared difficult for candidates to clearly identify and explain a physiological process in question one, with several candidates using either Phineas Gage to state that changes in personality or behaviour were a physiological process or Bouchard's twin study to address genetics as a physiological process. Similarly, Phineas Gage was also used to address a cognitive process in question two; this was very vaguely justified as brain damage resulting in an inability to make decisions about how to behave. As in question one, Bouchard's twin study was inappropriately used with IQ being addressed as a cognitive process. In addition, few candidates showed a good understanding of evolutionary theory.

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

In question three, most candidates were able to clearly identify a relevant principle of the sociocultural level of analysis and support it with appropriate and well outlined research. Candidates were quite well prepared with regard to supporting schema theory in the cognitive level of analysis essay with relevant and well described examples of empirical evidence and this was also the case regarding attribution errors in the sociocultural level of analysis essay.

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 a suitable study but few candidates described the study in sufficient detail (aim, method and findings) to reach the top markband. Many candidates struggled to explicitly consider the relevant physiological process. For example, they used Rosenzweig *et al.* and described the study, but they failed to articulate that a stimulating (versus deprived) environment resulted in an increase in dendritic branching. In addition, many candidates seemed to confuse a physiological process with a cognitive process, proceeding to describe how the rats "learned more" rather than focusing on the neuroplastic growth. When Avery *et al*.'s study was described, focus on melatonin was generally lost and instead SAD or sleep became the focus. Candidates who did poorly with this question discussed effects of the environment on *behaviour* using studies such as Bandura's, Milgram's or Zimbardo's.

Cognitive level of analysis

The vast majority of responses addressed the hippocampus as the biological factor and used either HM or Clive Wearing as case studies of amnesia. Martinez and Kesner's study of acetylcholine and memory or Newcomer's study of the role of cortisol on memory were also often selected. However, very few candidates presented a clear explanation of how the biological factor affects the cognitive process: the focus was on the description of the study. Many candidates did not identify a specific biological factor but used brain damage or disease in their

response. Several candidates used emotion as either a biological factor or a cognitive factor or used inappropriate studies such as Bartlett or Loftus and as such did not show any evidence of knowledge and understanding of the question. Some candidates used the Kasamatsu and Hirai's study and consequently found it difficult to justify the choice of the study with regard to a cognitive process vaguely referring to perception of the environment.

Sociocultural level of analysis

While most candidates could successfully identify a principle and provide a study to illustrate it, the vast majority of the responses did not outline the principle and very few candidates made meaningful links between their chosen principle and study that went beyond restating the principle. Again, the focus was on the description of the study. Some candidates were unable to correctly identify a principle stating general topics such as conformity or social identity theory as principles. There was a tendency to use a theory and a study to illustrate the principle such as social identity theory and Tajfel's study, ignoring the command for one piece of research (theory or study) so candidates did not get credit for the second piece of research. Some candidates used Bartlett's War of the Ghosts study to illustrate culture influencing behaviour.

Section B

Biological level of analysis

This was the less popular option. Most candidates that addressed this question considered mating behaviour or disgust and were able to present a few studies that support the evolutionary explanation. However, very few candidates were able to present their chosen evolutionary explanation in any great detail. In addition, it was challenging for most candidates to examine the explanation: there was a lack of critical thinking with regards to evolutionary approach as a whole. On the other hand, some answers were of high quality and discussed the underlying assumptions of the explanation, evaluated the empirical evidence and/or presented the strengths and limitations of the explanation.

Cognitive level of analysis

This question was the most popular essay with most candidates choosing to address schema theory related to memory and including classic studies such as Bartlett (though often inaccurately described), Loftus and Palmer, and Brewer and Treyens. Although candidates were mostly able to describe the studies well, not many were able to outline schema theory itself at the outset, with some candidates not able to address effectively what a schema is. Very few candidates were able to actually focus well on evaluating the theory as opposed to just evaluating the studies. Moreover, many candidates evaluated how useful schemas themselves are, pointing out that they help us organize things, but can also be misleading in terms of memory and stereotypes: these are good points, but they are not strengths and limitations of the theory. Answers were mainly descriptive with a few evaluative points addressed as a

list rather than a fuller explanation of each. Very few candidates were able to evaluate schema theory in depth.

Sociocultural level of analysis

This question was often addressed. Candidates could define two errors in attribution such as the fundamental attribution error, self-serving bias and actor–observer bias. Appropriate studies such as Ross *et al*. and Jones and Harris on the fundamental attribution error or Miller and Ross on the self-serving bias were offered in support of the answer. However, few candidates were able to give a good explanation of the biases in relation to attribution theory. Here again, most candidates did not discuss the errors in attribution, providing an evaluation of the studies that they reported but with little relevance or linking to the actual question. However, some candidates presented a thorough discussion addressing issues such as the reasons for the occurrence of the errors or the cultural differences in the expression of these errors, making reference to studies such as Kashima and Triandis's or Cousins's studies.

Recommendations and guidance for the teaching of future candidates

Candidates should be aware of the need for time management in the exam and allocation of marks in questions as some candidates wrote several pages for short answer questions and very little for the essay, thus losing marks in section B answers as they did not expand on any points to show deeper understanding.

Teachers should select appropriate studies for each learning outcome and should spend time in class unpacking each learning outcome so candidates can clearly identify and explain the key concepts involved.

Candidates should be well aware of what is required of them in terms of the command term so they focus the organization and development of their answers accordingly.

Candidates should be encouraged to think of the studies as only the supporting evidence. The response to the question itself happens above that. Candidates should be advised to answer the question first, and then support it with evidence.

Section A questions.

- Candidates should learn how to use their knowledge effectively. They should practise short answer questions with focus on the links from the study to the demands of the question.
- Candidates should be encouraged to get directly to the point. Too often, the candidates are writing trivial or irrelevant points as an inappropriate introduction to the short answer questions.
- Candidates should be advised that evaluation is not required in section A answers and that it can result in a loss of focus on the requirements of the question.

Section B questions:

- Teachers should ensure that candidates are given plenty of opportunity to write essays, as few candidates had well outlined introductions to explain key concepts and provide some theoretical context to the topic under discussion. Likewise, many essays tend to end rather abruptly without linking back fully to the question and actually answering the question.

- Candidates need to be encouraged to focus more on developing an argument. Clearly, candidates need more help in how to shape what they know, how to use only what they need, how to develop a better argument.

- More emphasis is needed on how to evaluate thoroughly. Critical thinking should be more than a list of evaluative points. For example, it is not sufficient to say that a study lacks ecological validity – candidates must explain why it is the case. It is not sufficient to identify the strengths and weaknesses of studies, candidates should also discuss wider implications. Candidates should be advised that critical analysis has to be related to the requirements of the question.

Higher and standard level paper two

Component grade boundaries

Higher level

Grade:	1	2	3	4	5	6	7
Mark range:	0 - 5	6 - 11	12 - 16	17 - 21	22 - 26	27 - 31	32 - 44

Standard level

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

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

No one option seemed to be more problematic than the others. There were both difficulties as well as excellence seen throughout all options.

Very few candidates attempted the sport psychology questions and therefore the evaluation of this option is limited.

Some candidates are still having some difficulty in addressing specific command terms.

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

Without variation, basic knowledge was better demonstrated than critical thinking by candidates. However, most responses were fairly organized and focused upon the question.

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

Abnormal psychology

Question 1

This question was a popular choice for candidates. In general, candidates seemed to select a range of psychological disorders and appropriate treatment methods. Unfortunately, a few candidates selected either a biomedical or group treatment method to discuss which resulted in a loss of marks. One issue for some candidates was also the command term “evaluate” which was not successfully accomplished by a minority of candidates, who chose to instead describe etiology and treatment modality rather than focusing upon the effectiveness of treatment.

Question 2

This was the most popular question both in the option and the entire exam. Although many candidates expressed some knowledge of the difference between abnormal and normal behaviour, this was done in a general way without benefit of supporting research. The most common researchers cited were Jahoda, Seligman and Rosenhan, and Rosenhan. Cultural and statistical norms were also commonly found within the answer, but many times were discussed in a general manner. Critical thinking was present, but was often not adequate to meet the demands of the command term “discuss”.

Question 3

This question also was popular. Most answers focused on depression and eating disorders, as well as a few centred on post-traumatic stress disorder. Many candidates had good knowledge of the different prevalence rates, but did not have good knowledge of the reasons for these differences. Better answers focused on sociocultural factors in producing different prevalence rates. Answers that focused on biological factors were often shallow and even reflected cultural stereotypes.

Developmental psychology

Question 4

This was not a popular question, and the lack of precision in responding may be an indicator as to why it was not more popular. Several candidates discussed identity development, but did not relate it well to physical change, or discussed physical changes without a clear link to identity development. A few astute candidates were able to link research such as Simmons and Blyth (1987) or Blyth, Bulcroft and Simmons (1981) with a reasoned argument linking identity and physical change.

Question 5

This question showed somewhat of a good range of research studies and theories regarding gender roles and culture. Many referenced Mead’s classic studies as well as more modern research. Generally candidates used the strategy of presenting a few studies in more depth, rather than discussing a larger number of studies in lesser detail.

Question 6

Ainsworth and Bowlby received the most discussion among candidates responding to this question. Generally, the discussion was fairly accurate, but in many cases the evaluation was not as precise. A problem arose with some candidates discussing Harlow and Harlow (1962), who did not use good critical thinking skills in linking animal studies to human behaviour.

Health psychology

Question 7

This question regarding stress was the most popular among the health psychology option. Although candidates had fairly good knowledge and understanding of what physiological and social aspects of stress are, the discussions tended to either be mere descriptions, or good evaluations of both aspects. Therefore, scores tended to vary widely between very good and poor.

Question 8

Not a very popular question, this did elicit some good responses. Astute candidates were able to discuss several biological factors present in health-related behaviours, and the more able did well in comparing biological, sociocultural and/or cultural factors in terms of relative importance. However, a wide range of discussion was not observed, with most responses related to either evolution or genetics.

Question 9

This question presented some difficulty for some candidates. The idea of prevention, as per the markscheme, might be interpreted as either prevention of beginning to engage in substance abuse or addictive behaviour, or as prevention from further substance abuse and/or addiction. However, many candidates instead discussed treatment strategies and effectiveness for those who suffer from substance abuse and/or addiction. A small number of candidates did a very good job of discussing prevention strategies and making clear that prevention from further abuse is also a viable strategy.

Psychology of human relationships

Question 10

This question was very popular, and definitely the most popular in this option. Most candidates were able to capably discuss two different theories, and generally selected one evolutionary and one psychological theory. The problem for some candidates came with the command term “contrast”. The descriptions of the theories were reasonably well done, but the contrast between the two was either lacking, or at a very limited level.

Question 11

This question was also popular within this option. Candidates seemed to know about strategies designed to mitigate violent behaviour. Most discussions centred on bullying behaviour, although some were more general. Some responses centred almost entirely upon the role of media without getting too deeply into actual strategies. As with other questions requiring the candidate to discuss effectiveness, this showed a lack of critical thought for some candidates.

Question 12

A wide range of human relations theories and studies were discussed by candidates with varying levels of success. Most did a very good job of description, although some were of course more detailed than others. The problem for some candidates involved measuring the strengths and limitations of the selected theory or study. When discussing studies, for example, some candidates focused the evaluation on methodology and did not venture into deeper thought.

Sport psychology

Question 13

This was the most popular question from the sport psychology option. The responses varied from those which have been seen in the past that are anecdote driven, and responses in which there was substantive research and theory referred to. In general, the centres teaching this option seem to have found more research studies to use when delivering this material.

Question 14

This question was also an improvement over past performance in sport psychology. By far, the most common techniques discussed were mental imagery, self-talk, and practice scheduling. The discussions generally referred more to research and relied less on anecdote than in past exam sessions.

Question 15

This was the least commonly answered question on the exam, and therefore no appraisal of significance may be made.

Recommendations for the teaching of future candidates

- Make clear the different requirements of the different command terms.
- Emphasize the need to use relevant psychological research in answers.

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

This year a number of candidates demonstrated very limited knowledge of qualitative research and/or the requirements for paper three. The purpose of paper three is to demonstrate knowledge of qualitative research methods as well as how to apply it to the stimulus material as outlined in the guide and overall, many candidates demonstrated that they knew this well. Most candidates appeared to have a fairly good understanding of the study in the stimulus material although the topic of romantic love in some cases stimulated candidates to use anecdotal evidence or reason based on their own beliefs. Unfortunately this resulted in quite a number of weak responses that failed to use knowledge of qualitative research methods to address the questions asked.

It appeared that a number of candidates were not at all familiar with the requirements for paper three and argued based on their own beliefs and/or on interpretation of the content of the stimulus paper instead of knowledge of qualitative research methods. The general characteristic of weaker candidates was that they addressed love and romance as the problem to address rather than focusing on the aim of the research study and the method used to investigate the difference of the portrayal of love and romance in Western culture and women's actual experiences.

Something noticed in weaker responses was that candidates did not use the information about “purposive sampling” in the stimulus paper for answering question one. Instead they suggested a range of other sampling techniques.

Also it was noticed that some candidates seemed to believe that researchers only want to have their own ideas confirmed in a study. This led to a very limited analysis where the main point in candidates' evaluation was to identify all the researcher's purposive scheming to confirm her own beliefs in this particular study on heterosexual females' experiences of romantic love and romance compared to the media's portrayal of the same.

Candidates continue to use vocabulary from quantitative research such as experiment for study, experimenter for researcher and hypothesis for aim and furthermore some candidates talked about variables, control of variables and even about cause-effect relationships and correlations in some cases. All this showed lack of knowledge of qualitative research methods and it is important to stress that paper three is actually testing candidates' knowledge and understanding of qualitative research methods – not quantitative.

There was some spread in the marks awarded. Candidates scored all along the mark range with many in the low range, most in the middle and fewer in the higher range. This is an

indication that while many candidates are well prepared to answer paper three questions there are also a substantial amount of candidates who seem less familiar with the requirements for paper three.

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

Many candidates seemed to have problems with critical thinking and evaluation of qualitative research. In weaker responses this was demonstrated in the very unbalanced evaluation of the use of purposive sampling in question one and in evaluation of the use of semi-structured interviews in question three. From reading the responses it could seem that candidates believe that all researchers are biased and only want to have their own ideas confirmed in a study. This is a concern as it demonstrates a limited understanding of research in psychology in general and in qualitative research in particular.

In question one, it appeared difficult for candidates to see that the main characteristics of participants in this study were that they should be heterosexual and women with experiences of romantic relationships. The fact that they were all related to psychology is not really part of the aim but just happens to be so because the researcher recruited them via her own social network. This is of course an extra source of bias, which the stronger candidates did also rightly observe.

In question three some candidates had problems understanding that they should actually evaluate the use of the semi-structured interview in the study and not suggest alternatives.

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

Many candidates appeared well prepared for question two on reflexivity and were able to explain why reflexivity was relevant to the study with relevant knowledge and examples from the stimulus paper.

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

Question 1

A number of candidates did not at all refer to purposive sampling – it appeared they had missed the reference to it in the stimulus material. Such responses referred to sampling methods such as convenience or snowballing. Although snowballing is a subtype of purposive sampling it was clearly stated in the stimulus material that the sample was purposive so saying it was snowballing was simply not correct. Some candidates wrote that purposive and snowballing is the same. Some candidates argued that the sampling was convenience sampling and supported this by saying that the participants were from the researcher's own social network. While that is true the sampling method was still purposive as the participants were selected based on specific characteristics relevant to the aim of the study.

Stronger responses demonstrated a good understanding of purposive sampling as well as the strengths and limitations of this sampling method in relation to the study in the stimulus material, for example referring to a sample that matched the requirements of the aim of the study (heterosexual women who had experience with romantic love relationships). When evaluating the use of purposive sampling in the study stronger candidates used knowledge of purposive sampling and combined it with information from the stimulus material, for example that purposive sampling always raises issues of generalization and in particular if the sample is small as in this case. Stronger responses also noted that generalization is not the main purpose of qualitative research but this was a point that weaker responses did not address.

Weaker responses did not mention purposive sampling at all in their response or mentioned purposive sampling but did not address strengths and limitations of purposive sampling and merely commented on the participants. Some weaker responses claimed that sampling was convenience or snowballing in spite of the fact that purposive sampling was explicitly mentioned in the stimulus material. A number of the weaker responses argued that purposive sampling was a bad choice and suggested that the researcher should have chosen a random sample instead because the sample could then be larger and more representative. This indicates limited understanding of sampling methods for qualitative research such as the study in the stimulus paper.

Weaker responses did not use knowledge of purposive sampling but only referred to information from the stimulus paper, for example that only eight women participated, that they were from the researcher's social network, all from New Zealand and all somehow related to psychology. Although this is all from the stimulus paper it must be combined with knowledge of purposive sampling to access the higher markbands.

Weaker responses also had issues around generalization from the purposive sample. For example, they could argue that because only heterosexual women were used the results could not be generalized to homosexuals or males but this was not the purpose in this study. This demonstrated very limited understanding of the aim of the study in the stimulus material and qualitative research in general.

Question 2

This question was by far the best answered as many candidates seemed well prepared for this question although quite a few also struggled with this question due to lack of specific knowledge of reflexivity in qualitative research. These candidates interpreted reflexivity in more general terms and argued that it was important for both researchers and participants to reflect on love and what it meant to individuals.

Stronger responses defined what is meant by reflexivity in qualitative research and could refer to both personal and epistemological reflexivity. Such responses gave good examples of why it was important in this study, *e.g*. referring to the researcher's personal interest in the topic, her job as a couples therapist and the fact that she was

herself a heterosexual female investigating women's experiences of love compared to the picture of romantic love in the media. The stronger responses also referred to purposive sampling as something that could justify reflexivity.

Weaker responses had difficulties focusing on WHY reflexivity was important in the study and spent most of the response on suggestions of strategies to reduce bias and increase credibility. While this could be a valuable part of an essay on reflexivity it cannot stand alone as the question asks for reasons why reflexivity was relevant.

Some candidates demonstrated no knowledge at all of reflexivity in qualitative research. For example, saying that reflexivity means cause and effect or trying to make up reasons like saying that reflexivity is about how the women talked about struggling for equality. It was also seen that candidates did have knowledge of reflexivity in qualitative research but had problems applying this to the stimulus material.

Question 3

This question seemed difficult for a number of candidates who appeared to lack basic knowledge of the semi-structured interview (SSI) as a qualitative research method. The question asked for an evaluation of the use of SSIs in this study, which meant that both strengths and limitations of using semi-structured interviews in this particular study must be addressed to access the higher markbands. However, it was often seen that candidates evaluated the study instead of the use of SSIs in the study, a strategy that often resulted in demonstrating very limited knowledge of SSIs.

Stronger responses outlined characteristics of the semi-structured interview and referred to both strengths and limitations of using SSIs in the study. For example, candidates argued that a good reason for using SSIs in the study was that the topic of perceptions of romantic love is rather sensitive and therefore could be more fully explored using this method because the combination of closed and open-ended questions gave participants the opportunity to talk more freely about their own experiences and how these were in line with the media's representation of romantic love or not. They also mentioned that the researcher could ask participants to elaborate on specific points and that all in all, this approach would result in richer data. As for limitations of this approach, stronger responses typically referred to problems in exploring themes that had not been prepared and that analysis of data could be time-consuming.

Weaker responses stated that SSIs were the best method of interviewing without arguing why as part of the evaluation and then continued to write about another method of interviewing without linking this to evaluation of SSIs. Many responses referred to the fact that the interview was recorded as part of the evaluation but did not link it to the use of SSIs and rather wrote about the ability to build rapport if the researcher recorded the interview or wrote about the process of interviewing in general rather than the use of SSIs (*e.g.* a lot of recording).

Another characteristic of weaker responses was that only limitations or only strengths of SSIs were mentioned or the response simply focused on suggesting alternative interview methods such as narrative or focus groups as a better approach because participants could talk without being interrupted. This showed a limited understanding of the requirements for paper three.

Recommendations for the teaching of future candidates

- Each exam paper is based on a brief description of a qualitative research study (the stimulus material) combined with three questions. All questions must be answered. The stimulus material is intended to serve as a starting point for analysis of how qualitative research methods could be applied to a specific study.

- Paper three is based on short answer questions, which means that candidates should not write long introductions of a more general kind. It is not recommended that candidates write a general introduction to sampling or interview methods listing all they know about sampling in qualitative research when the question is on one specific sampling method or one specific interview method, as this is not credited. Candidates should be trained in addressing each question asked in a straightforward manner and avoid filling in with general knowledge that is not directly relevant to the question asked and will therefore not receive any credit.

- Since a large number of candidates continue to use terms from quantitative research it is strongly encouraged that candidates are trained in qualitative research so that they become familiar with the methods and vocabulary associated with qualitative research. Likewise it is important that candidates are instructed in the difference between a qualitative and a quantitative approach to research.

- Teachers are encouraged to integrate teaching of qualitative methods as they go along in the course so that candidates get an idea of the differences between quantitative and qualitative research methods. It is obvious that most of the knowledge used to evaluate research in responses to paper three comes from the internal assessment. It could be a really good idea to have candidates practise "what it is like to be a qualitative researcher" in contrast to doing quantitative research, *e.g*. having candidates do activities that enable them to reflect on various aspects of qualitative research methods. This could be done before proper teaching of paper three starts. Teaching paper three should always 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. It is equally important that candidates have trained with previous exam papers so that they become familiar with the requirement of this paper.

- It seems that many candidates have difficulties using the stimulus material properly. A good many discuss the content of the stimulus material without much reference to relevant knowledge of qualitative research methods, or write about qualitative

research methods in a generic way without much reference to the stimulus paper. Using previous exam papers will give candidates an opportunity to better understand how to apply relevant knowledge and understanding of qualitative research methods to the study mentioned in the stimulus material. Every fifth line in the stimulus material is numbered so that candidates may refer to the lines without having to use extensive quotations. This could be used more effectively in the responses and teaching this paper should involve showing candidates how to find relevant parts of the stimulus material that could support explanation or discussion of qualitative research methodology.

- Overall, candidates should be prepared in such a way that they have both (1) a general knowledge of qualitative research methods as outlined in the guide; (2) competence in applying this knowledge in relation to the stimulus material; and (3) 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 claims and speculations with limited relevance to the questions asked.

- Finally, it is advised that teachers instruct candidates in what it means to address the command terms in relation to paper three, for example what "explain" or "evaluate" means. Too many candidates still have problems here so understanding what a specific command term requires should be part of effective teaching.