November 2014 subject reports

PSYCHOLOGY

Overall grade boundaries

Higher Level

Grade:	1	2	3	4	5	6	7
Mark range:	0 - 9	10 - 19	20 - 30	31 - 42	43 - 55	56 - 67	68 - 100

Standard Level

Grade:	1	2	3	4	5	6	7
Mark range:	0 - 9	10 - 21	22 - 33	34 - 46	47 - 57	58 - 69	70 - 100

Higher level internal assessment

Component grade boundaries

Grade:	1	2	3	4	5	6	7
Mark range:	0 - 2	3 - 5	6 - 9	10 - 13	14 - 17	18 - 21	22 - 28

The range and suitability of the work submitted

Overall the range and suitability of the work submitted was of high standard, with the majority of centres showing good understanding the requirements of the internal assessment (IA). The majority of IA reports submitted met the criteria for an experimental design, whereby there was manipulation of an independent variable with the effects of a dependent variable recorded. Very few centres submitted non-experimental work, or reports which breached ethical guidelines.

Most IAs were based on studies from the cognitive level of analysis, such as levels of processing, effects of interference and memory recall, reconstructive memory, experiments related to schema theory, and imagery versus rehearsal.

In general, the weaker reports shared the following characteristics:

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

It should be noted that it is not required to make an exact replication of an experiment; a partial replication is adequate. However, the candidate's experiment should be closely linked to a published experiment.

Candidate performance against each criterion

Criterion A: introduction

In many reports, the research presented was generally well explained and adequate in number. Many candidates, however, still find it challenging to explicitly state how the research is highly relevant to the hypotheses. Candidates should also be aware that the background research should always logically lead towards the candidate's own research hypotheses which in turn allows for the formulation of a clear research hypothesis.

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

Criterion B: design

The design (repeated measures or independent groups) was not always properly justified. A number of candidates did not operationalize the IV and/or the DV, that is, clearly making them measurable.

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

Criterion C: participants

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

Most candidates did identify the appropriate sampling technique, but often struggled to explain the use of this method, that is, why the method was chosen.

Criterion D: procedure

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

Criterion E: results – descriptive

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

Criterion F: results – inferential

Most candidates chose an appropriate test and justified the use of the test (based on the level of data and the design). At times, t-tests were chosen (which is acceptable) but often it was not the most appropriate test based on the particular aspects of the experiment (sample and/or variance of data). The statement of significance should also always be appropriate and clearly stated.

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

Criterion G: discussion

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

Almost all candidates presented limitations, but often in a superficial manner, without rigorous analysis. Limitations should be presented that are relevant to this particular investigation, not limitations of experiments in general. Modifications suggested by candidates should be based on the observed limitations of their study. It is also necessary that a conclusion is included for all IAs.

Criterion H: citation of sources

Referencing continues to be an area of weakness. Candidates often did not cite certain research or use a standard citation method (such as APA).

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. This was often not included.

Recommendations and guidance for the teaching of future candidates

- Assist candidates in selecting an appropriate experiment to replicate with an appropriate theoretical framework and background research. Finding relatively simple experiments to replicate is recommended. Again, it is advised that candidates do a partial replication of studies rather than try to 'create' their own study. Only two variables should be manipulated.

- For the sample, the number of participants in the experiment does not need to exceed 20 (independent design) or 10 (repeated measures design), and it is recommended to observe this.

- It would be helpful if candidates were given published research to read in order to familiarize themselves with the aspects of experimental research.

- It is recommended that candidates be given guidance in accessing appropriate psychological journals. Many candidates only used internet sources of a non-specialist nature as background literature.

- Candidates and teachers should be fully aware of the assessment criteria to ensure that all guidelines are met.

Standard level internal assessment

Component grade boundaries

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

The range and suitability of the work submitted

Overall the range and suitability of the work submitted was of a good standard this session. The majority of candidates were well aware of ethical issues and all candidates with a full report included a copy of informed consent in the appendices. The majority of candidates also provided a copy of the debriefing letter.

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

In general, the weaker reports shared the following characteristics:

- Weak and imprecise explanation of original research in the introduction. The aim was not clearly stated.
- Results were not always clearly related to the aim of the study and descriptive statistics were absent or not justified. Also, the results were not stated in words, the use of descriptive statistics was not explained, or there was no table.
- Discussions were superficial with no discussion of own results in the light of original research. Identification of limitations of own procedure was not linked to suggestions for modification.

Candidate performance against each criterion

Criterion A: introduction

Most candidates seemed familiar with the requirements of each assessment criterion although there were differences in achievement levels. In the stronger reports, the introduction was very well written with a clear focus and well-explained original research that clearly led to the candidate's own aim. It seemed difficult for weaker candidates to explain the original research in sufficient detail so that there was a clear relationship between a particular research study and the candidate's own aim. The

introduction is important in that it presents the rationale for the candidate's own experiment. Therefore the original research study should be explained and analysed in sufficient depth (for example, aim, procedure, findings) to stimulate discussion of own results in the light of the original research in the discussion section. The experiment should be simple and therefore it is acceptable to carry out a partial replication of a research study; for example, by reducing the number of variables. The level of depth of the explanation of the original research study was at times shallow, partly because weaker candidates used insufficient background material such as revision guides.

Criterion B: design

Most candidates correctly identified their experimental design but the choice of the design wasn't clearly justified at times. Justifications should be presented by stating why a specific design is preferred as the "best choice" for the specific topic and study conducted. Weak candidates still have problems in understanding the difference between a method (experiment) and design (repeated measures, independent samples, matched pairs design).

Most reports identified the independent and dependent variable. However, operationalization of these two variables was not always precise. Also, some weaker candidates confused the IV and DV.

Ethical guidelines were usually clearly followed and evidence of this was provided within the report.

Criterion C: participants

Description of characteristics of participants usually included some, but not always adequate, information. All samples were based on a student population usually coming from the candidate's school. Most candidates correctly identified the sampling technique but sometimes failed to justify the chosen sampling method. Some candidates had quite small or large samples in their study – there is no need for a very large sample. A sample of 15–20 participants is an appropriate size for a standard level study.

Criterion D: procedure

Most candidates wrote the procedure clearly and with enough relevant detail. However, some candidates did not include all relevant materials which they used when conducting their study (standardized instructions, informed consent forms, lists of words, and debriefing notes were not always included in the appendices). In order to make the procedure easily replicable, candidates should include information about the timing and location of their tasks.

Criterion E: results

Some candidates found it difficult to write the results section. Therefore, the results sections varied considerably in quality of presentation and the clarity of the information. In several weaker reports, only raw data was included in the results section. At their best they were both clear and informative, but there were occasions when the results were omitted or gave so little information that they were hardly worth presenting. In some cases, candidates presented and explained findings that were not part of the original aim. Usually the descriptive statistics chosen were mean, median, mode, standard deviation and range. It is highly recommended for candidates to choose only one measure of central tendency (the one which is appropriate for the scale of measurement used when obtaining the results). In addition, an appropriate measure of dispersion should be included. The use of descriptive statistics should be justified or explained. Presentation of tables and graphs needs to be clear so that data can be readily understood by the reader. Graphs should have a title, legend, and a label for each axis.

Criterion F: discussion

The quality of the discussion ranged greatly: from superficially written comments with general evaluation of the study to well balanced with an appropriate conclusion, offering modifications and improvements for further research. Many candidates found it difficult to relate their findings back to the original work stated in the introduction, and were seldom clear and specific in criticizing their own research. Another common problem was that some relevant information was provided in the discussion section but it wasn't fully elaborated by providing a clear connection between the method applied in the study and the results obtained. Furthermore, related ideas for future modifications of research were often undeveloped.

In addition, some reports reflected that candidates had problems with presenting a relevant conclusion. Although there were some concluding remarks embedded within the report, a final concluding statement was rarely clearly written.

Criterion G: presentation

Presentation in many reports was in accordance with requirements of criterion G – a lot of reports were impressively presented and scored full marks for this criterion. Unfortunately, there were still reports in which no reference section was provided. In some cases, although standard citation methods were used there was a lack of in-text citations or the references were not fully stated. Another common mistake was failure to label appendices and reference them in the body of the report. Also, raw data were sometimes incorrectly presented in tables in the results section.

Recommendations and guidance for the teaching of future candidates

- It is much easier for candidates to replicate an experiment if they have access to readings about experiments or summaries of them so that they can read about the materials, procedure and other important details.

- There should be more focus on the relationship between the aim of the candidate's study and the original research study so that these can be integrated in the introduction and the discussion of results. The original research study should be analysed in sufficient depth in the introduction so that the discussion of the candidate's own research is clearly linked to the material provided in the introduction.

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

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

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

- Before writing the discussion section candidates should carefully examine the procedure conducted in their study, and analyse ethical considerations or sampling biases that might have had an impact on the results obtained.

- Candidates should be encouraged to place all of their raw data and relevant materials in the appendices.

- Use of proper background readings relevant for the original research study must be encouraged. It is recommended that candidates be trained in critical use of internet resources. Many candidates only used internet sources of a non-specialist nature as background literature.

Higher and standard level paper one

Component grade boundaries

Higher level							
Grade:	1	2	3	4	5	6	7
Mark range:	0 - 4	5 - 8	9 - 13	14 - 19	20 - 25	26 - 31	32 - 46

Standard level **Grade:**	1	2	3	4	5	6	7
Mark range:	0 - 4	5 - 8	9 - 13	14 - 19	20 - 25	26 - 31	32 - 46

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

Overall, candidates' responses tended to be descriptive and lacked a clear focus on the demands of the command term both in section A and section B. It appeared to be a distinct challenge for many candidates to ensure their responses were developed with an explicit view to addressing the requirements for the various command terms. As a result, candidates lost marks in section A questions as the focus was on the content at the expense of the actual command term. In section B responses, despite the inclusion of very relevant theory and research, essays were on the whole descriptive and often any evidence of critical thinking was added as an afterthought or was not well integrated into the essay. Although there were some well developed and evaluative essays with relevant empirical evidence, many candidates struggled to build a clear and considered argument and therefore lost marks in criterion B.

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

There was a wide variety of research studies provided by candidates that were often well described and highly relevant to the question. There seemed to be fewer candidates focusing on outdated examples such as Phineas Gage and there was often evidence of knowledge and understanding of more current research. A good number of essay responses demonstrated well organized and structured responses in which ideas were presented in a clear and focused manner.

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

Section A

Biological level of analysis

A large number of candidates had difficulty explaining the concept of genetic inheritance in sufficient detail and in many cases there was little focus on the command term beyond a brief description of differences between monozygotic and dizygotic twins. Only a small minority of candidates showed a clear understanding of genes as affecting the physiological mechanisms which may lead to an explicit behaviour and most responses simply focused on a very general idea of behaviours being "influenced by both genes and the environment". Most candidates addressed intelligence as an example of behaviour but the majority struggled to link it well to inheritance so these responses were usually superficial. Candidates who chose depression as an example of a behaviour were often able to address the issue of gene expression more effectively. A small minority of candidates did not choose a relevant behaviour and focused on examples such as Down's syndrome or autism which did not earn them any marks.

Cognitive level of analysis

Many candidates did not provide an outline of the principles and simply stated them before going directly into descriptions of relevant studies to demonstrate the principles, with the result that the demands of the question were not met. Candidates who chose the "scientific study of cognitive processes" principle often provided little more than a link to a description of a study which was a laboratory experiment but did not address the features of the study which were "scientific". Stronger responses more clearly addressed the idea of hypothesis testing or that unobservable cognitive processes like memory could be scientifically investigated. Some candidates got confused with the sociocultural level of analysis principle that social and cultural factors affect behaviour so did not address the cognitive level of analysis and lost marks. Also, many candidates who chose the principle that mental processes guide behaviour gave overly general and detailed descriptions of schema theory and studies that investigate schemas without a clear or developed link back to the principle.

Sociocultural level of analysis

Weaker responses to this question did not outline social learning theory sufficiently and the focus of the response was on long but often inaccurate or poorly explained descriptions of Bandura's study on aggression at the expense of the theory. The idea of vicarious reinforcement was rarely addressed in such cases. In many cases though, candidates outlined social learning theory competently and provided clear links from the study to the theory.

Section B

Biological level of analysis

This was a popular choice and the majority of candidates focused on laboratory experiments as the method of choice, with case studies being the second most popular. The biggest issue for most candidates, regardless of the chosen method, was that the use of the method at the

biological level of analysis was rarely well addressed. Most candidates were able to explain the features of the method and the advantages and limitations but this was done generally and the appropriateness of the method for biologically based research was not well justified or was not well developed. Candidates also tended to provide fairly repetitive and superficial evaluation which was more often linked to the studies included than the method itself. Overly general evaluation concerned with cause and effect relationships and lack of ecological validity often led to a loss of marks in criterion B. Candidates choosing the case study method tended to be overly descriptive of studies such as that of HM at the expense of the actual use of the method to address cases of brain damage.

Cognitive level of analysis

This was the least popular essay choice this session. This essay was often well answered and many candidates showed good knowledge and understanding of relevant theories and research and the interaction aspect was explained and developed sufficiently. In some cases, however, the supporting studies were very descriptive and the interaction aspect of the question was not well developed. Weaker essays sometimes focused on flashbulb memory studies which did not address the demands of the question.

Sociocultural level of analysis

This was the most popular essay choice this session and, though there were several strong and well-evaluated essays, the majority addressed evaluation fairly superficially with some candidates struggling to address strengths of research in the field. Evaluation, for the most part, tended to focus on the studies of Asch (1951) and Sherif (1935) and in a large number of cases the evaluation was redundant and repetitive with most of the emphasis on the use of deception, lack of ecological validity and issues with the sample rather than addressing the bigger picture of conformity. Several candidates included Zimbardo's (1971) Stanford Prison study but in the vast majority of cases did not address the issue of conformity to social roles effectively at all.

Recommendations and guidance for the teaching of future candidates

- Candidates are still losing marks in section A answers due to including unnecessary evaluation which leads to a loss of focus on the command term in short answer questions. Candidates should be helped to demonstrate their understanding of how their chosen study links clearly and directly to the demands of the question. The command terms should be revised carefully so that candidates are aware of how they should approach each question so that the demands are fully met.

- On the whole, candidates need to be advised on how to make the best use of the studies they include, especially in section B responses. Research often tends to be presented in a very generic or stock manner and is not used effectively in light of the demands of the question. Studies should be carefully linked to the questions in both section A and section B so that the candidate's choice of empirical evidence is well justified. Evaluation was also rather generic and candidates should be encouraged to

integrate it well into their argument rather than just tacking it on to the end of a study. Standard evaluation points such as issues with ecological validity need to be developed and justified and candidates should be advised to avoid repetitive evaluation for studies which use the same research method. Likewise, critical analysis and evaluation should be directly relevant to the requirements of the question and should focus on supporting or challenging arguments/theories and go beyond a simple evaluation of studies.

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

Standard level **Grade:**	1	2	3	4	5	6	7
Mark range:	0 - 2	3 - 5	6 - 8	9 - 11	12 - 13	14 - 16	17 - 22

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

No single option appeared to be more difficult than the others. As in previous examinations, candidates demonstrated better knowledge and understanding of topics than the ability to apply critical thinking skills to the material covered in the options.

Many candidates were able to provide only basic evaluation of studies by focusing on issues such as lack of generalizability, insufficient sample size, lack of ecological validity, and various ethical concerns thereby earning marks in the middle range for criterion B.

Questions four and seven proved difficult for some candidates who were unclear on the meaning of the command term “examine”. Understanding of this command term was relevant for providing focused evidence of critical thinking. Although both of these questions were rather popular, many responses didn't receive the highest marks since they only focused on how certain factors affect development (or obesity) but did not include any interrelationships.

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

In the most popular options – abnormal psychology and human relationships – candidates seemed to be well prepared in terms of content knowledge. The majority of candidates were able to demonstrate a basic understanding of major concepts within the options and to use appropriate studies and theories in support of their responses. The responses of a significant number of candidates revealed that they had developed an understanding that a comprehensive explanation of human psychology requires attention be paid to all three levels of analysis: biological, cognitive, and sociocultural.

As in previous examinations, it was clear that some candidates were well prepared in how to write a response.

Evidence of analysis and evaluation was clearly present in the best essays. Some responses were exceptionally well written – these responses reflected that knowledge of relevant material was clearly applied to the questions and the answer was well organized.

In general, most responses provided some reference to psychological research, although precise and focused knowledge of research was not always present.

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

Abnormal psychology

Question 1

This question was the most popular within the option and the whole exam. Many candidates provided good responses. Most candidates addressed the following cognitive factors:

- negative cognitive schemas influencing depression
- distorted weight-related schema influencing bulimia.

Sociocultural factors tended to be less often addressed – when this was the case the following factors were addressed:

- vulnerability models
- cross-cultural differences influencing body dissatisfaction.

Many candidates appropriately addressed other factors (including biological factors) in order to respond to the command term “to what extent”.

The majority of candidates provided a response by discussing the extent to which cognitive or sociocultural factors influence one specific disorder – depression and anorexia/bulimia were the most popular choices.

Candidates who provided stronger responses tended to consider a small number of cognitive or sociocultural factors in order to demonstrate depth of knowledge – this approach was successful for many candidates who had good knowledge of a few studies/theories and could relate this to the question.

Question 2

This question was the least popular within the option. Most responses addressing this question tended to be of below average or average quality.

Candidates addressing this question tended to discuss the following issues:

- vulnerability models/life stressors (*eg* Brown and Harris, 1978)
- bias in diagnosis (Caplan, 1995)
- gender norms (Brown and Harris, 1978)
- differences in cognitive styles (*eg* Nolen-Hoeksema, 1994).

Evidence of critical thinking tended to be provided by addressing:

- cultural considerations
- role of historical context
- methodological considerations
- empirical evidence.

Although there was some evidence of critical thinking it was not always clearly focused on the specific question.

Question 3

This question was a rather popular choice. Many candidates provided good, detailed responses that were focused on the question and nicely organized. In most cases a discussion included both strengths and limitations as required, but it usually wasn't evenly balanced – strengths were more thoroughly addressed than limitations.

Candidates usually chose to evaluate a combination of CBT (cognitive behavioural therapy) and drug therapy as a popular choice and linked their response to specific disorders – depression or anorexia.

Strengths most often provided were:

- the overall treatment is tailored to the specific needs of the client
- it provides flexibility in treatment
- lower relapse rates.

Limitations most often provided were:

- too complex for one clinician to manage
- difficult to empirically study its effectiveness

- using too many approaches may reduce the effectiveness because several therapists are involved and they are not communicating about their progress or goals.

Developmental psychology

Question 4

This question was the least popular within the option and usually candidates seemed not to have enough content knowledge to provide a full essay response. The command term "examine" tended to be a challenge for some candidates since they failed to consider how biological factors influence human development in a way that uncovered the interrelationships of this issue. Most candidates decided to examine how only one biological factor influences human development.

Candidates usually addressed the question by referring to:

- the effects of maturation of the nervous system and cognitive development
- the role of neuroplasticity in brain development
- the role of sex hormones.

In some rare cases candidates discussed how cognitive and sociocultural factors interact with a biological factor. This approach is useful and appropriate but usually this was only done in a simple, superficial manner.

Question 5

This was the most popular question within the option. Many candidates provided good, detailed responses that were focused on the question and nicely organized.

Candidates choosing this question addressed the following theories:

- Piaget's theory of cognitive development
- Vygotsky's theory of cognitive development.

Evidence of critical thinking was provided by addressing:

- methodological, cultural and gender considerations
- controversies related to stages versus continuous process
- productivity of the theory in generating psychological research
- applicability of the theory.

Question 6

This was the second most popular question within the option. Most often candidates chose the following strategies:

- social programmes for youth such as Head Start or the Big Brothers Big Sisters programme
- stress inoculation training.

Most candidates provided good, detailed knowledge of two strategies but had more problems discussing them. When discussion was well addressed it tended to be given in the following way:

- by addressing that a child's ability to build resilience is dependent on their age and stage of development
- by discussing that a person's culture might have an impact on how he or she communicates feelings and deals with adversity.

Health psychology

Question 7

This was not a very popular question. The command term "examine" tended to be quite challenging for a number of less-prepared candidates. Most candidates gave an outline of several different factors related to overeating and the development of obesity. The material was usually presented in a simplistic manner.

Responses tended to focus on sociocultural factors – most often sedentary lifestyle and a high-fat diet was addressed. When physiological factors were addressed, genetic predisposition and the role of dopamine was presented. Psychological/cognitive factors most often addressed were self-esteem, and distorted body image.

In most cases candidates addressed a larger number of factors related to overeating and the development of obesity in order to demonstrate breadth of knowledge.

Question 8

This question was the most popular choice within the option. Many candidates provided good answers to this question. Most candidates chose to discuss adrenal responses to environmental stressors, the role of cortisol on hippocampal cell loss and the connection between stress and the immune system. Often candidates provided reference to Cannon's fight or flight theory (1914) and Selye's general adaptation syndrome model (1956) in their response. The most often stated research study was Kiecolt-Glaser *et al*.'s (1984) natural experiment to investigate whether the stress of an important exam had an effect on the body's immune functioning.

Well-prepared candidates tended to consider how psychological or social aspects of stress are interrelated to physiological aspects of stress.

Question 9

This question tended to attract well prepared candidates. Most candidates chose to evaluate nicotine replacement therapy, MBSR (mindfulness-based stress reduction) or group therapies.

Often, candidates who considered a small number of treatments for substance abuse and/or addictive behaviour managed to demonstrate depth of knowledge, and showed explicit evidence of critical thinking.

Psychology of human relationships

Question 10

This question was the most popular choice within the option and one of the most frequently addressed questions within this exam. Most candidates who wrote a response to this question had good knowledge of several relevant factors influencing bystanderism. Candidates often included Latané and Darley's (1968) research on the role of the number of people available to help (diffusion of responsibility) as well as referring to informational social influence (pluralistic ignorance). Also, Piliavin's costs versus benefits of helping (Piliavin *et al.*, 1969) was often addressed. Candidates provided good evidence of critical thinking by addressing cultural considerations and the role of historical context. Many candidates included some review of methodological considerations of research studies but often this needed further development.

Most candidates discussed a larger number of factors influencing bystanderism in order to demonstrate breadth of knowledge.

Question 11

This question was the least popular within the option. The command term "explain" requires candidates to give a detailed account, including reasons or causes, of why culture plays an important role in the formation and maintenance of relationships. Often candidates failed to address the command term clearly and instead provided a simple overview of several examples that show cultural differences in the formation and/or maintenance of relationships. For example, many candidates described how individualistic cultures focus on individual choice and romantic love whereas collectivist cultures often emphasize arranged marriages. When referring to studies, candidates most often provided some knowledge of Yelsma and Athappilly's (1988) comparative study of Indian arranged marriages and American love marriages and Levine *et al*.'s (1995) study on the role of love in the establishment of marriage. Evidence of critical thinking was usually only implicitly provided by offering simplistic evaluation of relevant research.

Question 12

This question was a rather popular choice within the option. Candidates often evaluated Wedekind's sweaty t-shirt study. Theories that were most often addressed were the

evolutionary explanation (Buss, 1996), social exchange theory (Kelley and Thibaut, 1959) and proximity theory. Evaluation tended to be explicit and included methodological, cultural and gender considerations of the research. Candidates that evaluated a small number of studies and/or theories in order to demonstrate depth of knowledge usually provided good detailed knowledge and therefore received quite a lot of marks in criterion A.

Sport psychology

Question 13

Not enough responses read to provide feedback.

Question 14

This question was the most popular among candidates responding to the sport psychology option. Candidates had knowledge of theories such as the inverted-U hypothesis, and Hanin's optimal arousal theory, as well as studies such as Fazey and Hardy's 1988 study on cognitive arousal and choking.

The question was well answered in terms of knowledge and understanding, but less so in terms of critical thinking. Some candidates did not focus entirely upon the question and lost focus during the response.

Question 15

Not enough responses read to provide feedback.

Recommendations and guidance for the teaching of future candidates

- Candidates need practice in writing high quality responses. Many candidates seem to have a lot of knowledge of key terms and explanations relevant for psychology but fail to apply their knowledge to the specified question. After writing their responses candidates should check whether they have addressed all aspects of the question, and most importantly, whether they have explicitly addressed the requirements of the command term.

- It could be beneficial to ask candidates to go through the exercise of marking their own or someone else's response. This exercise can help candidates understand how the examiner marks an essay.

- Candidates should be reminded that when the question includes a specific number (usually one or two), discussing additional information does not add marks to their response since only the first one/two will be considered. An exception to this rule is when candidates are using additional information to clearly and explicitly evaluate the one/two things that are the focus of the question. In this case, candidates are showing clear evidence of critical thinking.

Higher level paper three

Component grade boundaries

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

General comments

There were some very good responses indicating that some candidates are very well prepared. Overall it appeared that candidates understood the stimulus material quite well and were able to use it to some effect in their responses; however, there was a tendency in weaker responses to rely on anecdotal evidence or opinions instead of knowledge of qualitative research methodology.

In weaker responses candidates tended to analyse the study instead of using the stimulus material to address the methodological aspects of the study as was required in the three questions. The topic of health care for poor people raised various moral concerns and comparisons with healthcare systems in the countries of the candidates, which was not really relevant to the question. Candidates' opinions on the matter sometimes interfered with their use of knowledge of qualitative research methodology. There was also a tendency to focus mainly on the poor people rather than the settings and the treatment the poor people were exposed to.

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

In general, weaker candidates tended to comment on the study in the stimulus material instead of addressing the questions related to methodology in qualitative research. Weaker candidates also had problems integrating the stimulus material into their response in a meaningful way. For example, they relied on citations from the stimulus material to answer the questions or they treated the stimulus material as a text that they had to analyse rather than using it as the account of a research study that they should use as documentation for their methodological and ethical comments related to the questions asked. Weaker candidates also seemed to have problems understanding what was required when the command term "evaluation" was used as in question one. Many candidates in the lower markbands did not seem to know that they should evaluate the method used by the researchers in the study (that is, make an appraisal by weighing up strengths and limitations). Instead they either offered their opinion on the study or pointed only at strengths of using the particular method.

In question one quite a few candidates focused only on the covert aspect of the study and forgot to evaluate the part related to participant observation. This approach meant that they scored rather low as part of the question was not addressed.

Question three seemed overall the most difficult to answer for candidates. First of all the command term "to what extent" appeared difficult to address effectively for many candidates. Weaker responses just explained what generalization is with reference to populations and had very limited, if any, knowledge of generalization in qualitative research. Many candidates ended up concluding that the results of the study could not be generalized at all for reasons such as the small sample, that only poor people participated, that it was wintertime or that the sample had not been randomly selected.

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

Many candidates used the stimulus material well and could integrate their knowledge of qualitative research methodology with relevant parts of the stimulus material and stronger candidates did an excellent job here. Most candidates demonstrated a good knowledge of the covert aspect of the study in the stimulus material as well as the methodological and ethical issues related to this. Stronger responses also demonstrated sound knowledge of participant observation and combined covert and participant observation well.

Question two with reference to triangulation was by far the area where candidates seemed the best prepared. This question had the strongest responses overall although a number of weak responses were also seen. Stronger candidates seemed to have a good understanding of the role of triangulation in qualitative research and offered a balanced view of how triangulation was used in the study (researcher and data triangulation) and some also offered alternative ways to apply triangulation and gave good reasons for why this could be difficult in this particular study. As for question one it was clear that all candidates knew what “covert” meant and even the weaker responses addressed this aspect but this was often the only thing that was properly addressed in the weakest responses.

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

Question 1

This question was often well answered but there were also very weak responses. Stronger candidates were able to address both the covert aspect of the study as well as the participant observation aspect. Such responses were also able to account for relevant strengths and limitations of using this method in the study. Strong candidates were able to provide good reasons for why the researcher had to use a covert approach as well as why she had chosen participant observation in this case study. Many gave very good reasons (for example, that “walking in the shoes of the poor” was the only way to have a genuine insight into how these health centres treated the poor clients seeking medical help, or that going under cover was necessary as the staff would not behave in the same way if they knew they were being observed). Most

candidates also pointed towards ethical issues in the covert aspect of the observation or explained the potential danger for the researcher of doing such a study.

Weaker candidates had a tendency to forget to address the aspect of "participant observation" of the study. Or they did not evaluate properly – that is, presenting both strengths and limitations of the method used in the study.

Question 2

Stronger responses provided a good explanation of how triangulation was applied in the study using relevant material from the stimulus material, for example referring to line 16 where it is stated that the researcher asked another researcher to check the field notes and her interpretation of the data, and explained this with reference to potential researcher bias in this sensitive topic. Candidates often suggested other approaches to triangulation and gave good reasons for this with reference to ensuring credibility/trustworthiness in qualitative research. However, there was a tendency in some responses to focus too much on credibility and reflexivity.

Weaker responses did not know what triangulation is or they did not identify how triangulation was applied in the study and merely went into (irrelevant) details of the study instead of focusing on triangulation. Such responses seemed to know very little about triangulation as a contribution to the credibility of the study.

Question 3

Stronger responses had a good grasp of the specific conditions for generalization in qualitative research as well as an understanding of how this differs from generalization in quantitative research. Many of the stronger responses were able to provide specific terminology for generalization in qualitative research, that is, inferential generalization, representational generalization, or theoretical generalization, and they could apply this knowledge to the stimulus material.

Weaker responses demonstrated no clear knowledge of generalization in qualitative research but all had some knowledge of statistical generalization. This knowledge seemed to override any knowledge they may have had of generalization in qualitative research. The weakest responses focused on factors from the study such as time of year, sex of the researcher, or quality of welfare systems and claimed that because of such factors the results could not be generalized to the whole world thus demonstrating very limited knowledge and understanding of generalization in qualitative research.

Recommendations and guidance for the teaching of future candidates

- The paper three examination is based on a short description of a qualitative research study (the stimulus material) accompanied by three questions related to the methodology used in that particular research study. Candidates must answer all three questions paying attention to the command term and using their knowledge of

qualitative research as well as information from the stimulus material to support their analysis. Paper three is based on short answer questions and each question can receive a maximum of ten marks. 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, as this will not receive any credit.

- It appeared that some candidates had difficulties using the stimulus material properly. Quite a few discussed the content of the stimulus material without much reference to relevant knowledge of qualitative research methods, or wrote about qualitative research methods in a generic way without much reference to the stimulus material. Some generic knowledge is necessary in answering the questions but if candidates merely describe what they know within a given area and forget to integrate this knowledge with the study in the stimulus material they will be awarded marks at the lower end of the mark range. Good preparation for the exam involves using past exam papers for training so that candidates have the opportunity to acquire an understanding of 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. Likewise, teaching paper three could 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 question papers so that they become familiar with the requirements of this paper.

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

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