PSYCHOLOGY TZ2

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

Overall grade boundaries

Higher level

Grade:	1	2	3	4	5	6	7
Mark range:	0 - 8	9 - 19	20 - 29	30 - 41	42 - 53	54 - 65	66 - 100

Standard level

Grade:	1	2	3	4	5	6	7
Mark range:	0 - 9	10 - 22	23 - 32	33 - 43	44 - 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 - 13	14 - 18	19 - 23	24 - 28	29 - 46

Standard level

Grade:	1	2	3	4	5	6	7
Mark range:	0 - 4	5 - 9	10 - 13	14 - 18	19 - 23	24 - 28	29 - 46

General comments

Candidates showed a wide breadth of knowledge throughout paper one. It was impressive to see many candidates showing that they understand a good mix of both classical studies and more modern research being done in psychology.

Many candidates had difficulty differentiating between the different levels of analysis. For example, candidates used biological studies for the cognitive level of analysis or vice versa. Studies used at each level of analysis should reflect the principles that define it.

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

Candidates often did not meet the demands of the command term. Candidates often did not make explicit links between the research and the question. In addition, candidates often had difficulties in defining terms, often using the term to define itself: for example, conformity is when someone conforms.

Critical thinking was often not linked to the question. Psychological terminology was used but not well explained.

In several section A responses, the structure was confusing and unfocused. Writing skills were often poor, with several candidates writing one long paragraph to answer the extended response (section B) questions.

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

Candidates often wrote very strong descriptions of research and theories. There were several strong responses to the conformity essay, demonstrating a strong breadth of knowledge and understanding. Time management also seemed better this session as candidates allocated

sufficient time for the short answer (section A) questions and extended response (section B) questions respectively.

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

Section A

Biological level of analysis

Many candidates did not outline the principle, but simply stated it. Many candidates used the principle that animals are used to understand human behaviour, but often candidates did not make a link between the study and human behaviour, and instead wrote "we cannot generalize from animals." In addition, several candidates used animal studies from behaviourist psychology, such as Pavlov or Thorndike, with no link to the biological level of analysis.

Cognitive level of analysis

This question had several strong responses. However, there were also several responses where no specific biological factor was identified. The question asked for the effect of one biological factor on cognition, but several candidates described Maguire's study of neuroplasticity which is a study of how cognition may affect physiology. Some responses also included trust or emotion as a cognitive process, which is also incorrect.

Sociocultural level of analysis

Several candidates did not know the difference between a stereotype and a schema – giving examples of Loftus and Palmer or Bartlett as examples of stereotyping. A significant number of candidates simply described studies of stereotyping or the formation of stereotyping, rather than clearly identifying and describing how stereotypes affect behaviour.

Section B

Biological level of analysis

Very few candidates chose this question. Many strong responses focused on emotions or how biological factors affect memory. In weaker responses, the role of either cognitive or biological factors were not explicitly addressed, or the interaction between the two was only minimally referenced.

Cognitive level of analysis

There was a broad range of responses to this question, with the most popular choices being classic memory models, Flashbulb Memory Theory and schema theory.

Several studies were very well described, but the focus of a significant number of responses was on evaluating research, rather than evaluating the theory itself.

Sociocultural level of analysis

There were many strong responses that addressed factors that influence conformity. Some candidates tried to do too much, alluding to many factors, but not discussing them effectively. Three factors well described and discussed was usually better argued than responses that attempted to discuss five or six factors. Candidates often merely described or identified factors with little emphasis on the actual discussion which resulted in low marks for critical thinking.

Recommendations and guidance for the teaching of future candidates

- Candidates need more training in how to evaluate theories. The vast majority of the responses to the cognitive question focused on evaluating the research. In addition, more attention needs to be paid to the actual details of the theory. Very few candidates were able to clearly explain schema theory and how it actually worked. This led to incorrect evaluation points – for example, that we are born with all the schema that we will ever have or that flashbulb memories are the result of a lot of practice.

- Psychological terminology needs to be more carefully explained. Many candidates have a very superficial understanding of terms like ecological validity, reliability and demand characteristics. When these terms are used, the link between the term and the study must be explicitly explained.

- When asked for a factor or an effect, the factor or effect must be explicitly identified. Also, when asked for one behaviour, candidates should explicitly identify one particular behaviour.

- Candidates should focus on the demands of the question, rather than including irrelevant information. Candidates should not automatically write out the principles for the level of analysis as a response to each question. In addition, in the short answer (section A) responses there should not be any evaluation. This is poor practice and often leads to an unfocused response. In addition, it may leave the candidate feeling that they have written a well-developed response when, in fact, the majority of the response is irrelevant and will not merit marks.

Further comments

As has been mentioned in several subject reports, teachers are discouraged from using old research. Several candidates used Berthold's study of roosters. This study was done in the 1840s and what we know about testosterone today is not in line with his original findings. Psychology is a dynamic and developing field of scientific inquiry. By teaching candidates old, outdated research, we do them a disservice. It is recommended that teachers focus on research from the 20th or 21st century.

Finally, even though all teachers have favourite studies that they like to teach, it is important for us to consider the hazards of teaching material outside of the curriculum. For example, in the current psychology curriculum, there is no learning outcome for obedience research. In spite of this, Milgram is commonly taught. Several candidates misused Milgram on the exam as an example of conformity research. It is advised that teachers stay within the framework of the curriculum in order to guarantee the highest levels of clarity with regard to the learning outcomes.

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.