THE UNITED REPUBLIC OF TANZANIA
MINISTRY OF EDUCATION, SCIENCE AND TECHNOLOGY
NATIONAL EXAMINATIONS COUNCIL OF TANZANIA

CANDIDATES' ITEM RESPONSE ANALYSIS REPORT ON THE DIPLOMA IN SECONDARY EDUCATION EXAMINATION (DSEE) 2021

EDUCATIONAL RESEARCH, MEASUREMENT AND EVALUATION

THE UNITED REPUBLIC OF TANZANIA

MINISTRY OF EDUCATION, SCIENCE AND TECHNOLOGY

NATIONAL EXAMINATIONS COUNCIL OF TANZANIA

CANDIDATES' ITEM RESPONSE ANALYSIS REPORT ON THE DIPLOMA IN SECONDARY EDUCATION EXAMINATION (DSEE) 2021

762 EDUCATIONAL RESEARCH, MEASUREMENT AND EVALUATION

Published by:

The National Examinations Council of Tanzania,

P.O. Box 2624,

Dar es Salaam, Tanzania.

TABLE OF CONTENTS

FOREWORD

The National Examination Council of Tanzania is pleased to issue this report on Candidates' Item Response Analysis on the Diploma in Secondary Education Examination (DSEE) in Educational Research, Measurement and Evaluation subject for the year 2021. The report provides feedback to student teachers, tutors, parents, policy makers and the public in general on the performance of the candidates and the extent to which the instructional goals and objectives were met.

The Diploma in Secondary Education Examination marks the end of the diploma in education. It is a summative evaluation which shows the effectiveness of the education system in general and education delivery system in particular. The report indicates what the education system was able or unable to offer to students during their study on Diploma in Secondary Education.

In this report, factors which led the candidates to answer the questions correctly or incorrectly have been analysed. The analysis shows that, the candidates with good performance understood the demands of questions; had basic knowledge of the subject matter and good mastery of the English Language and possessed essay writing skills. However, the candidates who performance poorly demonstrated insufficient knowledge especially in the three topics which were; Educational Assessment and Evaluation, Educational Research and Assessing Achievement.

The feedback from this report is expected to help education administrators, college principals, tutors and student teachers to identify proper measures for improving candidates' performance in future examinations administered by the Council.

Finally, the Council is quite grateful to all stakeholders who provided valuable assistance in preparing this report.

Dr. Charles E. Msonde
EXECUTIVE SECRETARY

1.0 INTRODUCTION

This report presents the performance of the candidates who sat for the Diploma in Secondary Education Examination (DSEE) in Educational Research, Measurement and Evaluation subject in 2021. The examination tested the candidates' competences in explaining scales of educational measurement, developing research skills, carrying out projects and action research as well as disseminating the findings to others and using assessment skills and tools for improving the teaching and learning process.

A total of 2,095 candidates sat for the examination. The general performance of candidates was good since 97.97 per cent passed. The candidates' performance in 2021 with different grades as compared to that of the year 2020 is summarized in the table 1.

Year	**Sat**	**Number of Candidates and Percentage**					
			Grades				
		Passed	**A**	**B**	**C**	**D**	**F**
2020	2,810	2,781	92	680	1559	450	10
		99.64%	3.3%	24.2%	55.5%	16%	0.36%
2021	2,095	2,029	2	215	1,291	521	42
		97.97%	0.09%	10.4%	62.3%	25.1%	2%

The analysis of data in the Table 1 depicts that, the general performance has dropped by 1.67 per cent when compared to that of 2020. Moreover, the number of candidates who passed with grade A and B has decreased by 3.2 per cent and 9.8 per cent respectively.

In this report, the detailed analysis was done on the performance of the candidates in each question and topics based on the total number of candidates who sat for examination.

The examination paper consisted of two sections, A and B, with sixteen (16) questions in total. Section A had ten (10) questions set from the topics; Assessing Achievement, Test Construction, Educational Assessment and Evaluation, and Educational Research. All questions in this section were short answer questions and were all compulsory. Each question carried four (4) marks making a total of 40 marks for the section. Section B had six (6) questions set from the topics of Analysis and Interpretation of the Test Results, Educational Measurement, Educational Research, Test

Construction, and Qualities of Test. The questions in this section were of essay type and the candidates were required to attempt four (4) questions, where Question 11 was compulsory. Each question carried 15 marks, making a total of 60 marks for the section.

In this report, the analysis of the question is based on the category of the short answer items in Section A and essay type items in Section B. For Section A, the performance of the candidate is regarded as *Weak* if the scores range from 0 to 1.5 marks, *Average* if the scores range from 2 to 2.5 marks and *Good* if the scores range from 3 to 4 marks. For Section B, which contains essay questions the performance of the candidate is regarded as *Weak* if the scores range from 0 to 5.5 marks, *Average* if the scores range from 6 to 10 marks, and *Good* if the scores range from 10.5 to 15 marks. Also, general performance of the candidates is regarded as *Weak* if the scores range from 0 to 39%, *Average* if the scores range from 40 to 69%, and *Good* if the scores range from 70 to 100% paper wise.

The samples of the candidates' answers in each question have been attached to illustrate their responses. Also colours have been used to indicate the performance of the candidates in each questions and topics whereby *green* indicates good performance, *yellow* average performance, and *red* poor performance.

2.0 ANALYSIS OF THE CANDIDATES' PERFORMANCE IN EACH QUESTION

2.1 Section A: Short Answers Questions

This section had ten (10) questions and the candidates were required to attempt all the questions. Each question carried four (04) marks making a total of forty (40) marks on this section.

2.1.1 Question 1: Test Construction

The question had two parts. In part (a) the candidates were required to describe extended response items as used in educational measurement and evaluation, while part (b) required the candidates to describe restricted response items as used in educational measurement and evaluation.

The question was attempted by all 2,095 candidates corresponding to 100 per cent. Generally, the performance in the question was good as 1,796 (85.7%) candidates scored from 2 to 4 marks. Figure 1 illustrates candidates' performance in this question.

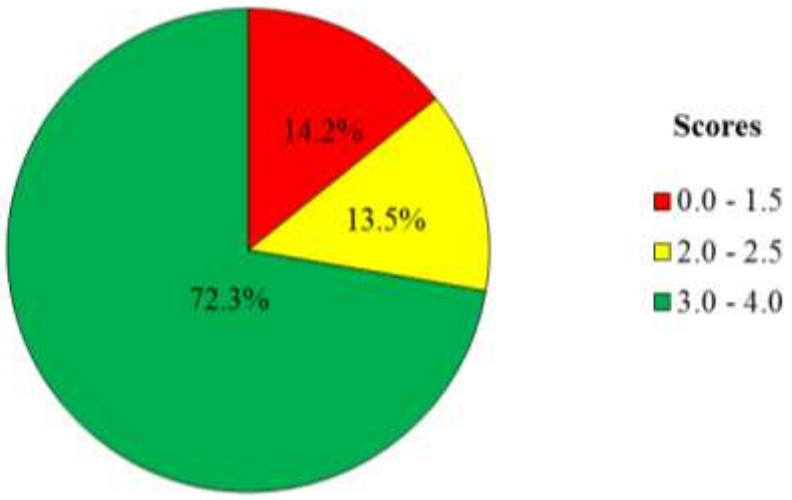
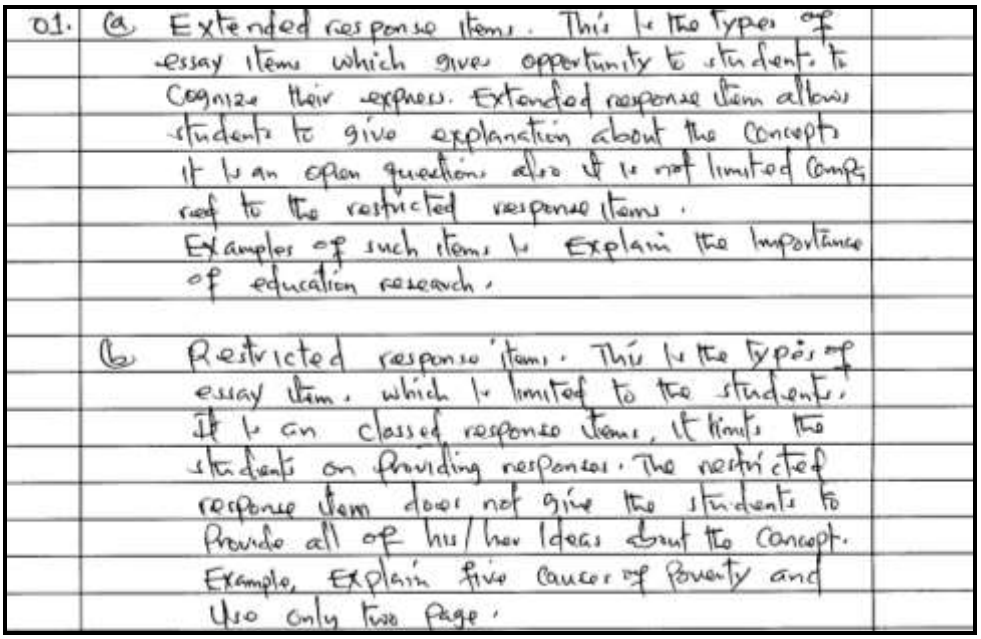


Figure 1: *The Candidates' Performance in Question 1*

Figure 1 shows that 1,513 (72.2%) candidates scored 3 to 4 marks among them 1,160 (55.4%) candidates scored 4 marks. Furthermore, 283(13.5%) candidates scored 2 to 2.5 marks. On the other hand, 298 (14.2%) candidates scored from 0 to 1.5 marks, of which, 196 (9.4%) scored 0 marks.

The analysis shows that 1513 (72.2 %) candidates who scored 3 to 4 marks correctly described the two concepts of extended and restricted response items as used in educational measurements and evaluation. In part (a), they correctly described the extended response items as*: (a) type of test item which give a candidate room to provide wider explanation or information of a concept, facts or principle, etc; extended responses help to give students freedom in making their responses; is the type of subjective item in which tends to provide room for an individual to express their views without restriction, for example when responding an essay without point limit.* Other candidates wrote: *(a) it is a type of subjective items which give students chance to respond directly without restrictions.*

Moreover, in part (b), they correctly defined the restricted response item as: *the type of test item which limits candidate's freedom to provide the intended information towards a particular concept or event; it is the type of item that limits an individual's ability to explain the concept where the examinee is given the limit for example, points, content and sometime time limit.* Therefore, these responses demonstrate that the candidates had sufficient knowledge on restricted and extended types of response. Extract 1.1 shows a sample of best responses from one of the candidates.

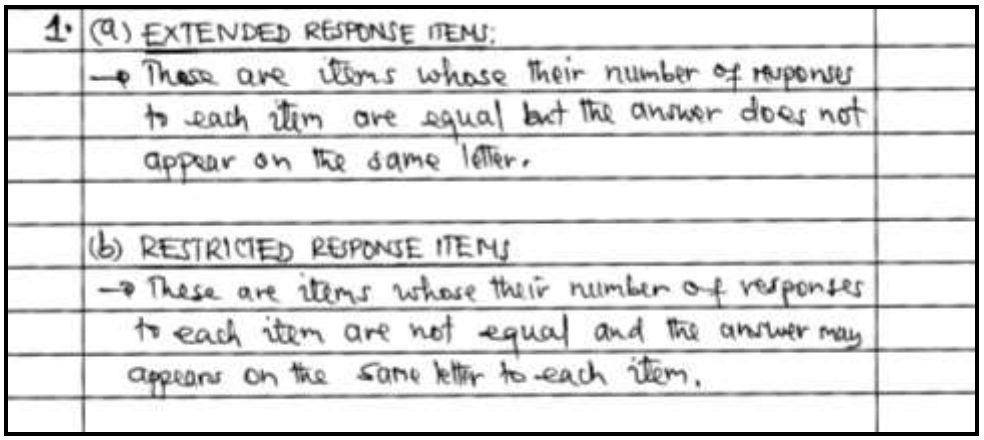

01. (a) Extended response items. This is the types of essay items which give opportunity to students to Cognize their express. Extended response item allows students to give explanation about the concepts it is an open questions also it is not limited compa red to the restricted response items.
Examples of such items is Explain the importance of education research.

(b) Restricted response items. This is the types of essay item, which is limited to the students. It is an classed response items, it limits the students on providing responses. The restricted response item does not give the students to Provide all of his/her ideas about the concept. Example, Explain five causes of poverty and Use only two page.

Extract 1.1: A sample of the correct responses in Question 1.

Further analysis of candidates' responses reveals that 283 (13.5%) candidates who scored 2 to 2.5 marks partially described the two concepts (extended and restricted). They described only one part of the two concepts. For example, one candidate gave responses such as: *(a) extended response items give freedom in selection of responses where a learner is able to explain many points. While in part (b) restricted response item is the one that limit or restrict in provision of response where there is no freedom.* Therefore, these descriptions indicate the partial knowledge on the concepts. However, many candidates in this category provided correct descriptions in one part of the items. For example, one candidate wrote: *(a) extended response items are the types of subjective items in which the test or question is prepared without concerning the length of the test; it is unstructured or opened essay. (b) Restricted response items are the item in which concerned with the length of the test or question is prepared according to length of test also level of the learner, it is structured or closed.* Part (a) had relevant answers while part (b) had irrelevant answers. Looking at these candidate's responses, it is obvious that, the candidate had insufficient knowledge and skills on extended and restricted response items as used in educational measurement and evaluation.

Furthermore, the analysis in this question revealed that 298 (14.2%) candidates who scored 0 to 1.5 marks failed to describe two concepts. Many of them wrote incorrect descriptions such as: *(a) extended response items are items that show the correct answers (b) restricted response items are items that do not show the correct answers.* Other candidates wrote that: *(a) extended response is the test items which give a chance of choice during the whole process of answering. (b) Restricted response item is the type of item which does not give chance of choice.* Additionally, another candidate described incorrect responses as: *(a) extended response is the type of subjective test items which tends to measure many learning outcomes it may cover two or more topics or all content in the subject matter, (b) restricted response is the type of subjective test items which tends to measure few language outcomes where as it covers a small area may be a topic or sub topics*. These descriptions show that candidates did not understood the question as they linked with other educational measurement and evaluation concepts of the subject such as objective and subjective type of test. Extract 1.2 is a sample of weak performance from one of the candidates.



Extract 1.2: A sample of response by a candidate who failed to describe extended and restricted response items in Question 1.

2.1.2 Question 2: Test Construction

This question consisted of two parts (a) and (b); each carried two (2) marks. Part (a) of the question asked the candidates to define taxonomy of educational objectives as used in educational measurement and evaluation, and part (b) candidates were asked to explain three domains of instructional objectives as proposed by Benjamin Bloom.

A total of 2,095 (100%) candidates attempted the question. The overall performance in the question was good as 1,540 (78.3%) candidates scored from 2 to 4 marks. Figure 2 is illustrative.

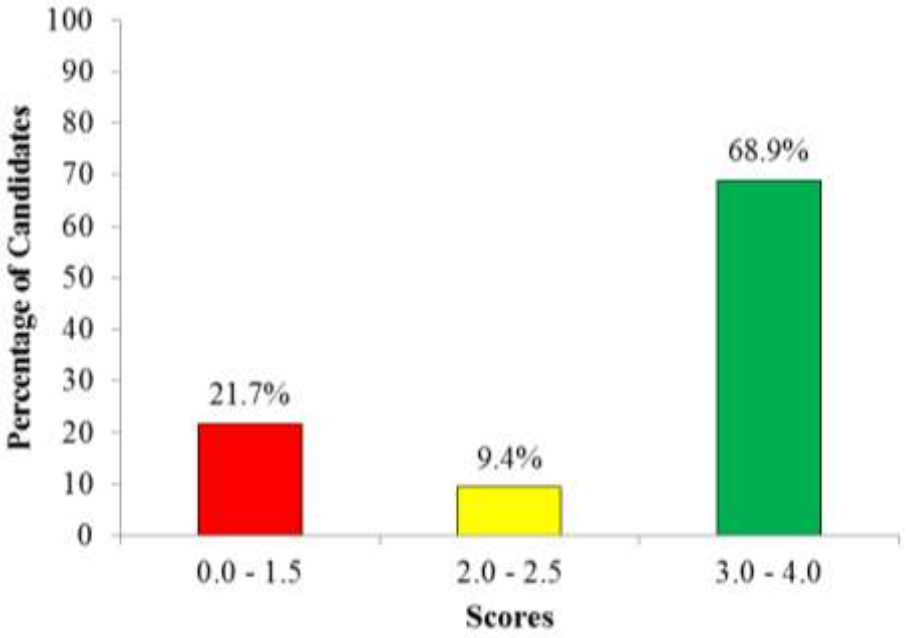
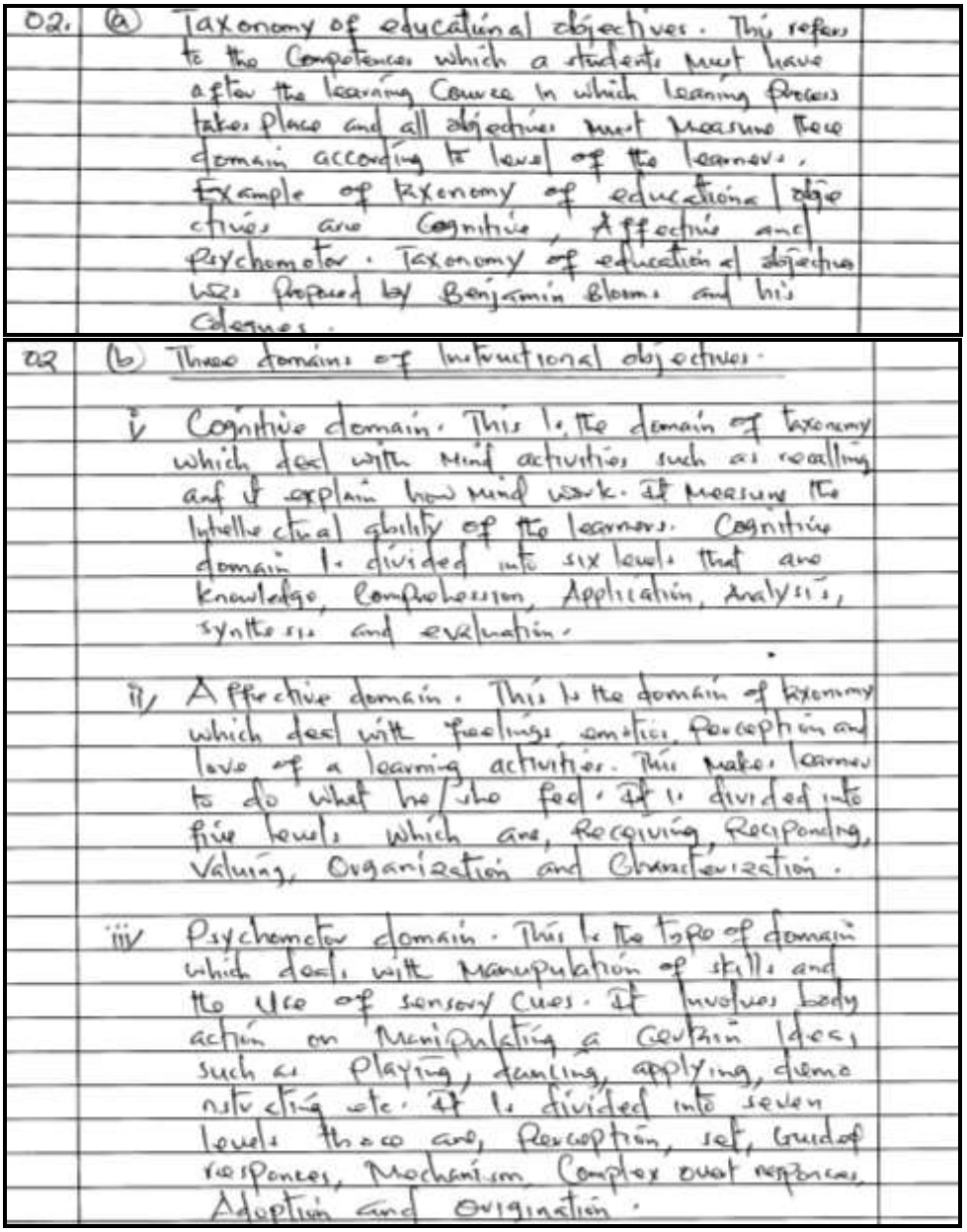


Figure 2: *The Candidates' Performance on Question 2*

The statistical data in Figure 2 shows that 1,444 (69%) candidates scored 3 to 4 marks, among them, 1,099 (52.5%) candidates scored full marks. 196 (9.3%) candidates scored 2 to 2.5 marks and 455 (21.7%) candidates scored 0 to 1.5 marks.

The analysis shows that 1,444 (69%) candidates who scored 3 to 4 marks provided correct definition of taxonomy of educational objectives as used in educational measurement and evaluation in part (a). For example, they defined it as: *the criteria of classifying objectives into three major domains.* Other candidates wrote that: *taxonomy of educational objectives is the system by which a teacher classifies objectives of learning domains.* Others wrote: *taxonomy of educational objectives is the structure/series of learning objectives that facilitates teaching and learning according to levels of the learners.* In part (b), the candidates explained three domains of instructional objectives as: *(i) cognitive domain is concerned with intellectual abilities of the learner, (ii) affective domain is the domain that is concerned with feelings or emotion of the learner, (iii) psychomotor domain is the domain which is concerned with perceptual motor skills.* Others wrote that: *(i) cognitive domain is the domain which deals with different levels of understanding e.g. knowledge, applying, analysing,...etc (ii) Affective domain is the domain that help to teach students about*

attitudes which can be positive or negative (iii) Psychomotor domain is the domain that deals with teaching skills and ability to acquire an individual. The analysis of this question shows that candidates had sufficient knowledge on taxonomy of educational objectives and domains of instructional objectives proposed by Benjamin Bloom. Extract 2.1 shows a sample of correct responses from one of the candidates.


02. (a) Taxonomy of educational objectives. This refers to the Competences which a students must have after the learning Course in which learning process takes place and all objectives must measure these domain according to level of the learners. Example of taxonomy of educational objectives are Cognitive, Affective and Psychomotor. Taxonomy of educational objectives was Proposed by Benjamin Bloom and his Colegues.

02 (b) Three domains of Instructional objectives.

i) Cognitive domain. This is the domain of taxonomy which deal with mind activities such as recalling and of explain how mind work. It measure the Intellectual ability of the learners. Cognitive domain is divided into six levels that are knowledge, Comprehension, Application, Analysis, synthesis and evaluation.

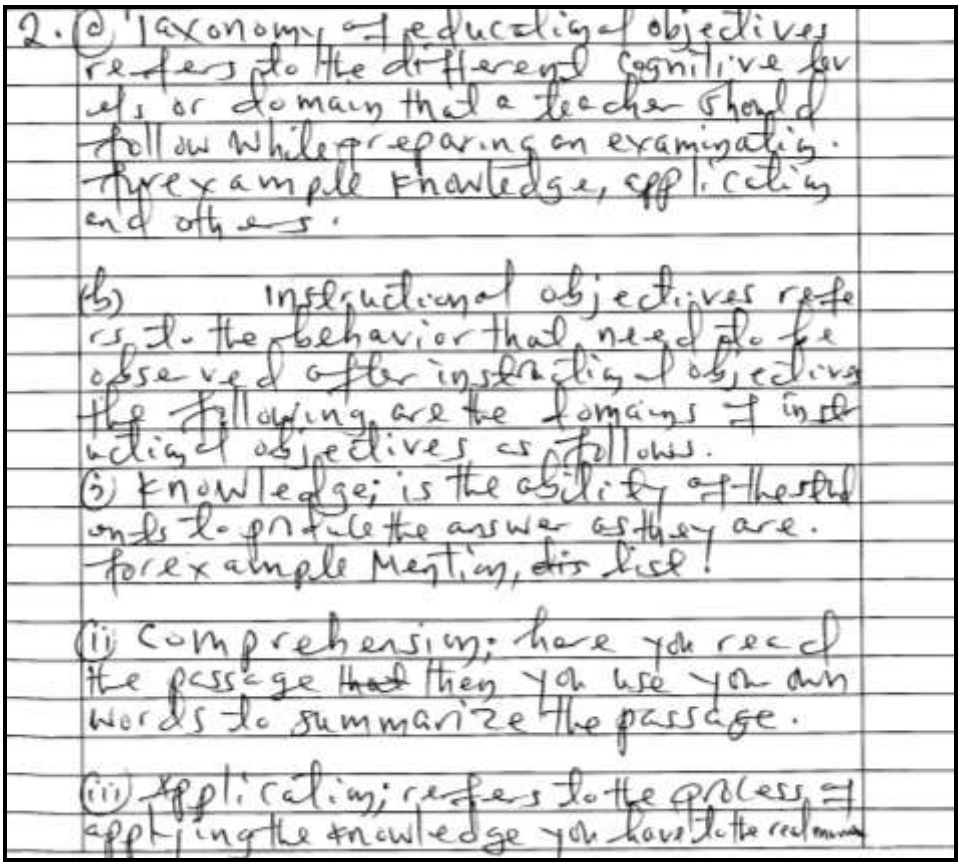
ii) Affective domain. This is the domain of taxonomy which deal with feelings, emotion, Perception and love of a learning activities. This make learner to do what he/she feel. It is divided into five levels which are, Receiving, Responding, Valuing, Organization and Characterization.

iii) Psychomotor domain. This is the type of domain which deals with Manupulation of skills and the use of sensory Cues. It involves body action on Manipulating a Certain Ideas, such as playing, dancing, applying, demonstrating etc. It is divided into seven levels these are, Perception, set, Guided responses, Mechanism, Complex overt responses, Adoption and Origination.

Extract 2.1: A sample of correct response in Question 2.

Moreover, the data analysis shows that 196 (9.3%) candidates who scored average marks (2 - 2.5) had insufficient knowledge on the meanings of taxonomy of educational objectives and domains of instructional objectives proposed by Benjamin Bloom. In part (a), some of them provided partial definitions such as: *taxonomy of educational objectives are the levels which were arranged according to domains of learning from simple to complex; taxonomy of educational objectives are the rank of domain of instructions used to measure learner' achievement in learning process* but they were unable to explain correctly the three domains of instructional objectives as they wrote: *(i) Knowledge- this domain measure the ability to recall what have been learnt in the previous course, (ii) Application-is the domain which measure the ability of the learner to relate concepts, (iii) Evaluation-is the domain of instructional objectives aims to measure the ability of learner to make value judgement on different issues.* The response shows that these candidates failed to differentiate domains of learning and levels as per each domain of instructional objectives and this lowered their points in this question.

Further analysis shows that candidates who scored low marks (0 - 1.5) either wrote incorrect responses or skipped the question though it was compulsory. For example some of the candidates wrote incorrect responses such as*: (a) taxonomy of educational objectives refers to the process of attaining or measuring domain of knowledge and its developed by some experts like Bloom which tends to show stages like comprehension, knowledge, application, analysis and evaluation (b) (i) Knowledge- this instructional objective tends to measure the knowledge attained by the learner and how to use it, (ii) Application-ability of the students to apply the knowledge and skills acquired, (iii) Comprehension-also it indicates how learners do express themselves in responding the question.* In addition, other candidates mentioned some of the cognitive levels only without explaining them. Looking at these responses, it proves that candidates were not aware about taxonomy of educational objectives and three domains of instructional objectives proposed by Benjamin Bloom. Extract 2.2 indicates poor performance from one of the candidates in this question.


2. (a) Taxonomy of educational objectives refers to the different cognitive levels or domain that a teacher should follow while preparing an examination. For example knowledge, applicating and others.

(b) Instructional objectives refers to the behavior that need to be observed after instructional objectives the following are the domains of instructional objectives as follows.

(i) knowledge; is the ability of the students to provide the answer as they are. for example Meeting, this list!

(ii) Comprehensing; have you read the passage then you use your own words to summarize the passage.

(iii) Applicating; refers to the process of applying the knowledge you have to the real

Extract 2.2: A sample of response by a candidate who failed to answer Question 2.

2.1.3 Question 3: Test Construction

The question required the candidates to identify four physical environmental factors that can affect an individual's performance. This question tested candidates' understanding on physical environmental factors that affect performance.

The question was attempted by 2,095 candidates equivalent to 100 per cent. The general performance in this question was weak as illustrated in Figure 3.

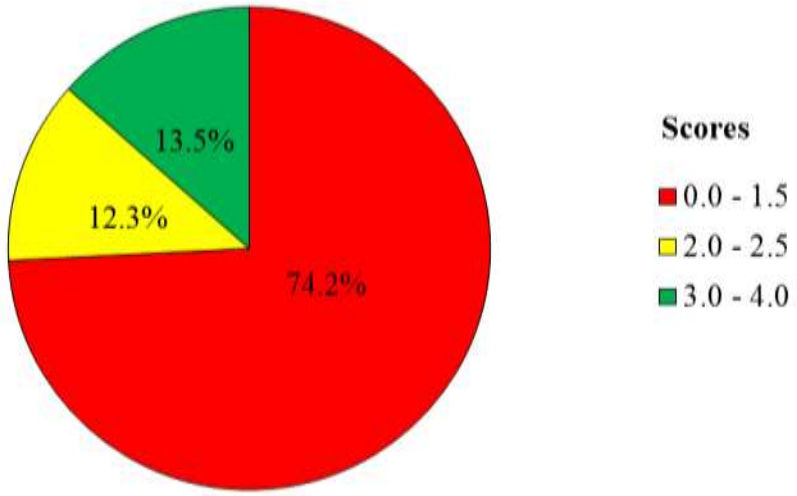
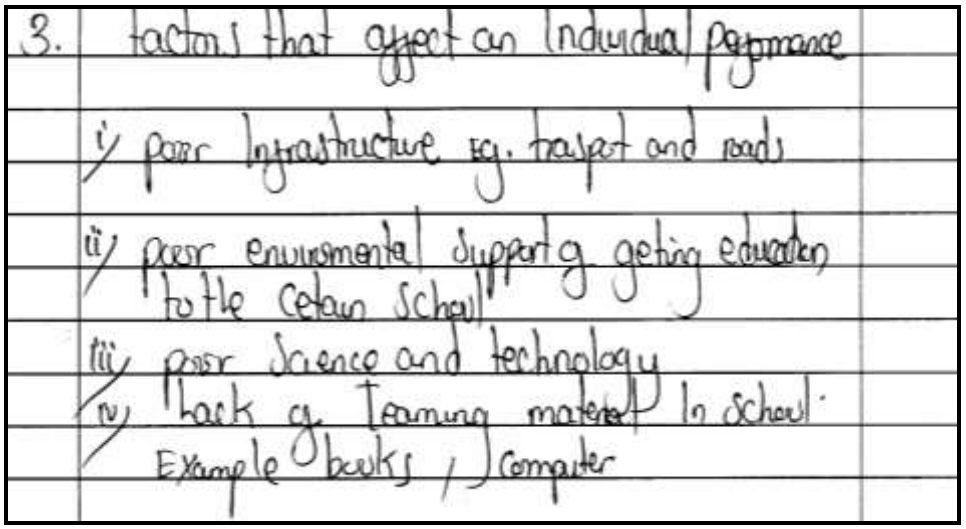


Figure 3: *The Candidates' Performance on Question 3.*

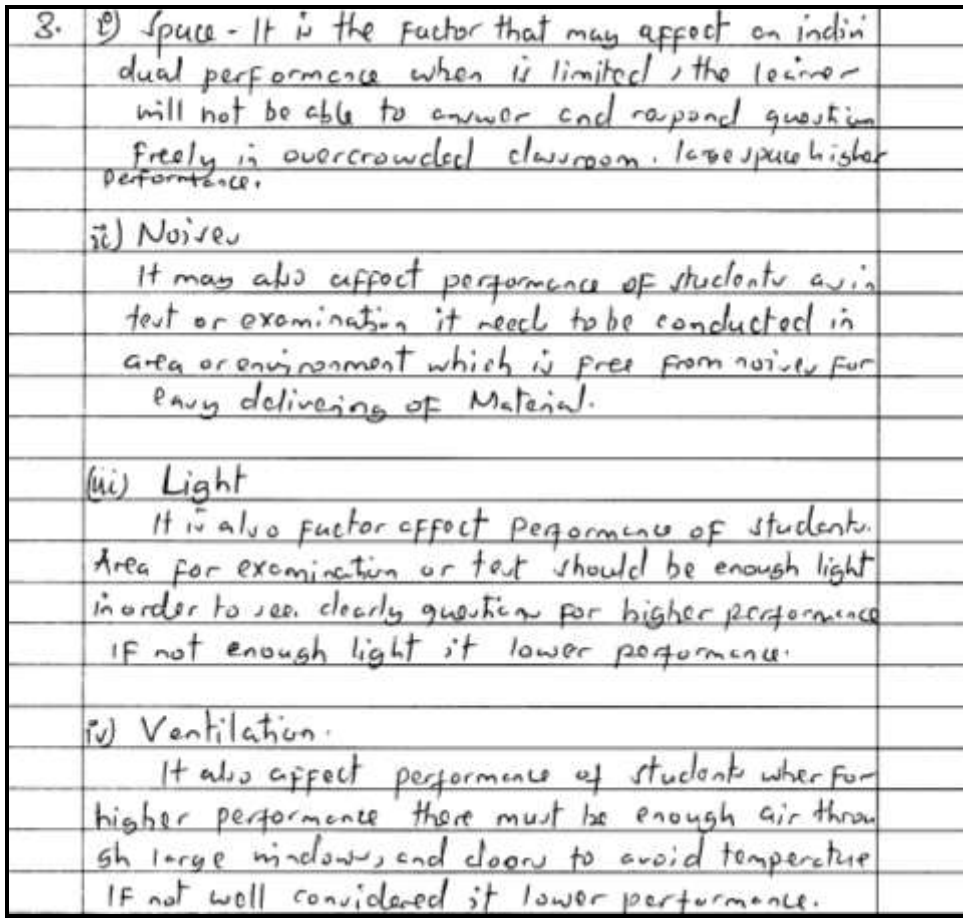
Results in Figure 3 indicate that 1,554 (74.2%) candidates scored 0 to 1.5 marks, among them 933 (44.5%) candidates scored 0 mark. 283 (13.5%) candidates scored 3 to 4 marks among them 133 (6.3%) candidates scored 4 marks, and 258 (12.3%) candidates scored 2 to 2.5 marks.

This question's analysis shows that 74.2 per cent of the candidates with weak performance were not careful in identifying physical environmental and some candidates' responses were irrelevant to the demands of the question. For example, some of the candidates identified the factors as: *(i) Lack of adequate learning materials, (ii) Lack of teachers, (iii) Poor laboratories, (iv) Un-friendly learning environment.* Others wrote: *(i) Environmental factors, (ii) Nature of examination, (iii) Too much announcement from supervisor, (iv) Nature of students themselves i.e having stress and psychological problems.* Extract 3.3 presents a sample of wrong responses from one of candidates in this question.



Extract 3.1: A sample of response by a candidate who failed to identify physical environmental factors that can affect individual's performance in Question 3.

Further analysis indicates that 283 (13.5%) candidates with good scores (3 – 4) identified four physical environmental factors that can affect an individual's performance correctly as follows: *(i) Physical noises from nearby classes and offices for instance, presence of garages and welding offices from which noises are inevitable due to nature of their works may affect individual's performance,(ii) Suddenly changes of weather can affect individual's performance when learners experience high or low temperature, (iii) If the room is small compared to number of examinees may also affect their performance, (iv) Insufficient air caused by poor ventilation may lower student' academic qualities.* Others wrote: *(i) If the examination room lacks appropriate light to help visibility of texts, diagrams, pictures, and objects may lead to failure in reading and writing. (ii)Improper sitting arrangement during test administration, (iii) Too much instructions from test invigilators out of those written on question if test was not well moderated and full of typing errors.* Others defined action research as*: the type of an applied research which conducted with the aim of finding solution to the problem immediately.* The responses imply that the candidates had sufficient knowledge and skills required by the question. Extract 3.2 shows good responses from one of the candidates.


3. (i) Space - It is the factor that may affect on individual performance when is limited, the learner will not be able to answer and respond question freely in overcrowded classroom. Large space higher performance.

(ii) Noises
It may also affect performance of students during test or examination it need to be conducted in area or environment which is free from noises for easy delivering of Material.

(iii) Light
It is also factor effect performance of students. Area for examination or test should be enough light in order to see clearly question for higher performance IF not enough light it lower performance.

(iv) Ventilation.
It also affect performance of students wher for higher performance there must be enough air through large windows and doors to avoid temperature IF not well convidered it lower performance.

Extract 3.2: A sample of response from candidates who correctly identified the four physical environmental factors that can affect performance correctly in Question 3.

However, the candidates' performance analysis reveals that candidates (12.3%) who had an average performance either failed to identify correctly all four physical environmental factors that affect individual's performance or listed the factors without descriptions. These candidates were also identified two out of four factors required. For example, they identified factors such as: *(i) Adequate working place, (ii) Ventilation, (iii) Temperature, and (iv) Light.* Others wrote*: (i) High or low temperature, (ii) High or low ventilation, (iii) Diseases, (iv) Poor infrastructure.* Thus, some of candidates' responses were correct and the other were incorrect suggesting that they lacked sufficient knowledge and skills on physical environmental factors that affect individual performance.

2.1.4 Question 4: Assessing Achievement

This question required candidates to explain the given concepts as used in assessing student's achievement in education. The six concepts were (a) Rating scales (b) Checklist (c) Socio-metric (d) Attitude test (e) Guess who techniques and (f) Anecdotal record.

A total of 2,095 (100 %) candidates attempted this question The general candidates' performance was weak since only 225 (10.7%) candidates scored 2 to 4 marks as summarised in Figure 4.

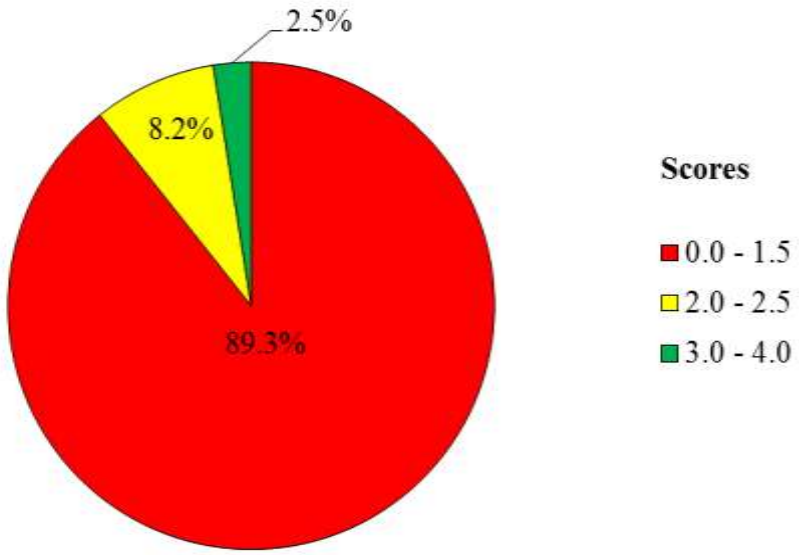
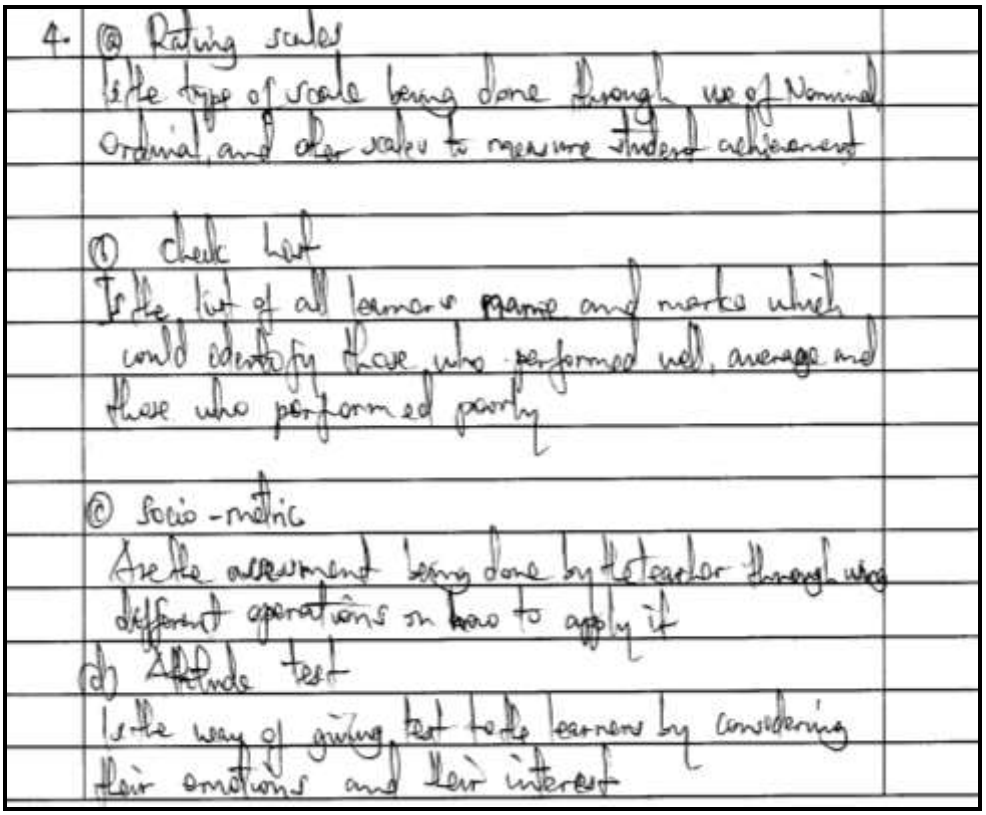


Fig. 4: *The Candidates' Performance on Question 4*

Figure 4 shows that among the candidates who attempted the question; 1,870 (89.3%) candidates scored 0 to 1.5 marks, 172 (8.1%) candidates scored 2 to 2.5 marks, and 53 (2.5%) candidates scored 3 to 4 marks.

The data analysis shows that 89.3 per cent of candidates who scored low marks (0 -1.5) failed to define all the two concepts while others defined correctly only one concept. For instance, in defining the attitude test, they wrote, *refers to the classroom test by which the previous learnt content are measured*, checklist *refers to the list that shows the performance of students from the chronological order of their position*. On the rating scale they wrote, *it is a scale which contains all scales and have true zero*. These responses were incorrect. This might be due to inadequate knowledge of

the concepts provided. Extract 4.1 illustrates the sample from a candidate with poor responses in this question.



Extract 4.1: A sample of response by a candidate who wrongly defined the concepts in Question 4.

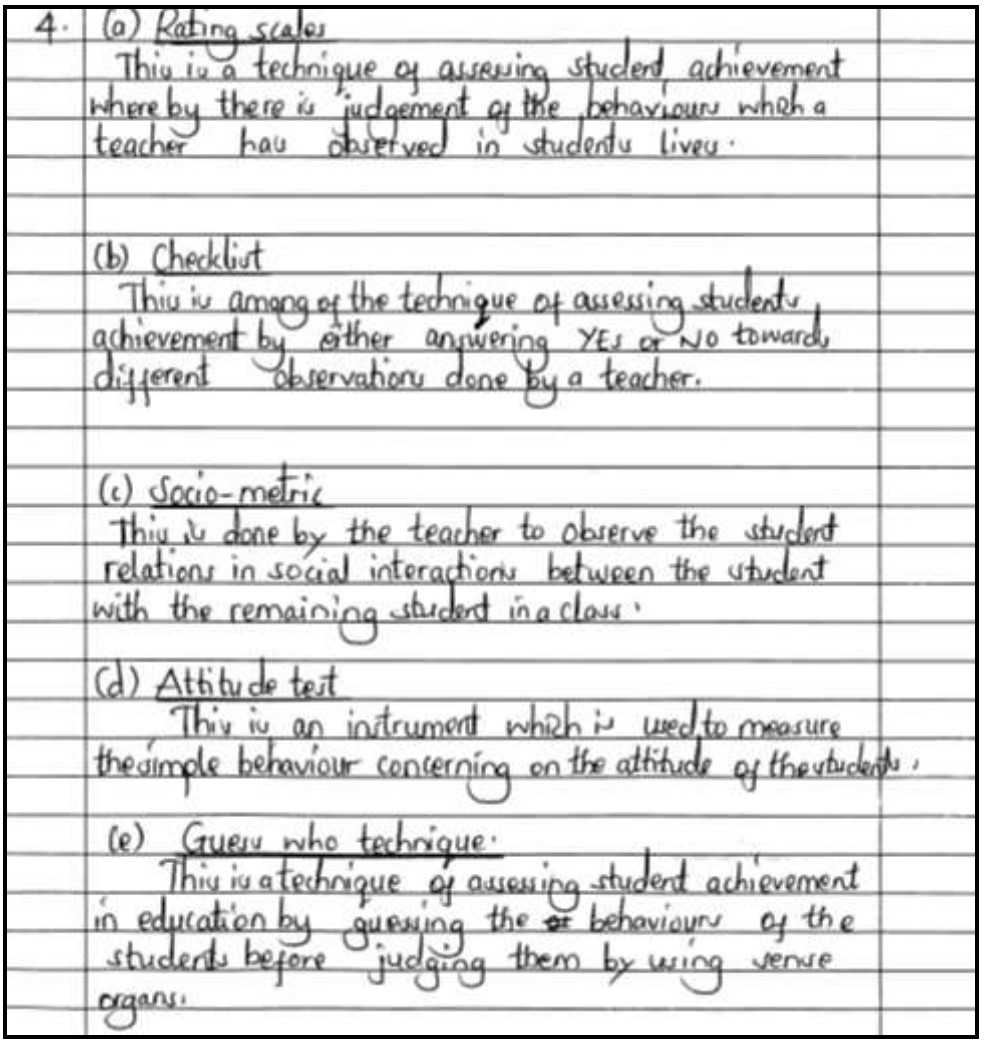
Furthermore, the analysis shows that candidates (8.1%) with average scores (2 - 2.5 marks) had partial knowledge on the six given concepts as used in assessing students' achievement in education. They explained the concepts as: *rating scale is the observation techniques where by observer check the rate the occurrence of attribute* and on checklist *is an observational technique where by the observer check the presence or absence of a certain attribute.* On the other hand, some of these candidates provided partial explanation of responses while other points were incorrect.

Additionally, candidate's responses indicate that 2.5 per cent of the candidates who scored good marks (3 - 4) had adequate knowledge on the concepts provided. The candidates were able to describe the concepts as follows: (a) *rating scales is a systematic procedure of report observable*

judgement eg. average, excellent; (b) *checklist refers to the techniques used in assessment that involve the recording whether the characteristic is present or not, it involves the use of yes or no answers;*(f) Anecdotal record that, *it deals with recording the observation of the learner in natural settings;* (d) *attitude test is the type of test item which used to measure the attitude of the learner in a particular academic area/program*. Others provided the definition on the other concepts as: (e) Guess who techniques *are the technique in assessment which describes the action is conducted by somebody /a particular person*. (c) *socio-metric is the technique is the technique in assessment which looks the interaction among students in the classroom.*

Moreover, other candidates explained the terms differently as: anecdotal record *is the observed judgement which kept a student in a positive character,* while the rating scales *is the scale where by the performance of the student is in a grade form. Example A-Excellent, B-very good, C-good. D- Certificatory, F- failed* or *is the techniques in assessment that involves the systematic process of recording observer's judgement on a particular behaviour of the learner.* They treated check list *as the techniques used in assessment that involve the recording whether the characteristic is present or not. It involves the use of yes or no answers,* and on the socio-metric they wrote that *is the technique in assessment which looks the interaction among students in the classroom.*

All these responses justify that candidates understood the demands of the question and had adequate knowledge and skills on the concepts asked as used in assessing students' achievement in education. Therefore, they provided relevant responses. Extract 4.2 shows a sample of good responses from a candidate who correctly explained four among six concepts, as they are used in assessing students' achievement in education.


4. (a) Rating scales
This is a technique of assessing student achievement where by there is judgement of the behaviours which a teacher has observed in students lives.

(b) Checklist
This is among of the technique of assessing students achievement by either answering YES or NO towards different observations done by a teacher.

(c) Socio-metric
This is done by the teacher to observe the student relations in social interactions between the student with the remaining student in a class.

(d) Attitude test
This is an instrument which is used to measure the simple behaviour concerning on the attitude of the students.

(e) Guess who technique.
This is a technique of assessing student achievement in education by guessing the ~~or~~ behaviours of the students before judging them by using sense organs.

Extract 4.2: A sample of response by a candidate who correctly explained the concepts in Question 4.

2.1.5 Question 5: Educational Research

This question required candidates to examine four characteristics of an action research.

The question was attempted by 2,095 candidates equivalent to 100 per cent. The performance was generally weak as 415 (19.8%) candidates scored 2 to 4 marks as shown in Figure 5.

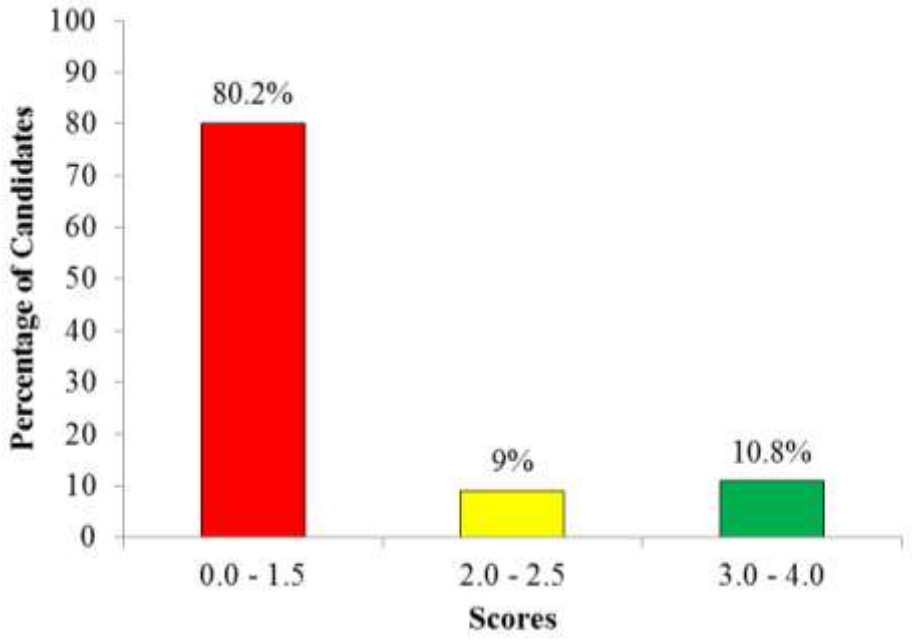
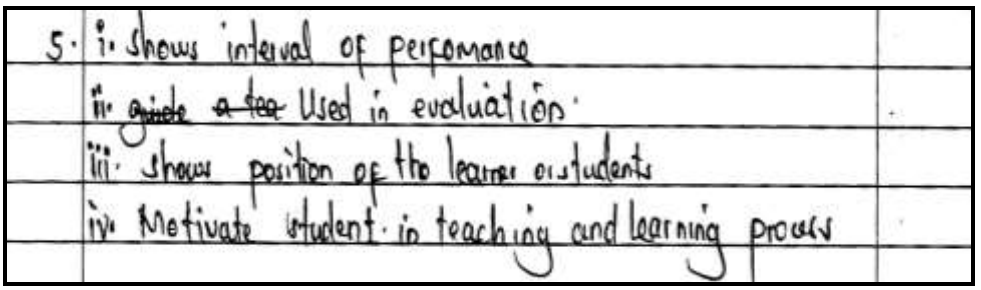


Figure 5: *Candidates' Performance on Question 5*

Figure 5 shows that 1680 (80.2%) candidates scored 0 to 1.5 marks, 227 (10.8%) candidates scored 3 to 4 marks and 188 (9%) candidates scored 2 to 2.5 marks.

The analysis reveals that, candidates who had weak performance (0 - 1.5 marks) provided incorrect characteristics of an action research like: *(i) it should be meaningful, specific, researchable, and verifiable; (ii) it should usually include field work; (iii) speed processing of data; (iv) it applies both quantitative and qualitative methods of collecting data.* Extract 5.1 shows such incorrect responses from one of the candidates.

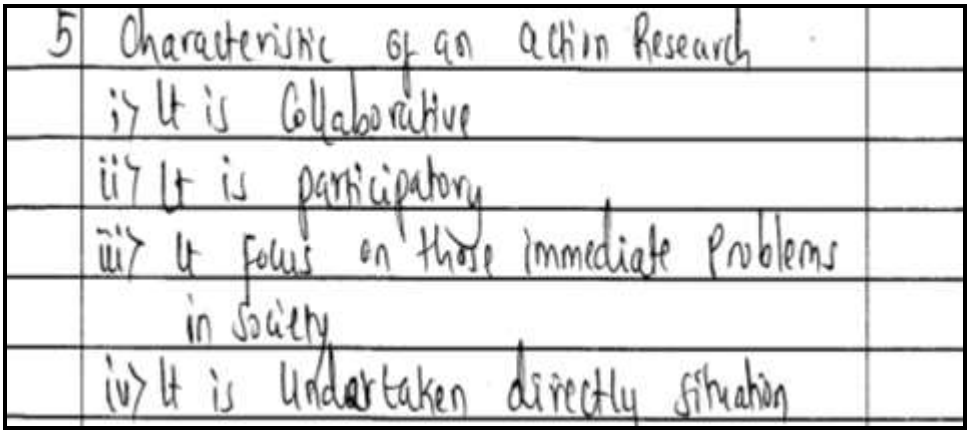

5. i. Shows interval of performance
ii. ~~a tea~~ Used in evaluation.
iii. Shows position of the learner o. students
iv. Motivate student in teaching and learning process

Extract 5.1: A sample of response by a candidate who failed to examine the characteristics of an action research in Question 5.

Furthermore, the analysis indicates that the candidates (9%) with average performance (2 - 2.5 marks) had inadequate knowledge on characteristics

of an action research. They examined relevant characteristics of an action research though lacked adequate clarifications as: *(i) it involves interaction between the researcher and the sample study; (ii) it is conducted more in institutions i.e colleges, schools, in solving existing problem; (iii) it aims at improving particular skills; (iv) its objectives are clearly defined.* they therefore provided partially correct points.

On the other hand, the candidates (10.8%) who scored good marks (3 - 4) provided relevant characteristics of an action research and presented in an organized manner. The candidates correctly gave the characteristics as: (i) *it is collaborative; (ii) it is undertaken directly in situation; (iii) it solved the problem which immediately occurs; (iv) it is participatory in nature.* Other candidates added the characteristics as: *it is a reflective process*, which is right. Extract 5.2 presents a correct response from a candidate who scored full marks.


5 Characteristic of an action Research
i) It is Collaborative
ii) It is participatory
iii) It focus on those immediate Problems in Society
iv) It is Undertaken directly situation

Extract 5.2: A sample of response by a candidate who correctly examined the characteristics of an action research in Question 5.

2.1.6 Question 6: Educational Research

This question had four concepts (a) Longitudinal study (b) Cross-sectional study (c) Type I error and (d) Type II error. The question required the candidates to give their understanding on the four concepts used in educational research.

A total of 2,095 candidates corresponding 100 per cent attempted the question. Generally, the performance of this question was weak, as only 34

(1.6%) candidates scored from 2 to 4 marks. Figure 6 illustrates the candidate's performance.

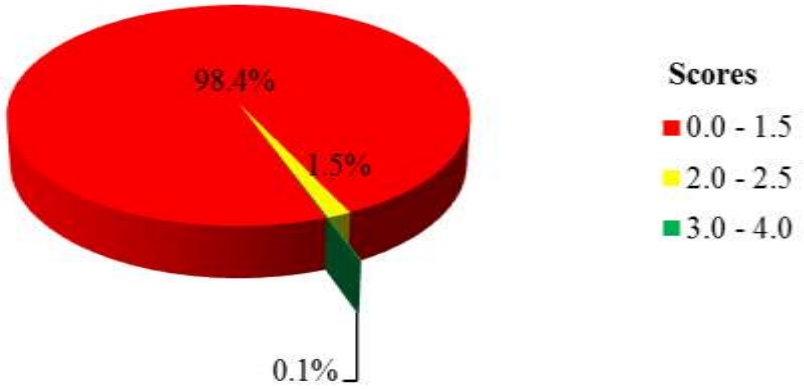
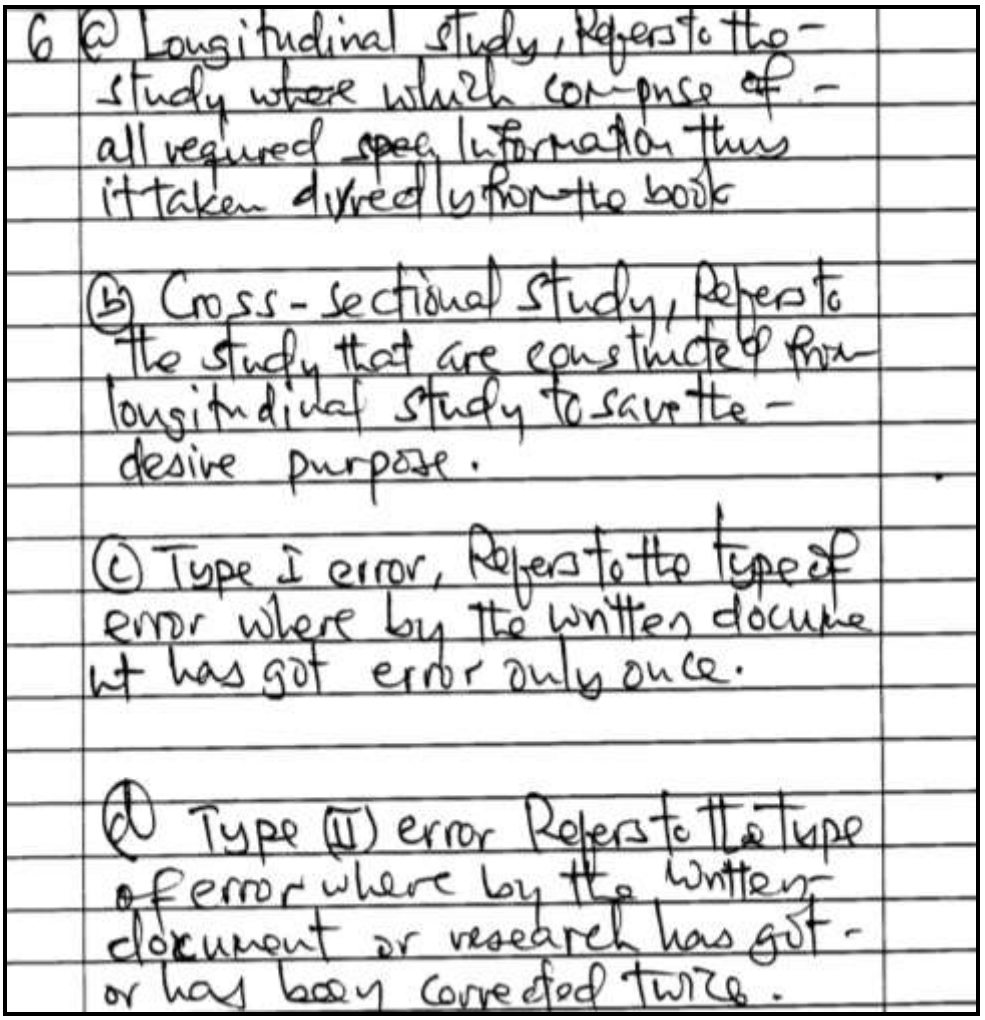


Figure 6: *The Candidates' Performance on Question 6*

Figure 6 shows that among the candidates who attempted the question; 2061 (98.4%) candidates scored 0 to 1.5 marks, 31 (1.5%) candidates scored 2 to 2.5 marks and 3 (0.1%) candidates scored 3 to 4 marks.

The analysis shows that 98.4 per cent of the candidates who had low performance lacked knowledge in defining the four concepts used in educational research. They gave wrong definition such as: *(a) longitudinal error: is the mistakes that occur before the action done in educational research; (d) type II error: - is the mistake that occur after the action done; (b) cross-section: - is the study which acts as a certain course of study; (c) type i error: - is the kind of error which occurs easy during the typing or writing words and also in arranging things.* These irrelevant definitions of the concepts from the candidates prove that, they were unaware of the concepts given. Extract 6.1 shows a sample of a candidate who performed poorly in the question.


6 (a) Longitudinal study, Refers to the study where which comprise of all required spec information thus it taken directly from the book

(b) Cross-sectional study, Refers to the study that are constructed from longitudinal study to save the desive purpose.

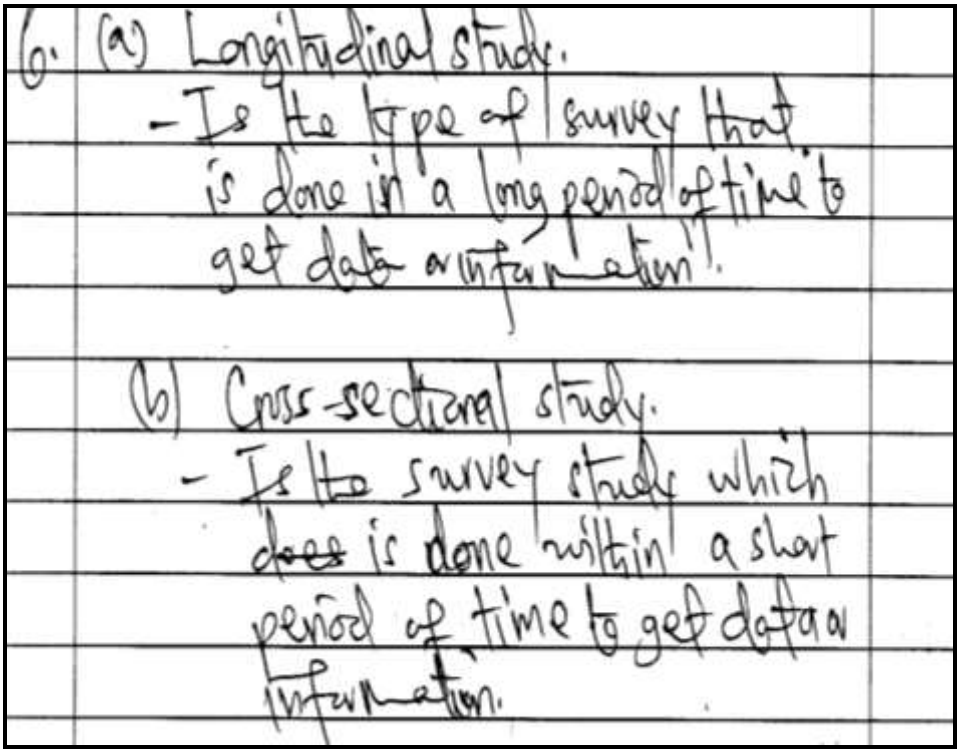
(c) Type I error, Refers to the type of error where by the written document has got error only once.

(d) Type (II) error Refers to the type of error where by the written document or research has got or has been corrected twice.

Extract 6.1: A sample of response by a candidate who failed to define four concepts as used in educational research in Question 6.

Moreover, the analysis indicates that candidates who had average performance defined correctly two concepts and missed the other two concepts. For example one of the candidates defined the given concepts as: *(a) longitudinal study:- is the survey research type where the research is going to be conducted; (b) cross-sectional study:- is the survey research design by which information from the sampled group is gathered in the population; (c) type I error:- is the error which occur first before conducting the research; (d) type II error:- is the error which occur as a results of failure of type I error.* Looking at the definitions above, it can be noted that this candidate was partially correct in (a) and (b) while had incorrect response in (c) and (d).

Further analysis shows that candidates (1.5%) who scored higher marks (3 – 4) had sufficient knowledge of the four concepts as follows: *(a) longitudinal study:- is the type of research design in which information is collected from a long period of time for the purpose of the entire population; (b) cross-sectional study:- refers to the research design in which information is gathered within a short period of time to be used in a population; (c) type I error:- this occur when the null hypothesis is rejected while it is actually true; and (d) Type II error:- this is used in hypothesis in which a researcher should accept the null hypothesis but the findings proved true.* Others defined the concepts as: (a) *longitudinal study:- is the kind of study which take long time to complete for example in making a research on diseases that rise in the society like corona viruses (covid-19); (b) cross-sectional study:- is the short way of study which help researcher to complete his/her course*. Extract 6.2 presents a sample of good responses by one of the candidates in question 6.


6. (a) Longitudinal study.
- Is the type of survey that is done in a long period of time to get data or information.

(b) Cross-sectional study.
- Is the survey study which ~~does~~ is done within a short period of time to get data or information.

Extract 6.2: A sample of response by a candidate who correctly defined two concepts in Question 6.

2.1.7 Question 7: Educational Assessment and Evaluation

This question required the candidates to examine four qualities of a good evaluation in teaching and learning process.

The question was attempted by 2,095 (100%) candidates. The candidates' performance was weak as 39.2 per cent scored 2 to 4 marks. Figure 7 illustrates the given information.

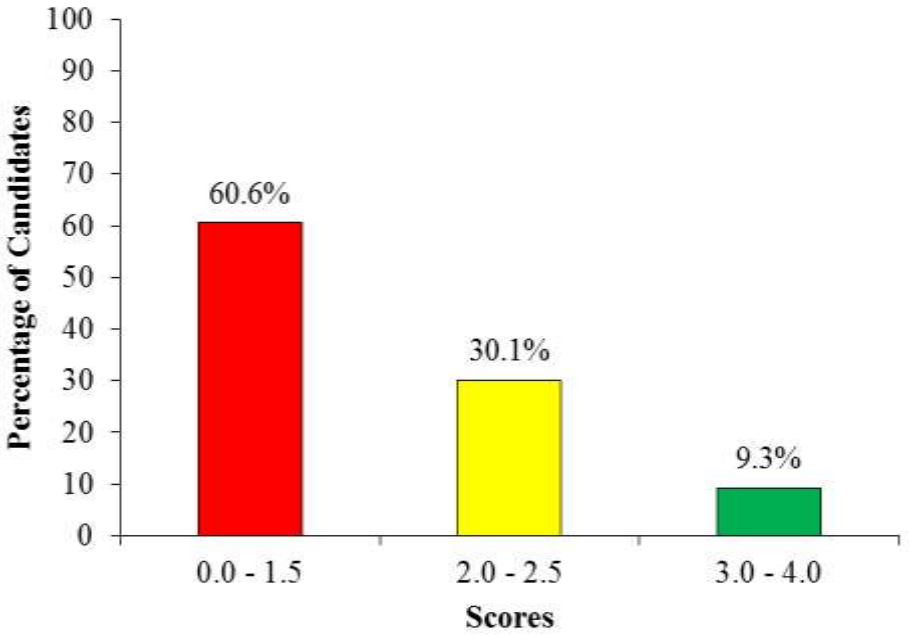
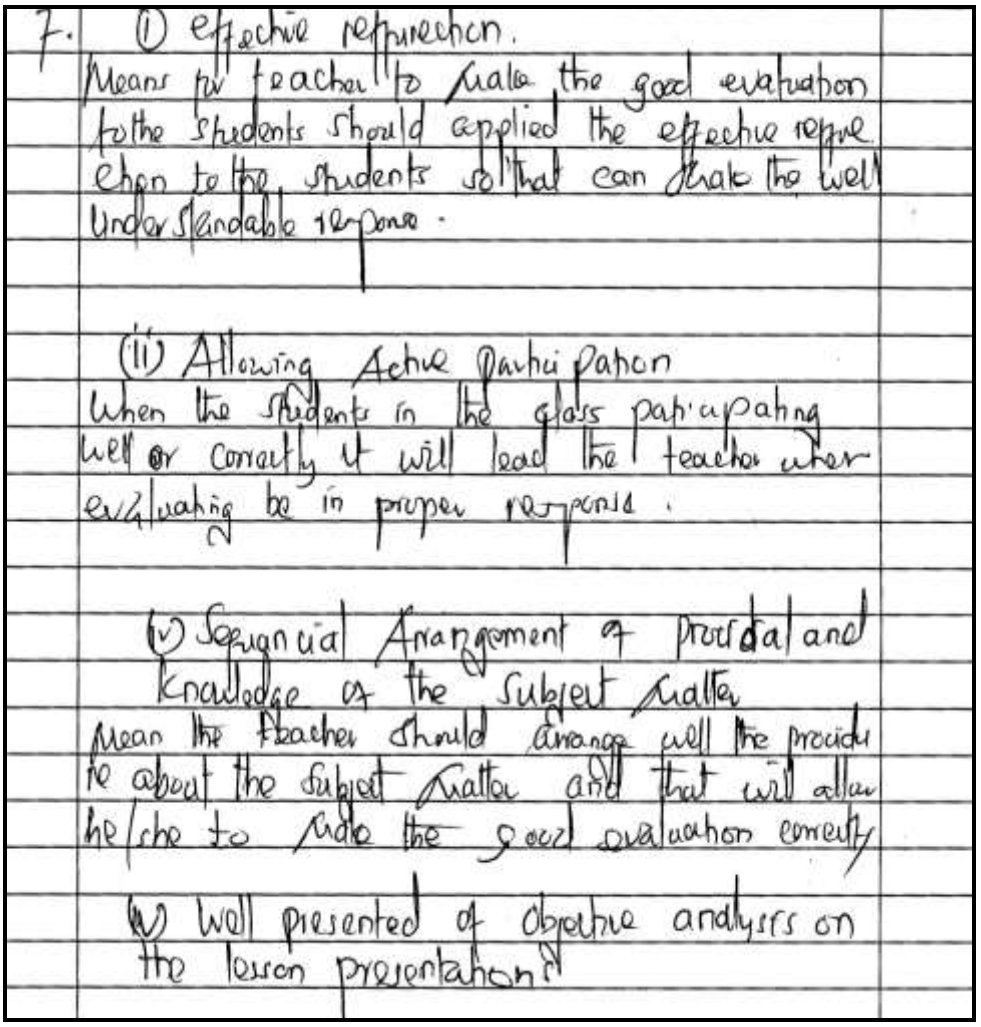


Figure 7: *Candidates' Performance on Question 7*

Figure 7 shows that among the candidates who attempted the question; 1,270 (60.6%) candidates scored 0 to 1.5 marks, 632 (30.1%) candidates scored from 2 to 2.5 marks and 193 (9.1%) candidates scored from 3 to 4 marks.

The analysis of the candidates' performance indicates that 1,270 (60.6%) candidates had weak performance in this question. These candidates had inadequate knowledge on qualities of a good evaluation in teaching and learning process. For example, some of them examined qualities as: *it aims at providing feedback; it aims at diagnosing teaching and learning difficulties; the process is pre-planned and properly administered; and its methods are clear and well defined.* Also other candidates gave the qualities as: *it should be objective to all students; it should involve all cognitive domain of learning; it should cover what the course required to*

be learnt. These responses imply that the candidate did not understand the demands of the question, they presented responses by guessing the qualities of a good evaluation due to lack of knowledge. Extract 7.1 shows a sample of responses from a candidate with poor performance.


7. (i) effective reflection.
Means for teacher to make the good evaluation to the students should applied the effective reflection to the students so that can make the well understandable response.

(ii) Allowing Active Participation
When the students in the class participating well or correctly it will lead the teacher after evaluating be in proper responds.

(iv) Sequencial Arrangement of procedural and Knowledge of the subject matter
Mean the teacher should arrange well the procedure about the subject matter and that will allow he/she to make the good evaluation correctly

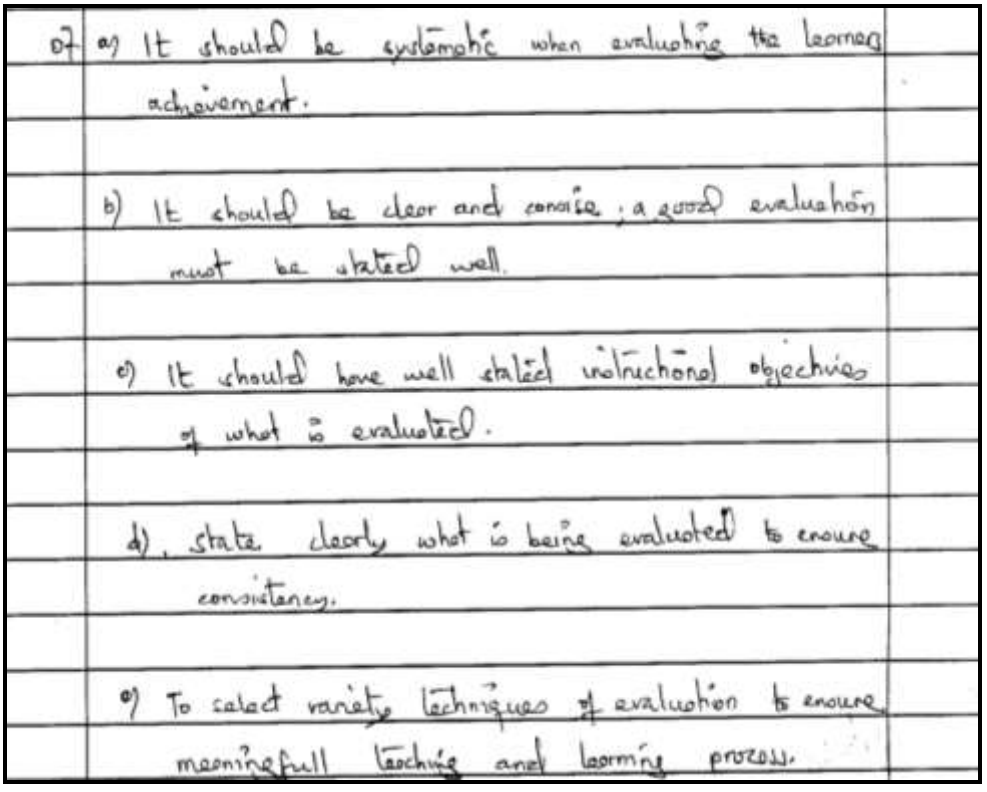
(v) Well presented of Objective analysis on the lesson presentation.

Extract 7.1: A sample of response by a candidate who failed to provided correct answers in Question 7.

Furthermore, candidates (30.1%) with average performance (2 – 2.5 marks) provided two out of four qualities of a good evaluation in teaching and learning process correctly. This shows that they had inadequate knowledge on the subject matter. For instance, one of the candidates wrote that: *the test to be evaluated should be free and fair to all examinees; good evaluation must be relevant for the content or area which is to be evaluated; good*

evaluation must be reasonable; good evaluation must be specific. This candidate did not meet exactly the demands of the question.

On the other hand, the analysis of the candidates' performance indicates that candidates (9.1%) who had good performance (3 - 4 marks) had adequate knowledge on the qualities of a good evaluation in the teaching and learning process. Their responses included the following: *methods of evaluating students should be well known to all examinees and should also be clearly defined; any evaluation activity should be well designed to promote students' academic achievement; good evaluation should be comprehensive and include multiple strategies.* Other candidates wrote that: *good evaluation should be based on SMART character meaning the responses to be evaluated should be Specific, Measurable, Attainable, Realistic and time bound*; and *in the process of teaching and learning the evaluation activity should be an integral aspect.* Extract 7.2 shows a sample of the good responses from the candidate's script.


07 a) It should be systematic when evaluating the learners achievement.

b) It should be clear and concise, a good evaluation must be stated well.

c) It should have well stated instructional objectives of what is evaluated.

d) State clearly what is being evaluated to ensure consistancy.

e) To select variety techniques of evaluation to ensure meaningfull teaching and learning process.

Extract 7.2: A sample of response by a candidate who satisfactorily examined the qualities of evaluation in Question 7.

2.1.8 Question 8: Test Construction

The candidates were required to give four strategies that examiners may use to control cheating in examinations.

A total of 2,095 candidates corresponding to 100 per cent attempted the question. Overall performance in this question was good as 1,746 (83.3%) candidates scored from 2 to 4 marks. Figure 8 illustrates the candidates' performance.

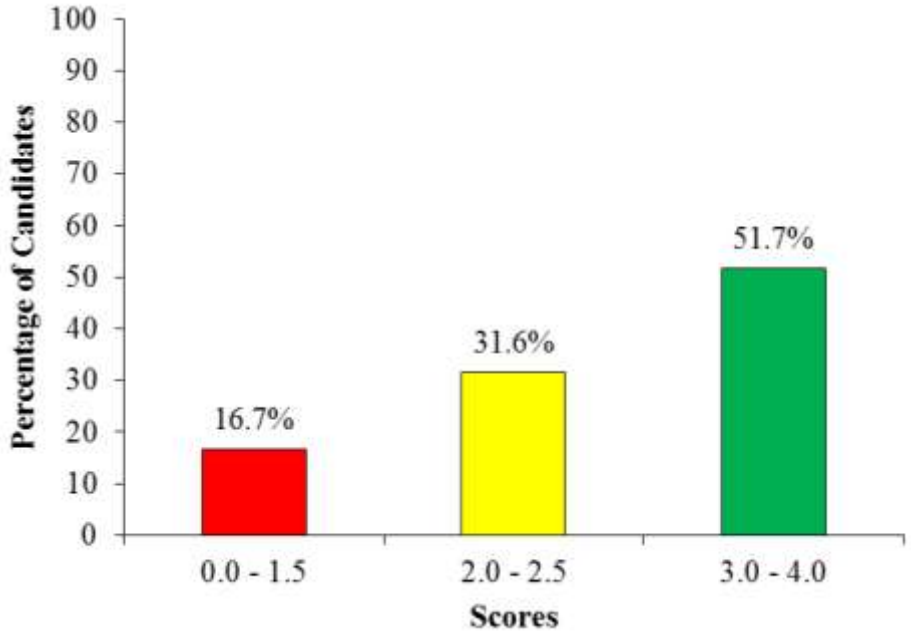
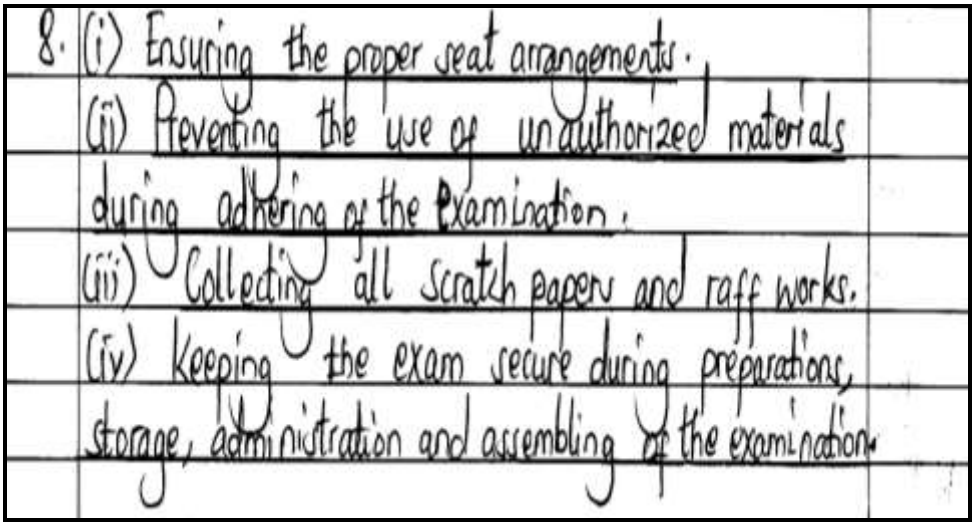


Figure 8: *The Candidates' Performance in Question 8*

Figure 7 indicates that the candidates' scores were as follows: 1,083 (51.7%) candidates scored 3 to 4 marks, among them 283 (13.5%) candidates scored 4 marks and 38.2 per cent scored 3 to 4 marks. Moreover, 663 (31.6%) candidates scored 2 to 2.5 marks, and 349 (16.7 %) candidates scored 0 to 1.5 marks.

The analysis of the candidates' performance shows that 1,083 (51.7%) candidates who scored 3 to 4 marks gave relevant strategies examiners may use to control cheating in examination. The candidates responses were such as: *The examiner should prepare the test that examinees will view all the items as relevant and fair to them; if possible the tester/examiner should use two forms of tests and give a different form to each row of students by*

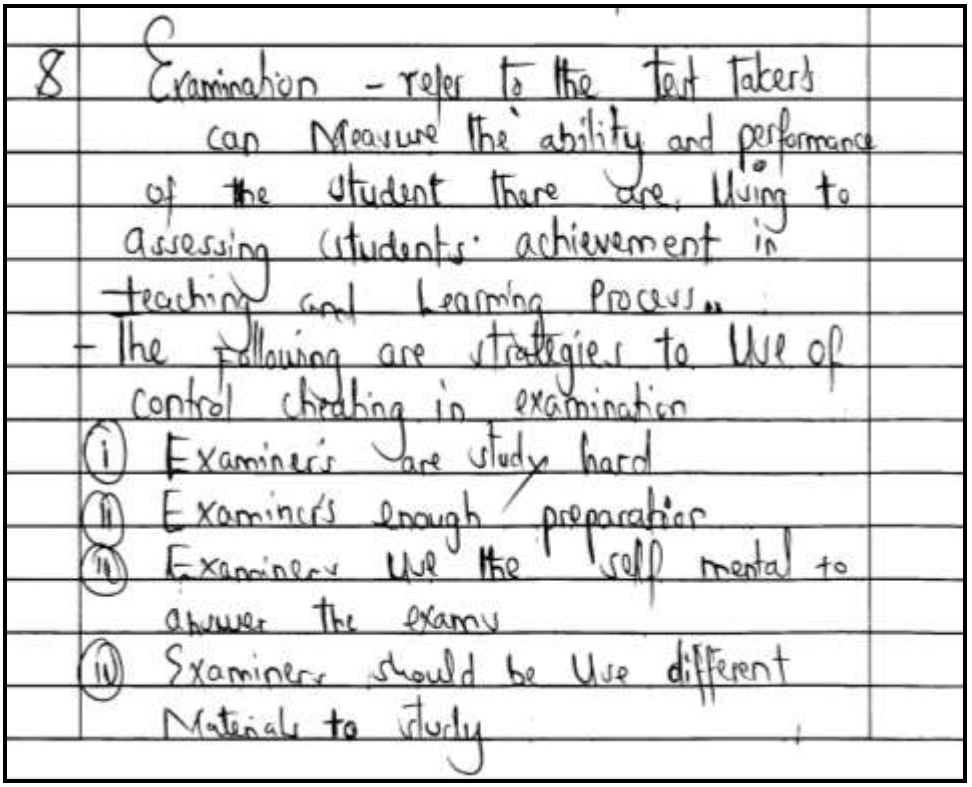
re arrange the order of items for the second question sheet/form; the invigilator should create and maintaining positive attitudes concerning the value of examination to different educational stakeholders; and the invigilator/ supervisor should periodically walk around the room and carefully observe how each examinee is doing the exam. Also other candidates had correct strategies as: *make examinees desks clean and remove all irrelevant materials in examination room before administration e.g. all scratched papers and rough work.* Therefore, candidates in this category had sufficient knowledge and were able to write strategies to control cheating in examinations. Extract 8.1 is a sample of good response from one of the candidates.


8. (i) Ensuring the proper seat arrangements.
(ii) Preventing the use of unauthorized materials during adhering of the Examination.
(iii) Collecting all Scratch papers and raff works.
(iv) Keeping the exam secure during preparations, storage, administration and assembling of the examination.

Extract 8.2: A sample of response by a candidate who was correctly wrote strategies to control cheating in examinations in Question 8.

On the other hand, the analysis of the candidates' performance shows 663 (31.7%) candidates had average performance, as they mixed both correct and incorrect strategies which can be used to control cheating in examination. For instance, one of the candidates wrote strategies as: *keeping secure the test during preparation simply because some students can see the test which can lead them to cheat***;** *use of special arrangement in sitting example the use of alphabetical order (A to Z).* Other candidates wrote that: *having them clear on their top of the table; clearly state the instructions where whoever go against the instruction will face the consequences.*

Further analysis of the candidates' responses in this question reveals that the candidates who scored 0 to 1.5 marks were 349 (16.7%). These candidates gave incorrect strategies that examiners may use to control cheating in examinations. Among these candidates, some of them neither gave one strategy nor attempted the question. Others wrote one strategy out of four. For example, one candidate wrote strategies as: *there should be special seating arrangement; keeping students aware about the instruction; to prepare the different examinations of the same content; to prepare a test or exams which are free and fair to all the learners.* Therefore, the candidate ended up scoring low marks. Extract 8.2 illustrates a sample of poor responses from one of the candidates.


8 Examination - refer to the test takers can Measure the ability and performance of the student there are. Using to assessing students' achievement in teaching and Learning Process.
- The following are strategies to Use of control cheating in examination
(i) Examiners are study hard
(ii) Examiners enough preparation
(iii) Examiners Use the self mental to answer the exams
(iv) Examiners should be Use different Materials to study

Extract 8.2: A sample of response by a candidate who failed to give strategies to control cheating in examinations in Question 8.

2.1.9 Question 9: Educational Research

The candidates were required to outline eight sources of literature review in research.

A total of 2,095 (100%) candidates attempted the question. The overall performance of the candidates was good as illustrated in Figure 9.

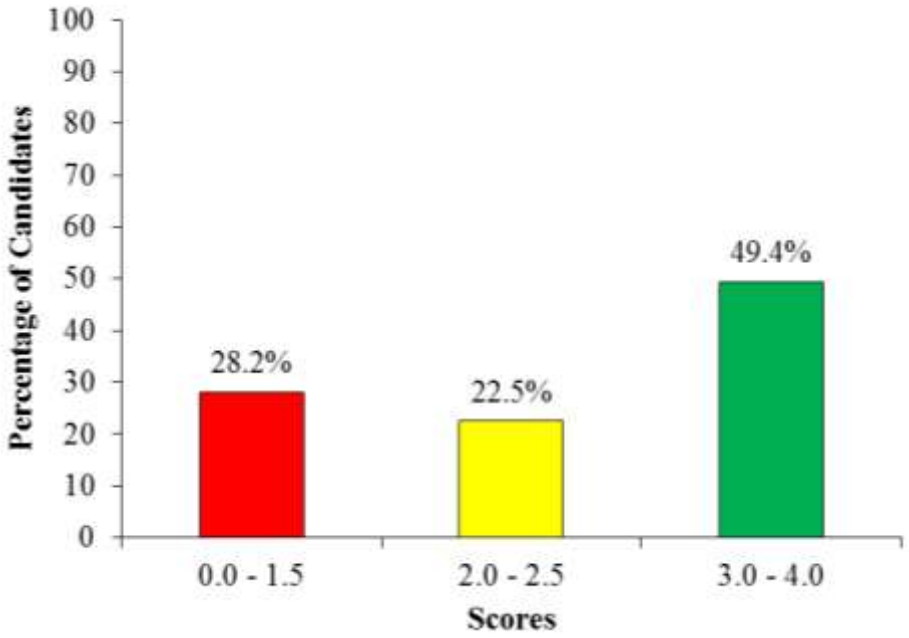
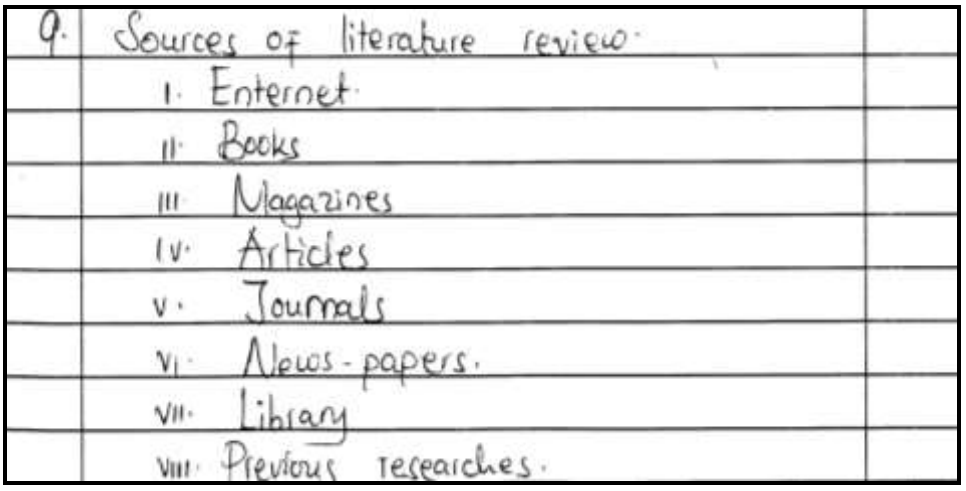


Figure 9: *The Candidates' Performance on Question 9*

Figure 9 indicates that 1,034 (49.1%) candidates scored 3 to 4 marks, 471 (22.5%) candidates scored 2 to 2.5 marks and 590 (28.2%) candidates scored 0 to 1.5 marks.

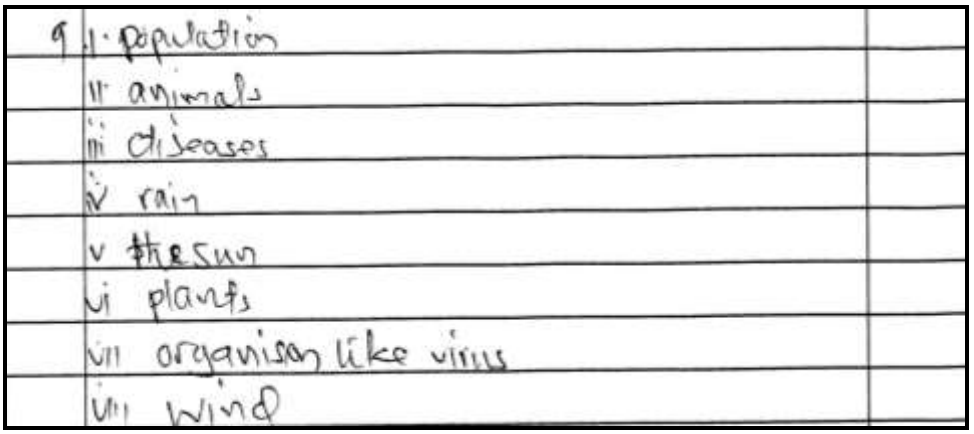
The candidates' response data analysis reveals that 1,034 (49.3%) who scored good marks (3 – 4) had adequate knowledge on the source of literature review, hence understood the question's requirements. For example, most candidates in this category outlined the sources of literature review as: *Books, Journal, E-material, Government report, Newspaper, Magazine, Mass media, Portfolio, Research work, Posters and Articles*. Others wrote sources such as: *conference documents; mass media such television; posters; sound records; memos; previous research work; encyclopaedia; and articles*. Extract 9.1 is a sample from a candidate who outlined sources of literature review correctly in Question 9.



Extract 9.1: A sample of response by a candidate who successfully identified sources of literature review in Question 9.

Moreover, the analysis shows that 471 (22.5%) candidates who scored average marks (2 - 2.5) outlined some of the sources of literature review while others provided both correct and incorrect sources. For instance, one candidate wrote responses as; *research record, journals, magazine, newspaper, books, syllabus and logbook*. Others wrote sources as: *pamphlets; text books; reference books; the use of direct reasoning; practical issues; journal; conference publications; and dissertations.* Looking at this candidate's responses it is obvious that the last two sources were not correct while the rest were correct. This example verifies that the candidates in this group had partial knowledge about sources of literature review.

Further analysis of candidates' responses indicates that 590 (28.2%) candidates with poor performance (scored from 0 to 1.5 marks) had inadequate knowledge and skills needed in this question. They presented incorrect sources of literature review such as: *practical issues, deducting from theories, inductive theories, personal experiences, pamphlets, projects work, and sensory experience.* This shows that the candidate misunderstood the question's requirements instead of identifying sources of literature review, they identified the sources of research problem. Extract 9.2 is a sample of such poor response from one of the candidates.



Extract 9.2: A sample of response by a candidate who failed to provide correct sources of literature review in question 9.

2.1.10 Question 10: Test Construction

The candidates were required to examine four factors to be considered by a teacher to ensure objectivity in scoring essay items.

The question was attempted by 2,095 candidates, equivalent to 100 per cent. Generally, the candidates' performance in this question was average, as 54.8 per cent of the candidates scored 2 to 2.5 marks. The data are presented in Figure 10.

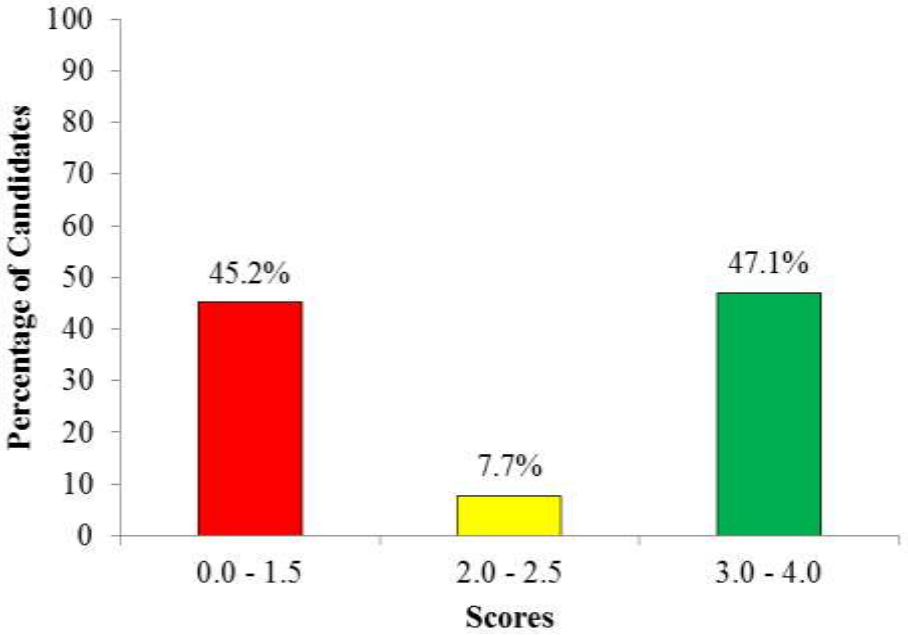
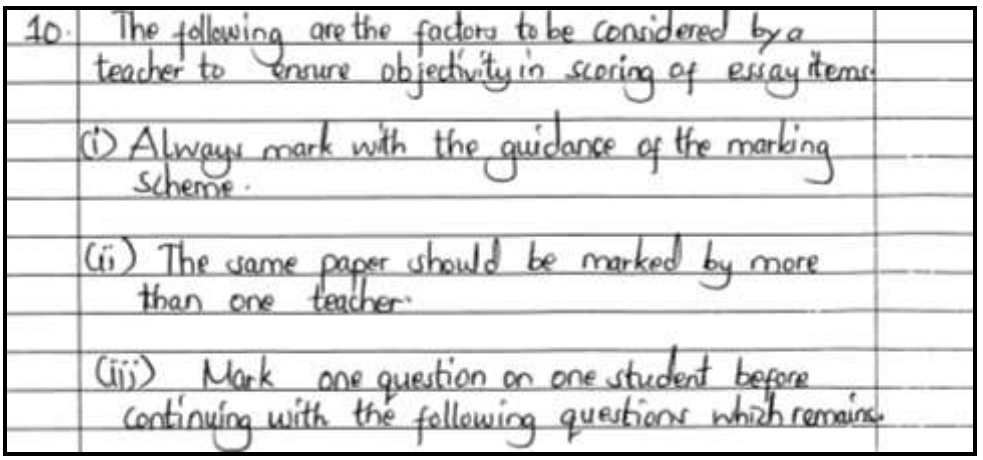


Figure 10: *The Candidates' Performance on Question 10*

The data in Figure 10 indicates that 986 (47.1 %) candidates scored 3 to 4 marks, 162 (7.7%) candidates scored 2 to 2.5 marks. However, the rest 947 (45.2 %) candidates scored 0 to 1.5 marks.

The candidates' response analysis shows that a few (7.7%) candidates who performed averagely (2 - 2.5 marks) mixed correct and incorrect factors that ensure objectivity in scoring essay items. Some of their responses were such as: *the marker should be physically fit because the activity requires high skills; there should be flexibility of scoring key or criteria; do not be influenced by irrelevant points; the marker should consider the main idea of the students.* These candidate's responses were correct, especially the last points while the first and the second points are incorrect.

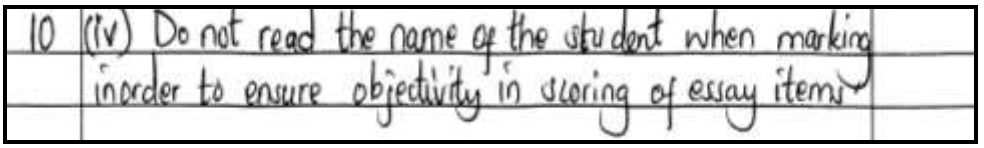
Furthermore, 47.1 per cent of the candidates who scored 3 to 4 marks understood the question requirement hence they provided relevant factors to be considered by teacher to ensure objectivity in scoring essay type items. The candidates' responses were such as: *to mark test by using a guide of marking scheme, marker should not refer to the name of candidate, mark one question to all answer scripts and last if possible essay test should be scored by more than one person.* Others wrote factors such as: *(i) the use of examination index numbers instead of student's name, (ii) set criteria for scoring each point of the essay item, (iii) don't mark candidate's scripts while you are unconscious such as being drunkard, (iv) the marker should be stress less and emotionally neutral*. These responses show that the candidates had sufficient skills about the factors to be considered by teacher to ensure objectivity in scoring essay items. Extract 10.1 is a sample of the responses from a candidate who performed correctly in this question.


10. The following are the factors to be considered by a teacher to ensure objectivity in scoring of essay items.

(i) Always mark with the guidance of the marking scheme.

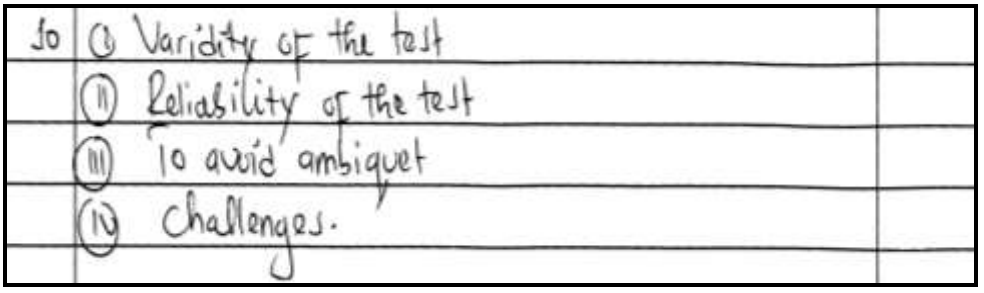
(ii) The same paper should be marked by more than one teacher.

(iii) Mark one question on one student before continuing with the following questions which remains.



Extract 10.1: A sample of response by a candidate who successfully examined factors to ensure objectivity in scoring essay items in Question 10.

Additionally, the analysis shows that 45.2 per cent of the candidates had weak performance (0 to 1.5 marks). They failed to explain even one factor while other candidates skipped this question. This shows that they were not able to understand while others lacked knowledge. For instance, those who managed to score at least 1 mark gave their responses like: *(i) introduction- the examiner should start reading the introduction of essay question to observe if it is matching with the question asked; (ii) points- all the examiners should check the points that the learner provided on the script; (iii) do not look on the name of the candidate; (iv) poor organization of points.* Looking at this response, only the third point was correct and resulted to scoring only 1 mark out of four. Extract 10.2 illustrates a sample of responses from a candidate who provided weak answers.



Extract 10.2: A sample of response by a candidate who failed to examine factors to ensure objectivity in scoring essay items in Question 10.

2.2 SECTION B: Essay Questions

This section had six (6) questions. Question eleven (11) was compulsory while other five (5) questions were optional. Candidates were required to choose only four questions out of five (5). Each question carried 15 marks. The total marks allocated to this section were sixty (60).

2.2.1 Question 11: Analysis and Interpretations of Test Results

Candidates were required to study the frequency distribution table representing English Language test results for 100 students from a certain Secondary School.

Scores	0-10	11-21	22-32	33-43	44-54	55-65	66-76	77-87	88-98
Frequency	2	8	10	12	18	20	15	10	5

From the table, the candidates were required to compute the following:

(a) The mean score of the distribution.
(b) The class interval size for the distribution.
(c) The highest and lowest score in the distribution.
(d) The modal class interval of the distribution.
(e) The variance of the distribution.

The question was compulsory and it was attempted by a total of 2,095 (100%) candidates. The overall performance in this question was average, since a total of 1,018 (48.6%) candidates scored 6 marks and above. Figure 11 illustrates the candidates' performance.

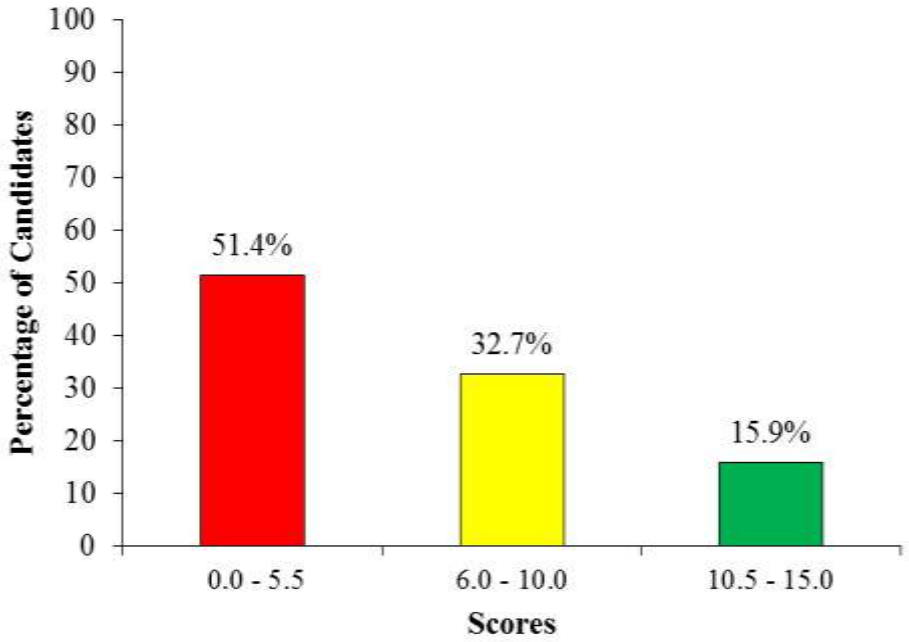
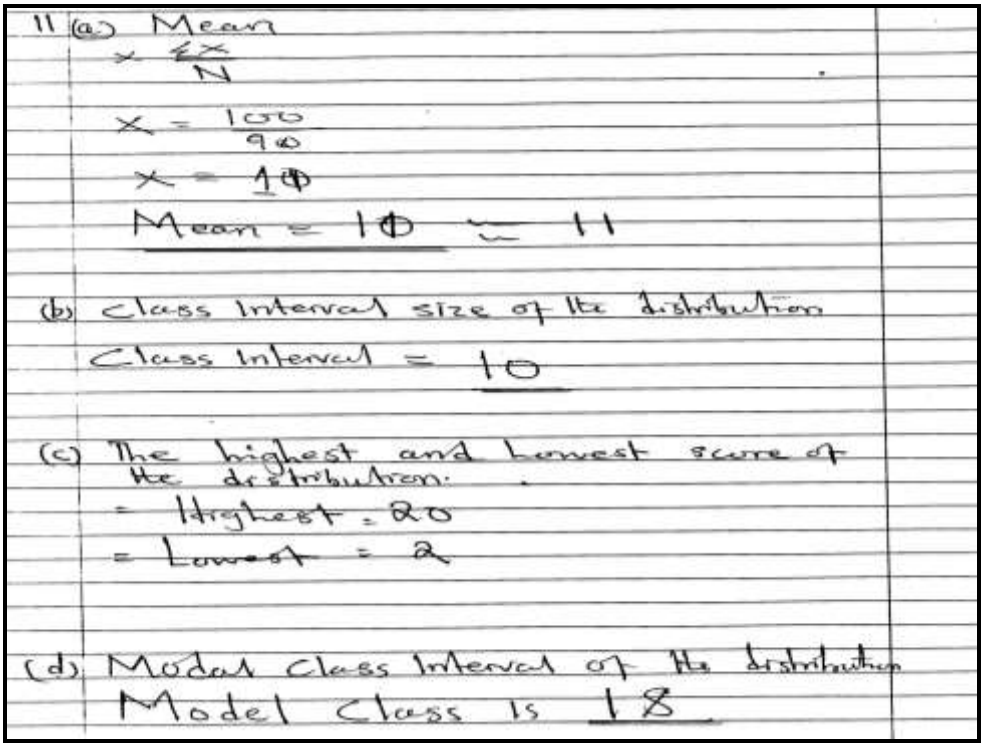


Figure 11*: The Candidates' Performance on Question 11*

Figure 11 indicates that 1,077 (51.4%) candidates scored 0 to 5.5 marks and 333 (15.9%) candidates scored 10.5 to 15 marks. The other 685 (32.7%) candidates scored 6 to 10 marks.

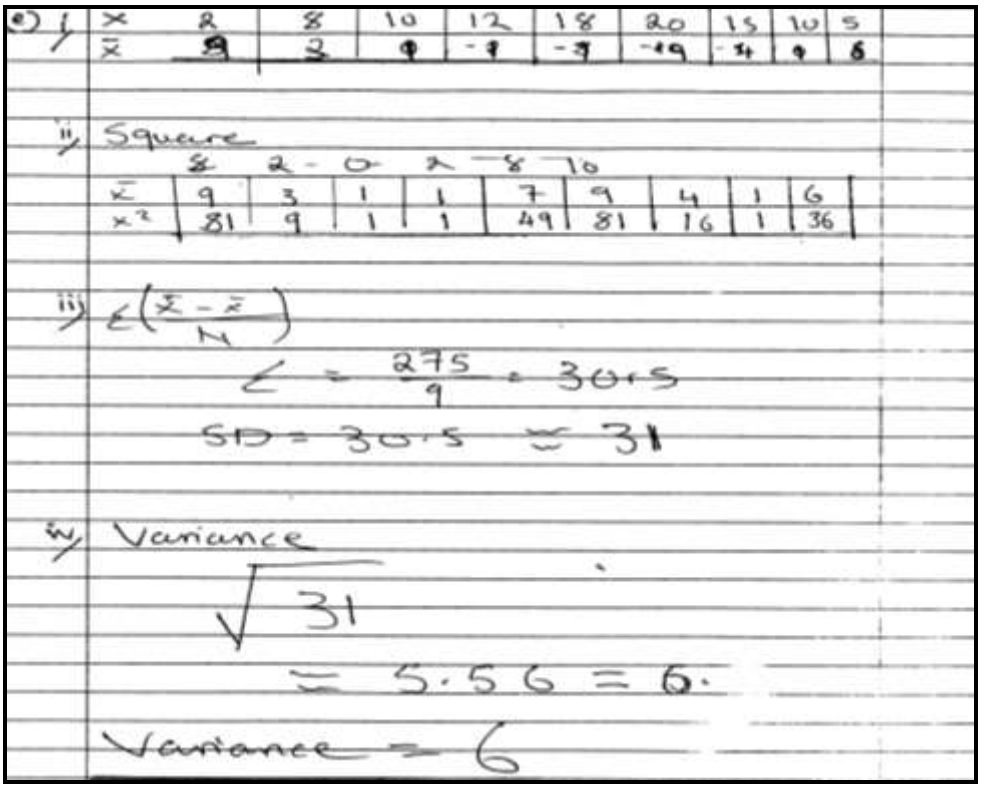
The analysis of the candidates' performance shows that the candidates who scored low marks (0 - 5.5) computed either one or two parts of the question while other candidates who scored 0 marks failed to respond to any part of the question. These candidates applied wrong formulae in computing the mean score and the variance of the distribution. For instance, one candidate applied wrong mean score formulae as: $\bar{x} = \frac{\Sigma f}{n} = \frac{100}{9}$ = 11.1 in part (a) and in part (b) the candidate wrote *median = highest score – lowest score (20-2=18).* In part (d) and (e) of the question the candidate wrote that, *the modal class interval is 18 and the variance is 43.8 w*ithout showing the processes towards these answers. Therefore, these responses, are evidence that candidates lacked sufficient knowledge on the area of content being tested. Extract 11.2 illustrates a sample of weak responses from a candidate.


11 (a) Mean
$\bar{x} = \frac{\Sigma x}{N}$
$\bar{x} = \frac{100}{90}$
$\bar{x} = 10$
Mean = 10 ≈ 11

(b) Class interval size of the distribution
Class interval = 10

(c) The highest and lowest score of the distribution:
= Highest = 20
= Lowest = 2

(d) Modal class interval of the distribution
Model class is 18



Extract 11.2: A sample of response by a candidate who provided incorrect answers in Question 11.

Further analysis of the candidates' responses reveals that 32.7 per cent of the candidates who scored average marks (6 - 10) calculated the first four parts (a, b, c, and d) of the question correctly, but they used improper variance formulae in part (e) of the question. Also some candidates provided incorrect responses without drawing appropriate table consisting data that would have guided them to arrive at the correct answer. For example, some candidates wrote the wrong variance formulae as:

Variance = $\frac{\Sigma\left(x-\bar{x}\right)^2}{\Sigma f}$ *and* $\frac{\sum fx}{\sum f} - \left(\frac{\Sigma fx}{\Sigma f}\right)^2$ instead of $\frac{\Sigma f\left(x-\bar{x}\right)^2}{\Sigma f}$

and $\frac{\Sigma f x^2}{\Sigma f} - \left(\frac{\Sigma fx}{\Sigma f}\right)^2$ *respectively.* These wrong formulae resulted to incorrect value of variance.

In addition, some candidates computed the mean score of the distribution incorrectly which resulted to incorrect value of variance. Most candidates

who attempted this question demonstrated good mastery of the content especially in parts (a), (b), (c), and (d).

Furthermore, the analysis shows that a total of 333 (15.9%) candidates who scored high marks (10.5 - 15) had good performance managed to compute correctly the: mean score; class interval size; highest and lowest score; modal class interval, and variance of the distribution as:

(a) The mean score = $\frac{\sum fx}{\sum f}$ *where by:* $\sum fx$ *= summation of frequencies times class marks, and* Σf *= summation of frequencies.*

Class Interval	*f*	*Class marks x*	*fx*	$x-\bar{x}$	$\left(x-\bar{x}\right)^2$	$f\left(x-\bar{x}\right)^2$
0-10	*2*	*5*	*10*	*-47.96*	*2300.162*	*4600.324*
11-21	*8*	*16*	*128*	*-36.96*	*1366.042*	*10928.336*
22-32	*10*	*27*	*270*	*-25.96*	*673.922*	*6739.22*
33-43	*12*	*38*	*456*	*-14.96*	*223.802*	*2685.624*
44-54	*18*	*49*	*882*	*-3.96*	*15.682*	*282.276*
55-65	*20*	*60*	*1200*	*7.04*	*49.562*	*991.24*
66-76	*15*	*71*	*1065*	*18.04*	*325.442*	*4881.63*
77-87	*10*	*82*	*820*	*29.04*	*843.322*	*8433.22*
88-98	*5*	*93*	*465*	*40.04*	*1603.202*	*8016.010*
Total(∑)	***100***	***441***	***5296***	***-35.64***	***7401.138***	***47557.88***

$\bar{x} = \frac{\sum fx}{\sum f} = \frac{5296}{100} = 52.96$

Therefore, the mean score of the distribution is 52.96

(b) The class interval size for the distribution can be computed by finding the real limits of any of the class intervals in the distribution and thereafter subtract the lower real limit from the upper real limit. For instance, class interval of 11- 21. The upper real limit is 21 + 0.5 = 21.5 while the lower real limit is 11- 0.5=10.5.Class interval size = 21.5-10.5

Therefore, the class interval size is 11.

(c) The highest and lowest score can be computed as follow:

(i) The highest score =finding the highest class mark in the distribution which is 93

(ii) The lowest score = finding the lowest class mark in the distribution which is 5

(d) The modal class interval can be computed by looking the class interval with the highest frequency compared to any other interval in the distribution which is 55-65

(e) The variance of the distribution is computed by using the following formula:

$$\text{Variance} = \frac{\Sigma f\left(x-\bar{x}\right)^2}{\Sigma f} = \frac{47557.88}{100} = 475.5788 \text{ or } 476$$

Therefore, the variance of the distribution is 476

Many candidates who attempted this question had some variations in approaching the question since the majority drew a table with columns and rows to illustrate the given data from the distribution and some additional data in order to meet the requirement of the formulae used, but they all reached to the correct answer. Meanwhile a few candidates had different approaches. They used more than one tables as per demands of each part of the question. Extract 11.1 illustrates the sample from a candidate with good responses in question 11.

11.

Scores	Frequency (f)	Mid-point (x)	fx	$x-\bar{x}$	$(x-\bar{x})^2$	$f(x-\bar{x})^2$
88-98	5	93	465	40.04	1602.80	8014
77-87	10	82	820	29.04	843.32	8433.2
66-76	15	71	1065	18.04	325.40	4881.6
55-65	20	60	1200	7.04	49.56	991.2
44-54	15	49	870	-3.96	15.68	235.24
33-43	12	38	456	-14.96	223.80	2685.6
22-32	10	27	270	-25.96	673.92	6739.2
11-21	8	16	128	-36.96	1366.40	10923.32
0-10	2	5	10	-47.96	2300.16	4600.32
	100		5296			43557.68

(a) The mean score $(\bar{x}) = \frac{\Sigma fx}{\Sigma f}$

$\bar{x} = \frac{5296}{100}$

$\bar{x} = 52.96$

∴ The mean score of the distribution is 52.96

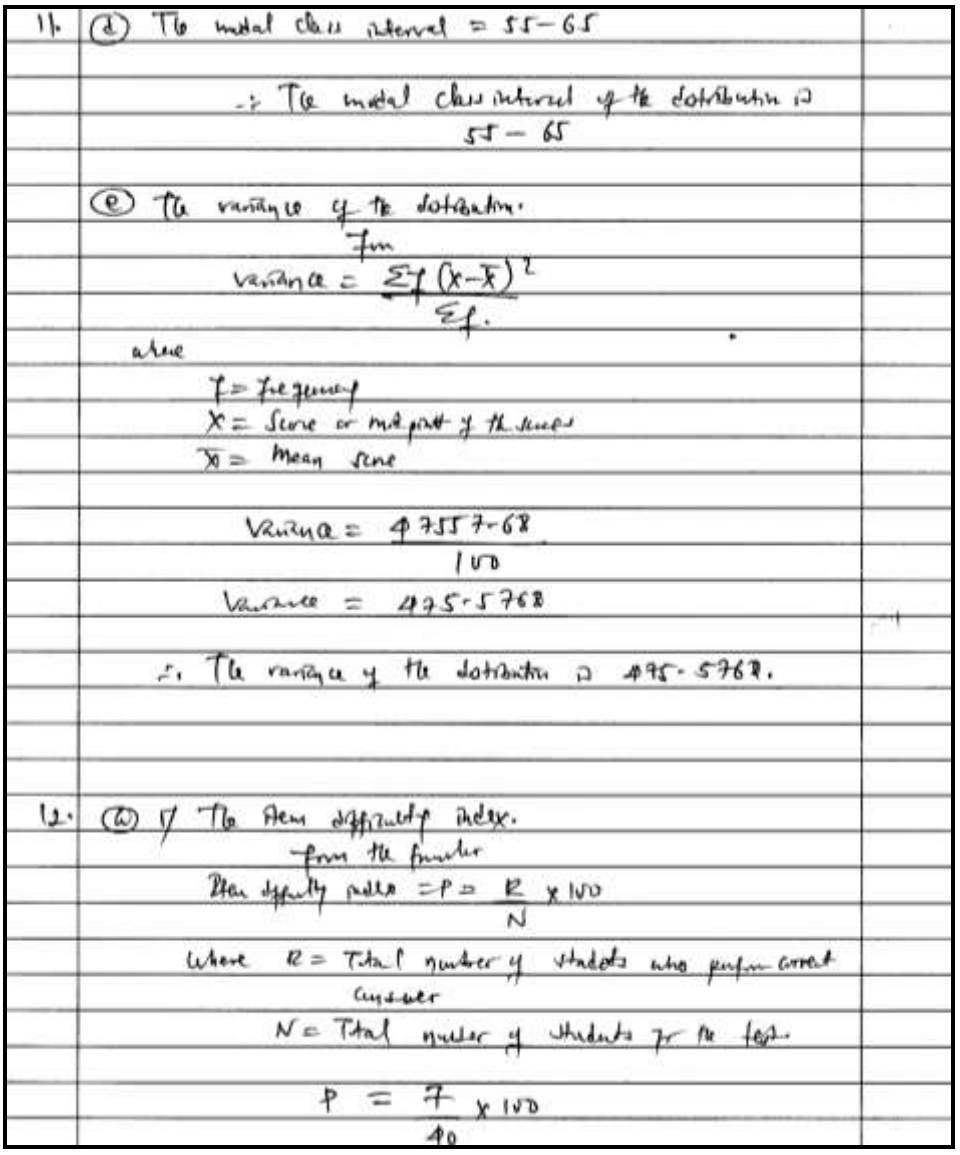
(b) The class interval size = 11 − 0

= 11.

∴ The class interval size = 11

(c) The highest and lowest score

∴ The highest score is 93 and lowest score is 5



Extract 11.1: A sample of response by a candidate who successfully computed all parts in Question 11.

2.2.2 Question 12: Analysis and Interpretation of Test Results

Candidates were required to study the results obtained from 40 students taken as a sample for item X analysis where letter ‘B’ was the correct answer.

Item X	Possible Responses and Choice						Total
	Omit	A	B*	C	D	E	
High Achievers	2	11	5	1	1	0	20
Low Achievers	0	12	2	2	4	0	20
Total	**2**	**23**	**7**	**3**	**5**	**0**	**40**

In Part (a), the candidates were required to compute: (i) the item difficult index and (ii) the discrimination index. In part (b), candidates were instructed to state the level of difficult of the item and to give the two reasons on the (a) (i) and (ii) computation results.

The question was attempted by 959 candidates corresponding to 45.8 per cent. Generally, the candidates' performance in this question was good as illustrated in Figure 12.

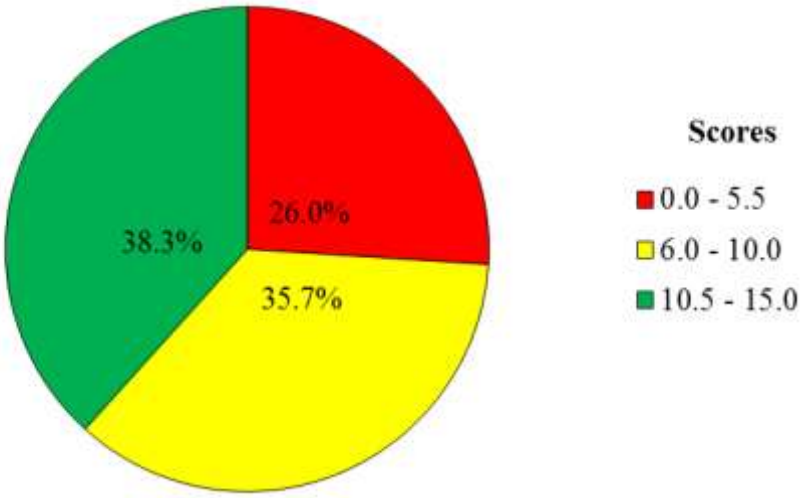


Figure 12: *The Candidates' Performance on Question 12*

Figure 12 shows the candidates' scores as follows: 364 (38.3%) candidates scored 10.5 to 15 marks, 343 (35.7%) candidates scored 6 to 10 marks, and 249 (26%) candidates scored 0 to 5.5 marks.

The analysis of the candidates' performance indicates that 38.3 per cent of the candidates performed well, since they computed correctly the two indices in part (a) and in part (b) they gave two reasons for the levels of indices obtained in part (a) (i) and (ii). For example, many candidates successfully computed two parts as follows:

(a)(i) The item difficult index $P = \frac{RU + RL}{T} \times 100\%$ *where by:*

RU= Representatives or sample of students who get the item right from the upper group.

RL= Representatives or sample of students who get the item right from the lower group.

T= Total number of students taken as sample

P= Item difficult index

Given that: RU= 5, RL=2, and T=40 (20 from upper and 20 from lower)

Solution:

$$P= \frac{RU + RL}{T} \times 100\% = \frac{5+2}{40} \times 100\% = \frac{7}{40} \times 100\%$$

$$P= \frac{700}{40}\% = 17.5\%$$

Therefore, the item difficult index is 17.5%

Also, in responding to part (a) (ii) of the question, the candidate managed to compute the item discrimination index as:

(a)(ii) The item discrimination index $D= \frac{RU - RL}{\frac{1}{2}T}$ *where by:*

RU= Representatives or sample of students who get the item right from the upper group.

RL=Representatives or sample of students who get the item right from the lower group.

T=Total number of students taken as sample

D=Item discrimination index

Given that: RU= 5, RL=2, and T=40 (20 from upper and 20 from lower)

Solution:

$$D= \frac{RU - RL}{\frac{1}{2}T} = \frac{5-2}{\frac{1}{2\times 40}} = \frac{3}{20} = 0.15$$

Therefore, the item discrimination index is 0.15

Candidates who chose this question especially parts of part (a) demonstrated good abilities in computing the two indices of item *X*, although there were different styles of writing the indices formulae. For

instance, some candidates wrote item difficult formula as: $\frac{R}{T}\times 100\%$ and item discrimination index formulae as: $\frac{RU-RL}{1/2N}$ instead of $P=\frac{RU+RL}{T}\times 100\%$ for item difficult index and $D=\frac{RU-RL}{1/2T}$ for item discrimination index respectively. Despite the different styles of using the two formulae, candidates finally ended up with the correct answer that the question required.

In part (b) of the question which instructed the candidates to state two reasons on the levels of indices obtained in part (a) (i) and (ii), candidates were able to state the reasons but the qualities of their reasons were slightly different. For example, one candidate gave the two reasons as:

(b) (i) The item difficult level of 17.5% shows that the item X was very difficult because the level is within the range of 0-29% which is interpreted as very difficult item.

(b) (ii) The item discrimination level of 0.15 shows that the item X was bad or poor in discriminating the higher and lower achievers because the level obtained is less than 0.4

Looking at those valuable reasons highlighted above, there were also a few candidates who added another reason and they stated it as:
(b) (iii) The item X was inconsistence and it should be removed from the bank of questions because it was poorly constructed by attracting few candidates from both groups (higher and lower achievers). Extract 12.1 is a sample of good response from one of the candidates.

12 (a) (i) The item difficulty index (P) = ?
From the formular.

Item difficulty index (P) = number of students with correct response (R).
Total number of students taken as a sample (T). ×100%

$$P = \frac{R}{T} \times 100\%$$

Since the correct answer was B,
then, R = 5+2 = 7.

T = 40.

Thus,

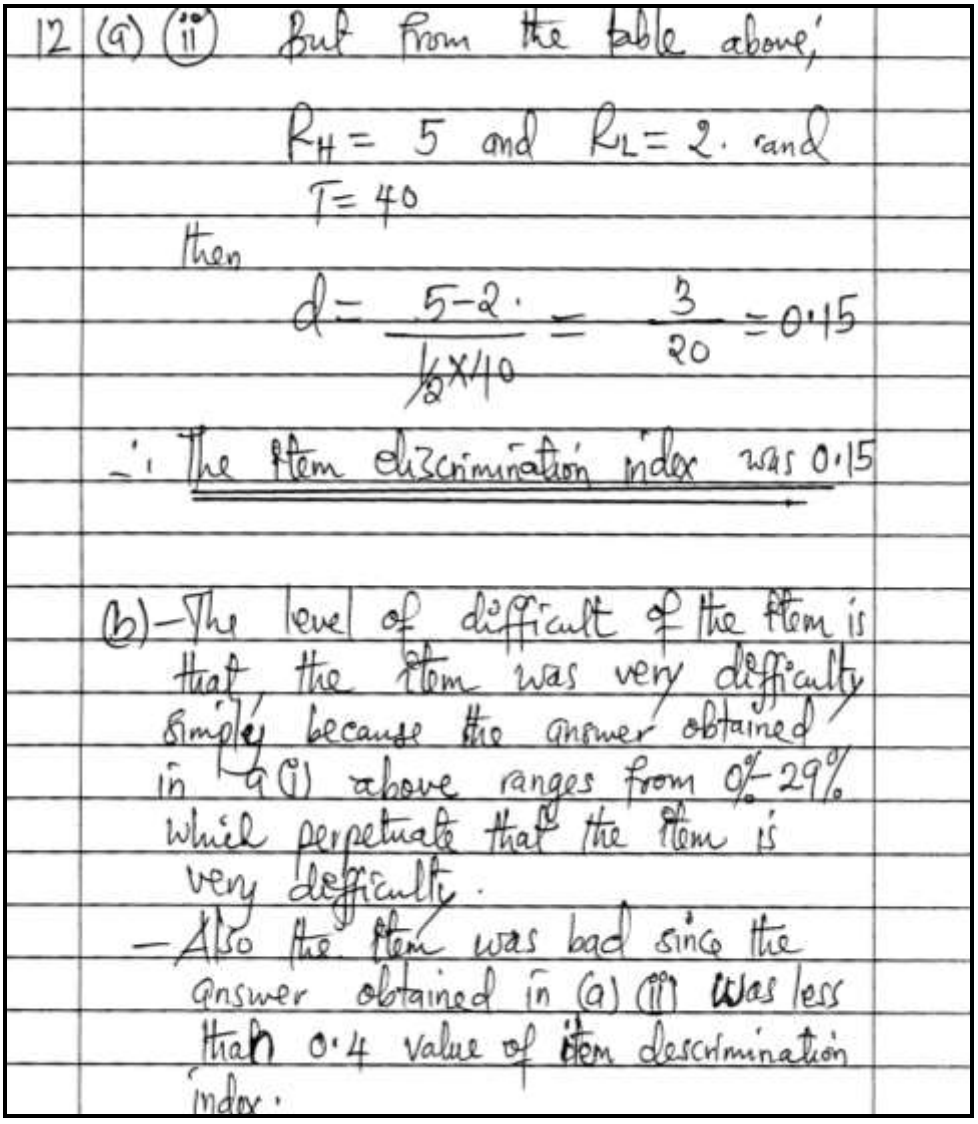
$$P = \frac{7}{40} \times 100\% = 17.5\%$$

∴ The difficulty index of the item was 17.5%

(a) (ii) Required, Item discrimination index (d)
= ?

From,
Item discrimination index (d) = Difference between correct response number of higher achievers and Low achievers
½ Total sample.

$$d = \frac{R_H - R_L}{\frac{1}{2}T}$$



Extract 12.1: A sample of response by a candidate who responded Question 12 correctly.

Moreover, the analysis shows that a total of 343 (35.7%) candidates whose score ranged from 6 to 10 marks, were able to able to provide correct responses either in part (a) or (b) of the question. But evidence from the candidates' scripts reveals that, candidates who attempted this question demonstrated competence in computing the values of two indices item difficult index and item of discrimination index, but failed to state two reasons on the level of each index obtained in part (a). Examples of their reasons were such as:

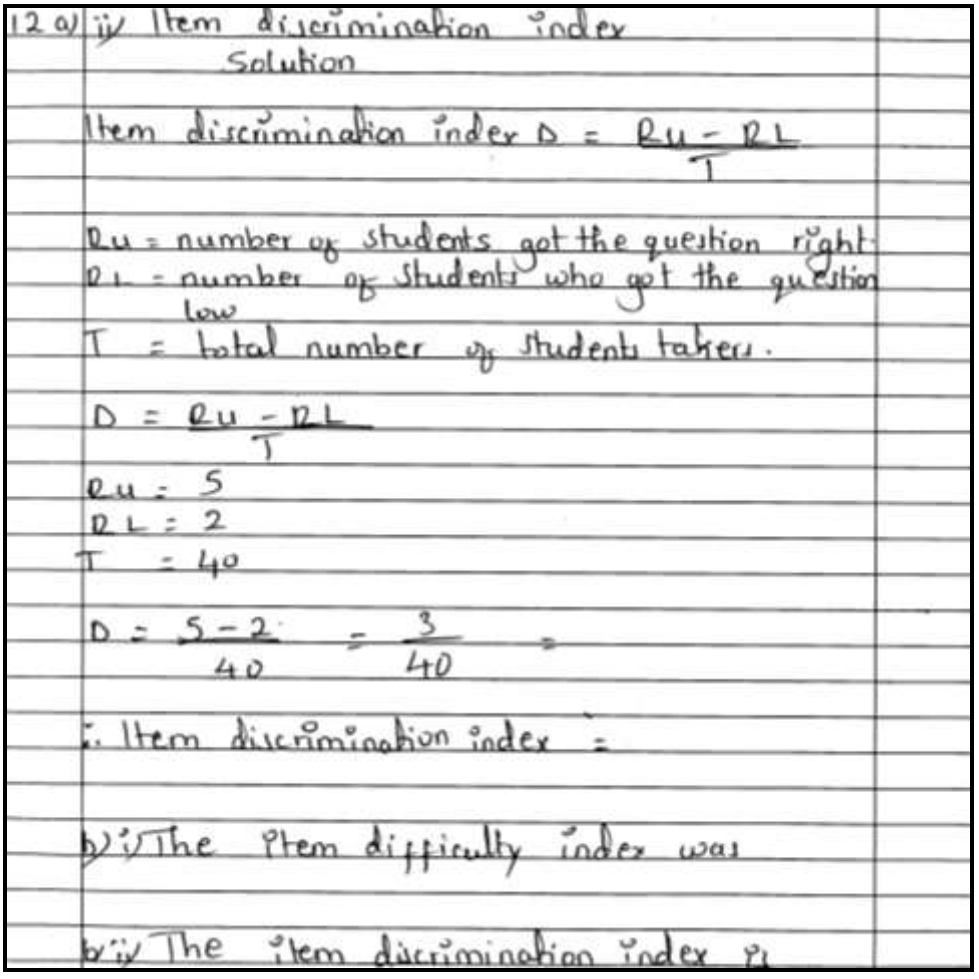
(b) (i) 17.5% shows that the performance is very low due to responses showed by levels and

(b) (ii) 0.15 is poor because the level is less than 0.4. Therefore, the performance is low.

It was noted that the reasons given were not correct, because when analysing a particular item in the test, the focus is not on the test performance but the level of item difficult and the level of discriminating the higher and lower achievers in the item. Therefore, failure to state two reasons correctly resulted to scoring lower marks. Other candidates gave wrong reasons as: *teacher should change the teaching methods in order to ensure high performance; the result was poor because the level is under 0.4; the level of difficult is bad because the performance is below 0.4*. In part (b), most of the candidates gave incorrect reasons due to insufficient knowledge of interpreting the value of indices.

Furthermore, a total of 249 (26%) candidates who had poor performance had scored ranging from 0 to 5.5 marks in this question. There were several reasons behind their poor performance. Some candidates were unable to use the appropriate formulae for computing two indices in part (a) of the question. Since the correct values from the computation in part (a) of the question determine the correctness of reasons in part (b) of the question, candidates who provided wrong responses of the two indices also failed to state two reasons based on the values of indices obtained in part (a). Also some candidates could only recall the appropriate formulae but failed to compute the indices. This tends to lower their marks. For instance, one candidate wrote the correct formulae of item difficult index as: $P = \left(\frac{Rupper - Rlower}{Total} \right) \times 100\%$ but failed to interpret a total number of students taken as sample by considering only one of the two groups (20 students from either group) instead of taking the totality of the higher and lower achievers which is 40 students. Therefore, the candidate missed some marks because of wrong interpretation and incorrect computation. In addition, other candidates gave the wrong reasons in part (b) of the question as: *part (b)(i) the level of difficult in item X is difficult because the difficult index is between 30-49% which define the level of difficult and in part(b)(ii) The level of difficult is bad item because the discrimination index is less than 0.4*. These two reasons are incorrect because of incorrect values of two indices obtained in part (a) of the question. This shows that

candidates did not understanding the demands of the question. Surprising in part (b)(ii) the question instructed candidates to state discrimination level with two reasons, but some of the candidate stated again the level of difficult which was already stated in part (b)(i). The analysis also proves that some responses in part (b) of the question lacked sufficient reasons about the level of indices because those who attempted this part were only ended by stating the levels of two indices as item difficulty level is very difficult, and item discrimination level is weak or bad in discriminating higher and lower achievers, without giving reasons as a matter of defending what they have stated. Extract 12.2 shows a sample of weak performance in this question.


12 a) ii) Item discrimination index
Solution

Item discrimination index $D = \frac{R_U - R_L}{T}$

R_U = number of students got the question right
R_L = number of students who got the question low
T = total number of students takers.

$D = \frac{R_U - R_L}{T}$

$R_U = 5$
$R_L = 2$
$T = 40$

$D = \frac{5-2}{40} = \frac{3}{40}$ =

∴ Item discrimination index =

b) i) The item difficulty index was

b) ii) The item discrimination index is

Extract 12.2: A sample of response by a candidate who failed to provide correct answers to Question 12.

2.2.3 Question 13: Educational Measurement

By using examples from owns experience, candidates were required to explain five roles of measurement in education.

The question was attempted by 1776 (84.6%) candidates. The general performance of the candidates was good as 1748 (98.4%) candidates scored 6 to 15 marks. Figure 13 summarizes candidates' performance in this Question.

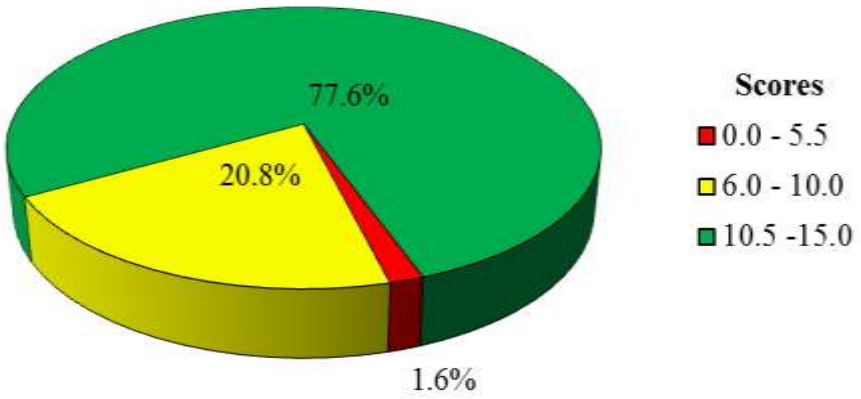
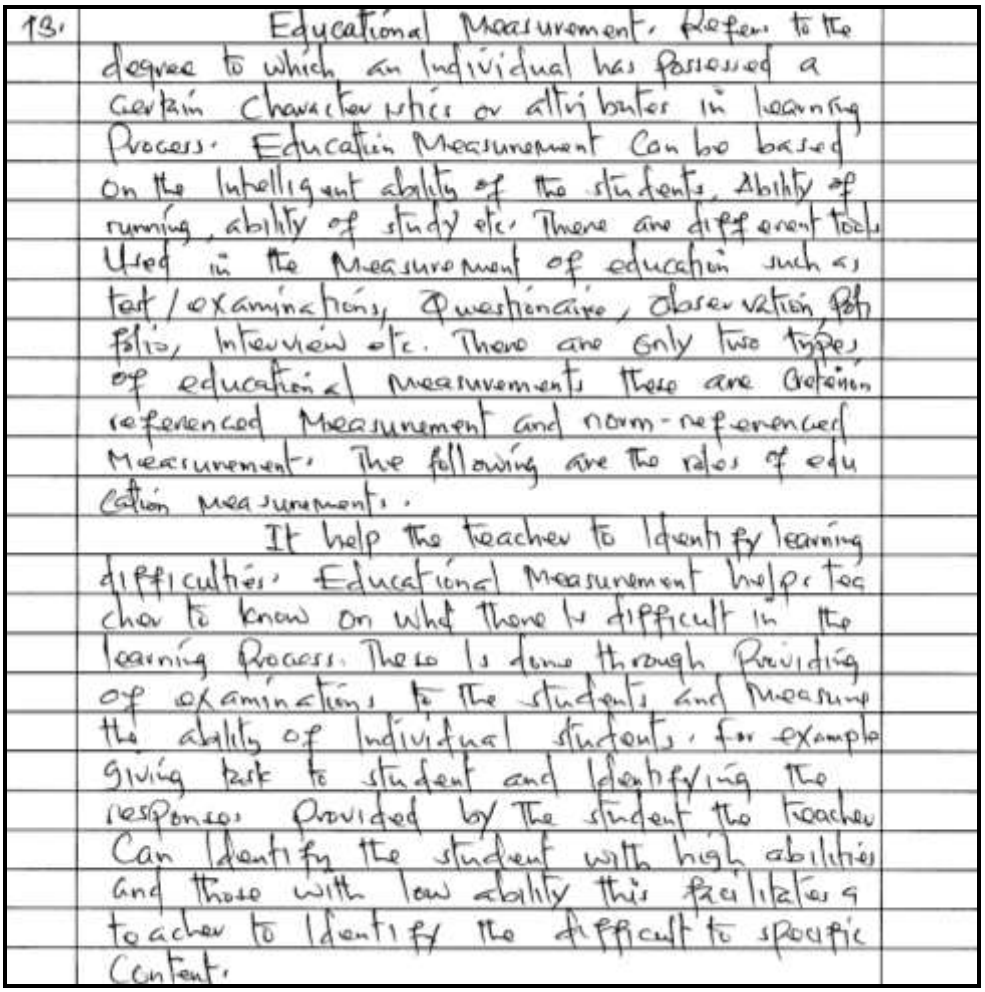


Figure 13: *The Candidates' Performance on Question 13*

Figure 13 indicates that among the candidates who attempted this question; 1379 (77.6%) candidates scored from 10.5 to 15 marks, 369 (20.8%) candidates scored from 6 to 10 marks, and 28 (1.6%) candidates scored from 0 to 5.5 marks.

The analysis shows that 1379 (77.6%) of the candidates who performed well in this question were able to explain five roles of measurement in education. They were also able to provide relevant examples for each role. Moreover, these candidates managed to provide relevant introductions and conclusion at the end their responses. The candidates in this category demonstrated good abilities in defining term measurement as: *the process of assigning numbers or numerical values to a particular event so as to judge on the degree or the extent to which a learner or an object possesses a particular characteristic or attribute.* In the main body, the candidates explained relevant roles of measurement in education as follows: *measurement used as a guide in decision making procedures when assessing the effectiveness and efficiency of the programme; measurement may also used as the basis for establishing and maintaining standard of*

learning; the results from measurement classifies or select individuals to join for special programmes of study, training, streaming or both; measurement also helps educational planners on issues related to curriculum development or educational improvement. However, there were variations in explaining roles of measurement in education explained: *measurement helps to determine criteria for recognition and awards of learners and teachers for the effort they put into the programme; it leads to the motivation and competition amongst students; it enables a teacher to evaluate his/her teaching and learning strategies.* Looking at all these roles, the candidates proved to possess adequate knowledge on roles of measurement in education. Extract 13.1 shows a sample of good responses in this question.



13. Educational Measurement. Refers to the degree to which an individual has possessed a certain characteristics or attributes in learning Process. Education Measurement Can be based on the Intelligent ability of the students, Ability of running, ability of study etc. There are different tools Used in the Measurement of education such as test / examinations, Questionaire, observation, Portfolio, Interview etc. There are only two types of educational measurements there are Criterion referenced Measurement and norm-referenced Measurement. The following are the roles of education measurements.

It help the teacher to Identify learning difficulties. Educational Measurement helps teacher to know on what there is difficult in the learning Process. This is done through Providing of examinations to the students and Measuring the ability of Individual students. for example giving task to student and Identifying the responses Provided by the student the teacher Can Identify the student with high abilities and those with low ability this facilitates a teacher to Identify the difficult to specific Content.

It helps teachers to determine the students behaviours. Through obtaining diffe rent characteristics ou attribute possesed by each individual a teacher can easy Under stand the behaviours to each students, for example, obeying of school regulation, school attendances and classroom activity participation.

13. It helps teachers to make self-evaluation. Through measurement a teacher can make self evalu ation on what is achievable or not achievable during the learning process. It is bother for a teacher to obtain the learning problems that appeared in order to make evaluation that makes a teacher to improve either in technique or method employed during learning and also helps a teacher to come with relevant knowledge.

It helps curriculum developers to set the learning objectives, determining of the learning difficults, and appropriateness of the subject content is done by measurement either through questions ou examination hence the curriculum developers is able to determine what is supposed to be learne d by the students and at which level.

It helps teachers to identify learners with abnormalities. Learners are different in intellectual level such as perception and recalling due to differe nt abnormalities to some individual so it is better to obtain the learners that needs special care during the learning process. Examples these learners with hearing impairment, low vision etc.

Generally. Education measurement helps a teacher to know the cognitive level of the stu- dents which will help them to know the materi al required and which methodologies must be used so as to achieve the learning obje- ctive as intended.

Extract 13.1: A sample of response by a candidate who correctly explained the roles of measurement in education in question 13.

Moreover, the analysis shows that 369 (20.8%) candidates who had an average performance (scoring from 6 to 10 marks) managed to explain different roles of measurement in education correctly, but failed to give relevant examples for each role they explained. These candidates briefly wrote the roles of measurement in education as: *measurement helps to know the students' report in learning process; measurement can be provided to students using exams, tests, quizzes or assignment; measurement helps teacher compare individual abilities and progress; measurement help solving and identify learning difficulties; measurement is useful in motivating learners during the course of study; measurement helps for certifying and awarding learners at the end of the programme.* Other candidates presented the responses like*: measurement helps teacher to predict future performance of the learners; measurement results help in streaming and ranking students.* These valuable points lacked explanations and examples which led them to score lower marks.

On the other hand, the candidates who performed weakly in this question failed completely to define the term measurement and also to explain the roles of measurement in education. For instance, some candidates did not understand the question's requirements. Therefore, they provided unclear answers which did not meet the demand of the question. For example, one of the candidates in this question responded that measurement: *helps teacher to know the degree of something; helps to know the length of something; helps to know the amount of something; to know the angle of something*. Similarly, another candidate wrote the role of measurement as: *to conduct presentation; to conduct group discussion; and to conduct project and research*. These responses from two candidates show that candidates' responses were completely deviated from requirement of the question due to irrelevance of their points. Extract 13.2 illustrates a sample of response from one of the candidates who performed this question poorly.

13 Measurement is the process of evaluating the knowledge and skills of the some one, so the measurement in learners deal with evaluating the knowlede of the learners; also the measurement in education have the following roles:

To know the learning process. The measurement help the teacher to know the learning process of the certain students in the class and can make the judgement about the progress of the certain learners.

Lead to inform the teacher if the lesson is understand to the learners: Through measurent the teacher can aquire the information if their lesson is not understand to the learners and can take action like use other method of teaching and other alternative of teaching the certain lesson.

Lead to improve the teaching and learning process; since the measurement lead to teacher to know if their lesson is not well understand to their learners and use the alternative way of teachnge well the certain lesson to their learners.

It lead to make the learner to be creative: Since the teacher some time can use the sorraunding environment or current issues to measure their learners and this can make the learner to thing beyond.

It Lead to improve learner interest: The measurand it lead the student to learning the interest in learn. like giving of motivation to those students who can perform well in class, so this make this kind of behavior to continuen.

Generally the above are the importance or role of measurement in education.

Extract 13.3: A sample of response by a candidate who poorly explained the roles of measurement in education in Question 13.

2.2.4 Question 14: Educational Research

The candidates were required to identify five characteristics which qualify educational research as a scientific process.

This question was attempted by 518 candidates corresponding to 24.7 per cent. Data analysis indicates that the question had average performance as that 309 (59.6%) candidates scored 6 to 15 marks. The data is summarised in Figure 14.

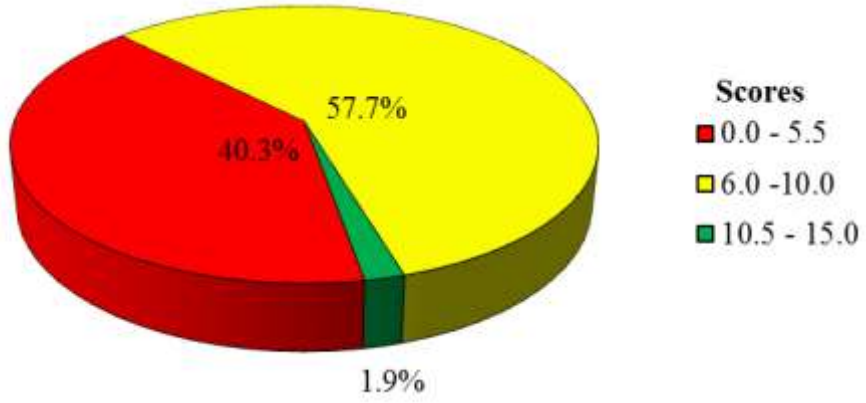


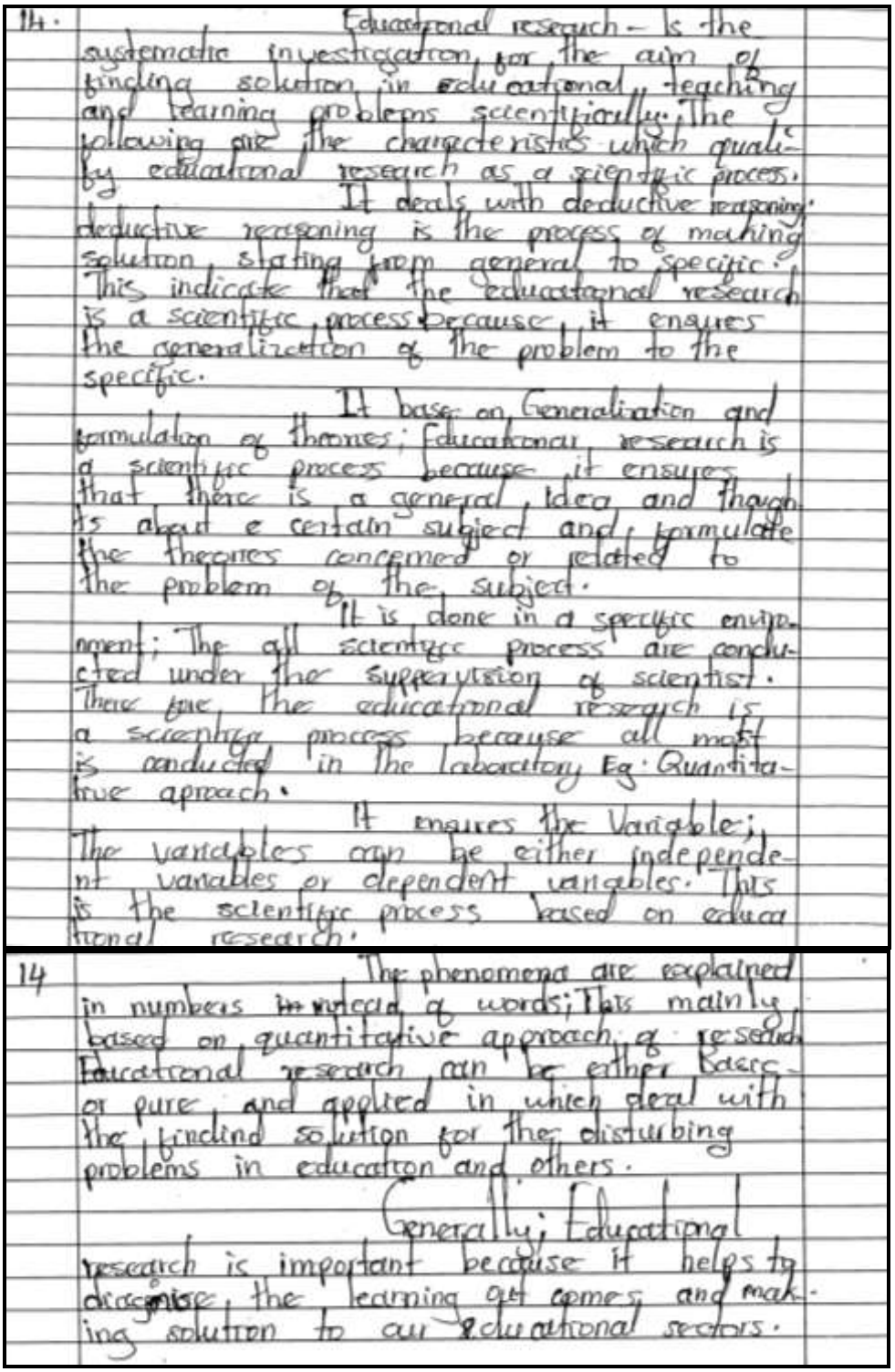
Figure 14: *The Candidates' Performance on Question 14*

The data in Figure 14 shows that 299 (57.7%) candidates scored from 6 to 10 marks, 209 (40.3%) candidates scored from 0 to 5.5 marks and the rest 10 (1.9%) candidates scored from 10.5 to 15 marks.

The candidates, who scored average (6 - 10 marks) had inadequate knowledge on characteristics which qualify educational research as a scientific process. The candidates also failed to provide good introduction, provided weak points as well as their conclusions were not relevant to the points that they had explained. They also provided fewer points out of the required points. Their responses were such as: *(i) educational research involves scientific procedures- in order to conduct educational research you should observe stages like problem identification, review the literature, decide on methodology, collecting data, analysing data, and draw conclusion; (ii) it should aim at solving educational problems; (iii) educational research should be applicable- it should provide useful information in solving educational problems e.g why mass failure of form iv ward secondary schools so as to improve educational programmes; (iv) educational research should be experimentally testable-the findings*

gathered from the field of education should be tested its validity;(v) educational research is cumulative that is bases from one generation to the next generation hence it focuses on the realities. These combined responses from candidates justify that they had insufficient knowledge about educational research though the last point could not fit from what the question needed.

Moreover, the analysis indicates that 40.3 per cent of the candidates who scored low marks (0 - 5.5) mixed correct and incorrect explanation of five characteristics which qualify educational research as a scientific process. For example, one candidate wrote: *(i) educational research deals with deductive reasoning by which the solution is made from general to specific; (ii) educational research involves hypothesis by which assumptions of the data being collected are to be rejected or accepted so as to get the truth; (iii) a good quality of educational research should be brief to a certain issues of education challenges; (iv) educational research should have a title or topic for making research; (v) any educational research should be clear by making research easier to find the problem or solutions.* The responses show that many of the candidates in this category lacked skills and knowledge of the characteristics which qualify educational research as a scientific process. Extract 14.2 presents weak responses from one of the candidates in this question.


14. Educational research - is the systematic investigation, for the aim of finding solution in educational, teaching and learning problems scientifically. The following are the characteristics which qualify educational research as a scientific process.

It deals with deductive reasoning, deductive reasoning is the process of making solution, stating from general to specific. This indicate that the educational research is a scientific process because, it ensures the generalization of the problem to the specific.

It base on Generalization and formulation of theories; Educational research is a scientific process because it ensures that there is a general idea and though. Is about a certain subject and formulate the theories concerned or related to the problem of the subject.

It is done in a specific environment; The all scientific process are conducted under the supervision of scientist. There fore the educational research is a scientific process because all most is conducted in the laboratory Eg. Quantitative aproach.

It ensures the Variable; The variables can be either independent variables or dependent variables. This is the scientific process based on educational research.

14 The phenomena are explained in numbers instead of words; This mainly based on quantitative approach of research. Educational research can be either basic or pure and applied in which deal with the finding solution for the disturbing problems in education and others.

Generally; Educational research is important because it helps to recognise the learning outcomes and making solution to our educational sectors.

Extract 14.2: A sample of response by a candidate who failed to provide correct characteristics which qualifying educational research as a scientific process in Question 14.

Further analysis indicates that 10 (1.9%) candidates who scored 10.5 to 15 marks answered the question correctly, and some of their points were relatively well presented. This justifies that the candidates had adequate knowledge of the topic from which the question was set. They satisfactorily identified and explained five characteristics which qualify educational research as a scientific process. Moreover, candidates of this category provided relevant introduction and satisfactory conclusion. For example, one candidate defined *Educational research as the systematic and scientific principles and procedures in collecting, analysing and interpreting data concerning educational issues*. Other candidates wrote valuable definitions as: *educational research is the systematic aiming at finding solution of educational issues scientifically; educational research is the process of identifying the problem and finally find solution for the improvement of education programmes*; *educational research refers to the systematic process of collecting, analysing, organising and interpreting data for the particular purpose.*

Additionally, in the main body part of the candidates' responses had variations of points that qualify educational research as a scientific process and their points were: *(i) educational research should be verifiable by which the information gathered have to be verified before drawing conclusion; (ii) educational research has to be reliable by yielding similar results over different time when similar procedures have to be observed; (iii) educational research should adhere the empirical information by using senses in collecting primary as well as secondary data from the field;(iv) educational research should be characterized with generalizability by drawing conclusion of the universe basing on the findings; (iv) educational research should be systematic since the process does employ scientific stages before drawing conclusion.* However, some candidates had contrary valuable points in addition from those already identified by others, these points were: *since different methods, techniques, approaches and experiments are used, the findings should be objective; and experimentation where by some educational researches should undergo experiments i.e. the effects of alcohol on academic performance.* Extract 14.1 illustrates a sample from a candidate who performed well.

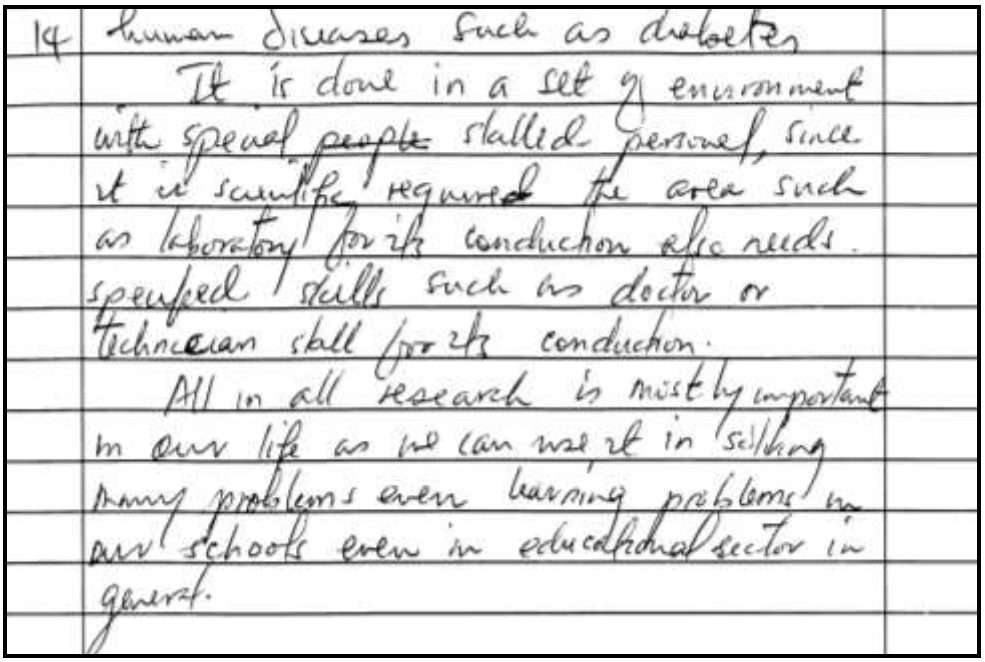
14 Research is systematic process of collecting, analysing and interpreting data in order to obtain the scientific solution for the existing phenomenon. It is done daily in our life whereby we face various problems. The following are the qualifications of a research as a scientific study!

It is systematic, this is because it has its defined procedures to follow step by step so as for a researcher to obtain the intended goal, those procedures are like problem identification, literature review, methodology, data collection and conclusion (TIEDA)

It is dealing with empirical data here by for the solution to be obtained there must be an application of data obtained and tested during a research process. Therefore it is generally stated as "it put theories into practices".

It is universal, if a research has been conducted in one part of the world and obtained the solution, that solution become accepted world wide, forexample medicines for various diseases and their vaccines.

It allows replication of data, as after the research has been conducted the obtained data can be modified and become used in many other places and sectors, forexample genetic engineering is nowadays used for plants and agricultural activities though firstly was done to cure


14 human diseases such as diabetes
It is done in a lab of environment with special people skilled personel, since it is scientific required the area such as laboratory for its conduction also needs specified skills such as doctor or technician skill for its conduction.
All in all research is most by important in our life as we can use it in solving many problems even learning problems in our schools even in educational sector in general.

Extract 14.1: A sample of response by a candidate who responded to Question 14 correctly.

2.2.5 Question 15: Test Constructions

This question required the candidates to evaluate three strengths and two weaknesses of using multiple choice questions in assessing students' achievement.

The question was attempted by 1,416 (68%) candidates. The overall performance of the candidates was good, as 96 per cent passed the question by scoring from 6.0 to 15.0 marks. The performance is illustrated in Figure 15.

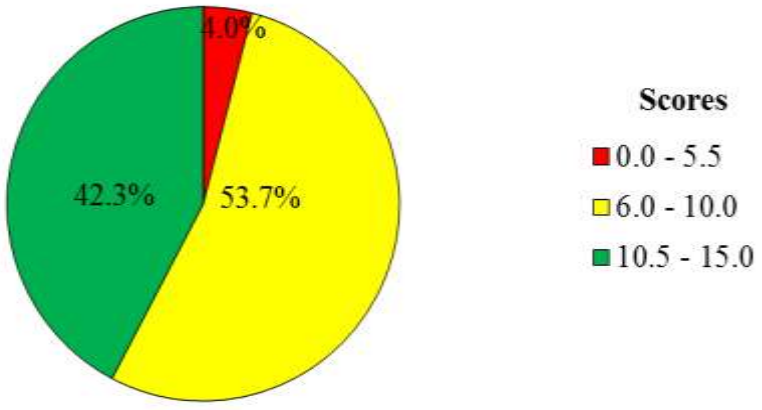


Figure 15: *The Candidates' Performance in Question 15*

Figure 15 indicates that the scores of candidates are as follows: 760 (53.7 %) candidates scored from 6 to 10 marks, 599 (42.3%) candidates scored from 10.5 to 15 marks, and the rest 57 (4%) candidates scored from 0 to 5.5 marks.

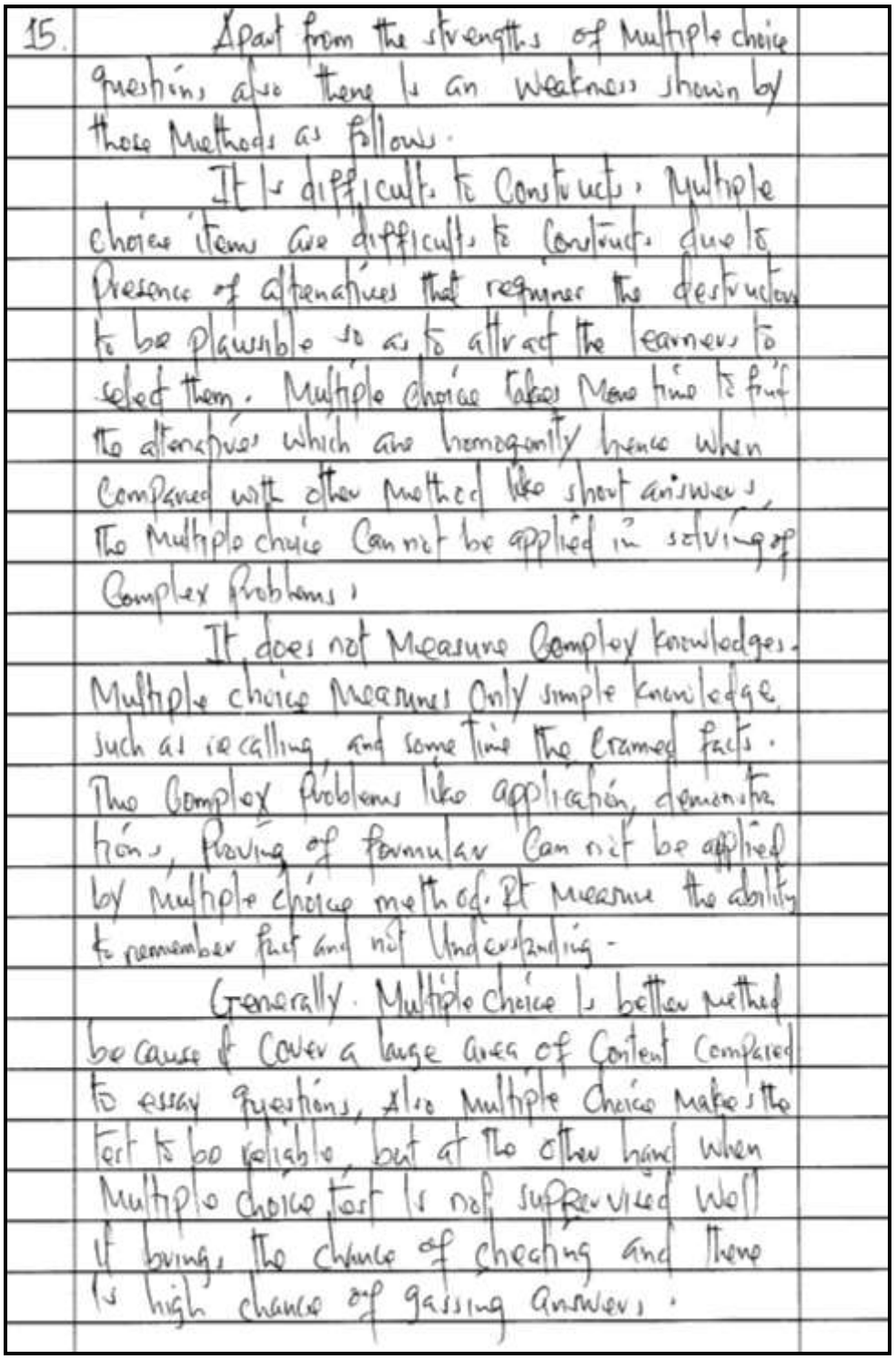
The analysis shows that 599 (42.3%) candidates who scored from 10.5 to 15 marks, showed good performance and demonstrated good organization of points, also they correctly evaluated the three strengths and two weaknesses of using multiple choice question when assessing student's achievement as the demand of the question. These candidates provided relevant introduction and conclusion in their responses. Their responses on strengths were such as: *(i) multiple choice questions are both cover the broad competencies; (ii) they are all easy to mark and take little time to complete; (iii) they are highly structured and measure simple recalling of information and facts.* Others wrote that: *(i) multiple choice questions can assess both simple and complex learning outcomes; (ii) some incorrect alternatives provide for diagnostic information; (iii) multiple choice questions cover large portion of the contents where many questions can be asked*. Likewise, in another part of the question, candidates were also able to evaluate two weaknesses of using multiple choice questions as follows: *(i) sometime it is hardly to find equally plausible destructors in multiple choice questions; (ii) constructing multiple choice questions is very difficult while techniques and skills are highly needed for effective question.* Some candidates provided other weaknesses as: *(i) it is time consuming composing good multiple choice questions; (ii) multiple choice question are not good in measuring problem solving, candidates' writing skills as well as expressive behaviour*. These varied valuable responses justify the possession of sufficient knowledge on multiple choice items. Extract 15.1 presents a sample of good responses from one of the candidates.

15. Multiple choice items. This are types of selection items which is most selection item than other Method, Multiple choice is most preferred in Constructing of test Compared to other selections items like Matching item and true-false items. There is an objective test which have Only one Predetermined responses. There is high reliability in the Multiple Choice items. The following, are the strengths of Multiplechoice questions in assessing student's achievements.

It have high reliability. Due to the Presence of only one predetermined answer, there is high chance of obtaining the test reliability. Also this leads to the reduction of bias during the test scoring. The teacher will reduce the degree of subjectivity and brings fair to all students.

It is easy to score. Due to availability of the pre-determined responses, the Multipechoice becomes easy in scoring Compared to the essay items which have no predetermined correct responses. The Multiple Choice makes the substitution of scoring that Can be done by other teacher to be more simple, hence the time taken to these kind of selection item in scoring will be at a minimum level.

It takes little time in scoring, The Multiple choice questions takes a little time in scoring that facilitate easy Interpretation of the test score, so the teachers is able to provide the results at a time planned. Compared to subjective test item which takes much time on Marking the Interpretation of the student results will be slow and hence the feedback will come late.


15. Apart from the strengths of Multiple choice questions also there is an weakness shown by those Methods as follows.

It is difficult to Constructs, Multiple choices items are difficults to Construct due to Presence of alternatives that requiree the destructors to be plausible so as to attract the learners to select them. Multiple choice takes More time to find the alternatives which are homogeneity hence when Compared with other Method like short answers, The Multiple choice Can not be applied in solving of Complex Problems.

It does not Measure Complex knowledges. Multiple choice Measures Only simple knowledge such as recalling and some time the learned facts. The Complex Problems like application, demonstration, Proving of formular Can not be applied by Multiple choice method. It measure the ability to remember fact and not Understanding.

Generally. Multiple choice is better method because it Cover a large area of Content Compared to essay questions, Also multiple Choice Makes the test to be reliable, but at the other hand when Multiple choice test is not supervised well it brings the chance of cheating and there is high chance of gassing answers.

Extract 15.1: A sample of response by a candidate who evaluated strengths and weaknesses of multiple choice questions correctly in question 15.

Further analysis shows that 760 (53.7%) candidates who scored from 6 to 10 marks had average understanding of the three strengths and two weaknesses of multiple choice questions and as a result, they could not score above 10 marks. Some of these candidates were able to evaluate either strengths of multiple choice questions correctly or weaknesses of multiple choice questions incorrectly or their vice versa. Also, some candidates in this group could not provide sufficient explanations of the strengths and weaknesses of using multiple choice questions. For example, one candidate wrote correctly the strengths and weaknesses as: strengths- *(i) they are simple to mark; (ii) they measure simple learning outcomes; (iii) they are simple in scoring.* Weaknesses- *(i) they are difficult to construct; (ii) they encourage cheating or guessing the answers since candidates have alternatives if candidates were not well prepared.*

In addition, the analysis indicates that 57 (4%) candidates who scored poorly from 0 to 5.5 marks had insufficient knowledge on strengths and weaknesses of using multiple choice questions. Most of them started well with good introduction but provide irrelevant responses and conclusion. For instance, one candidate gave the following wrong strengths and weaknesses as: Strengths*:-(i) it reduces language expertise in the form of hand writing and paragraph; (ii) it promotes the economy of time use time to mark rather than essay type which use much time; (iii) it ensures validity of the test.* Weaknesses: - *(i) problem of guessing by concentrating much on the guessing and copying each and every thing from students; (ii) cheating opportunities due to the distribution of materials.* As per the responses, it can be noted down that this candidate provided explanation as well as irrelevant points. Extract 15.2 is a sample of poor responses from one of the candidates.

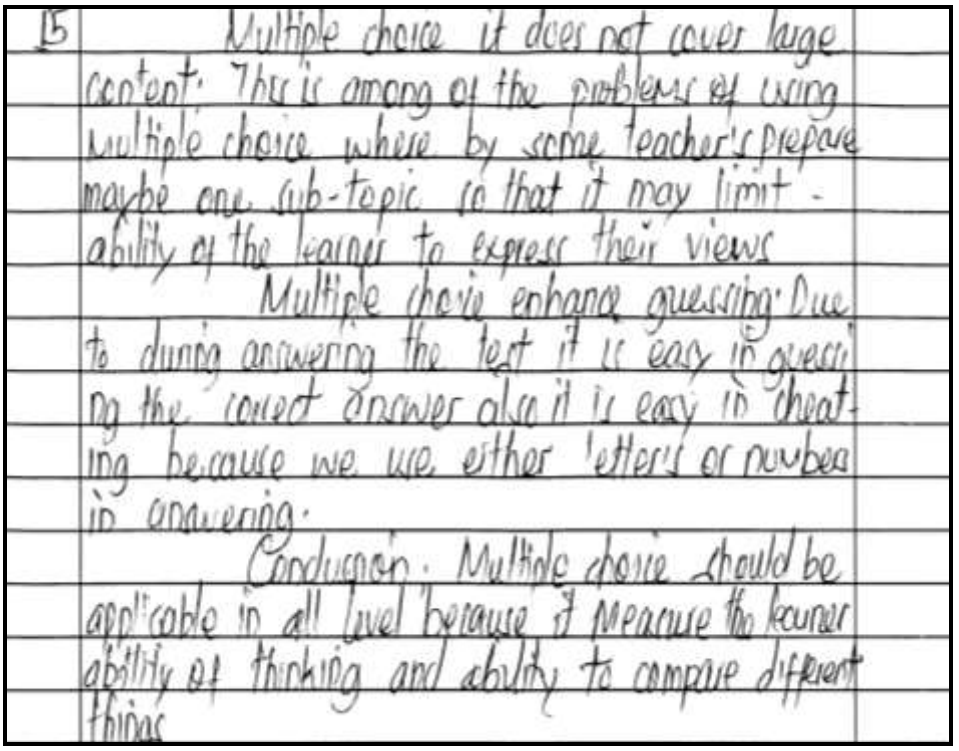
15 Multiple choice : Is a two way chart which contain questions in one side listed LIST A and answers in another side listed LIST B. This type of test items questions should collocate with their answers and teacher should avoid long sentences and the use of ambiguous language. The following are the strength of multiple choice questions in assessing student's achievement

Multiple choice it save time and energy. This is due to during test assembling and during scoring it take a short time because it is labelled either in letters or numbers.

Multiple choice create critical thinking to the learner, because during answering it need a ~~leade~~ learner to pay attention and to think the correct answer for a certain question

Multiple choice it help in making comparison. Due to the fact that because in answering the question a learner must compare the section A and section B so that it make a learner to be competent in making comparison of different things

The following are the weaknesses of using multiple choice in assessing student's achievement


15 Multiple choice it does not cover large content. This is among of the problems of using multiple choice where by some teacher's prepare maybe one sub-topic so that it may limit ability of the learner to express their views
Multiple choice enhance guessing. Due to during answering the test it is easy in guessing the correct answer also it is easy to cheating because we use either letters or numbers in answering.
Conclusion. Multiple choice should be applicable in all level because it measure the learner ability of thinking and ability to compare different things

Extract 15.3: A responses by a candidate who failed to evaluate the strengths and weaknesses of multiple choice questions in Question 15.

2.2.6 Question 16: Educational Measurement

The question required the candidates to examine five factors that affect test reliability.

A total of 1,611 candidates, equivalent to 76.9 per cent attempted the question. The general candidates' performance was good since 90.1 per cent scored from 6 to 15 marks. Figure 16 illustrates the performance of candidates.

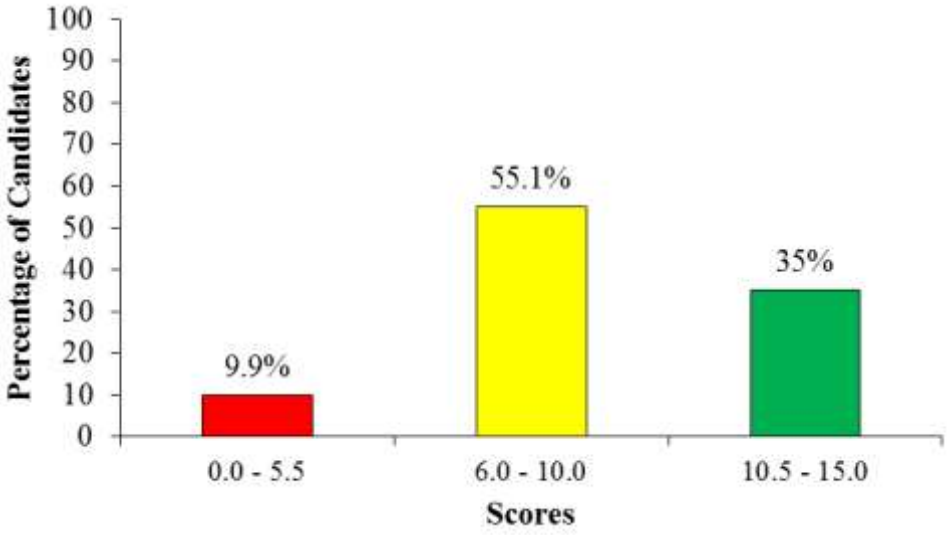


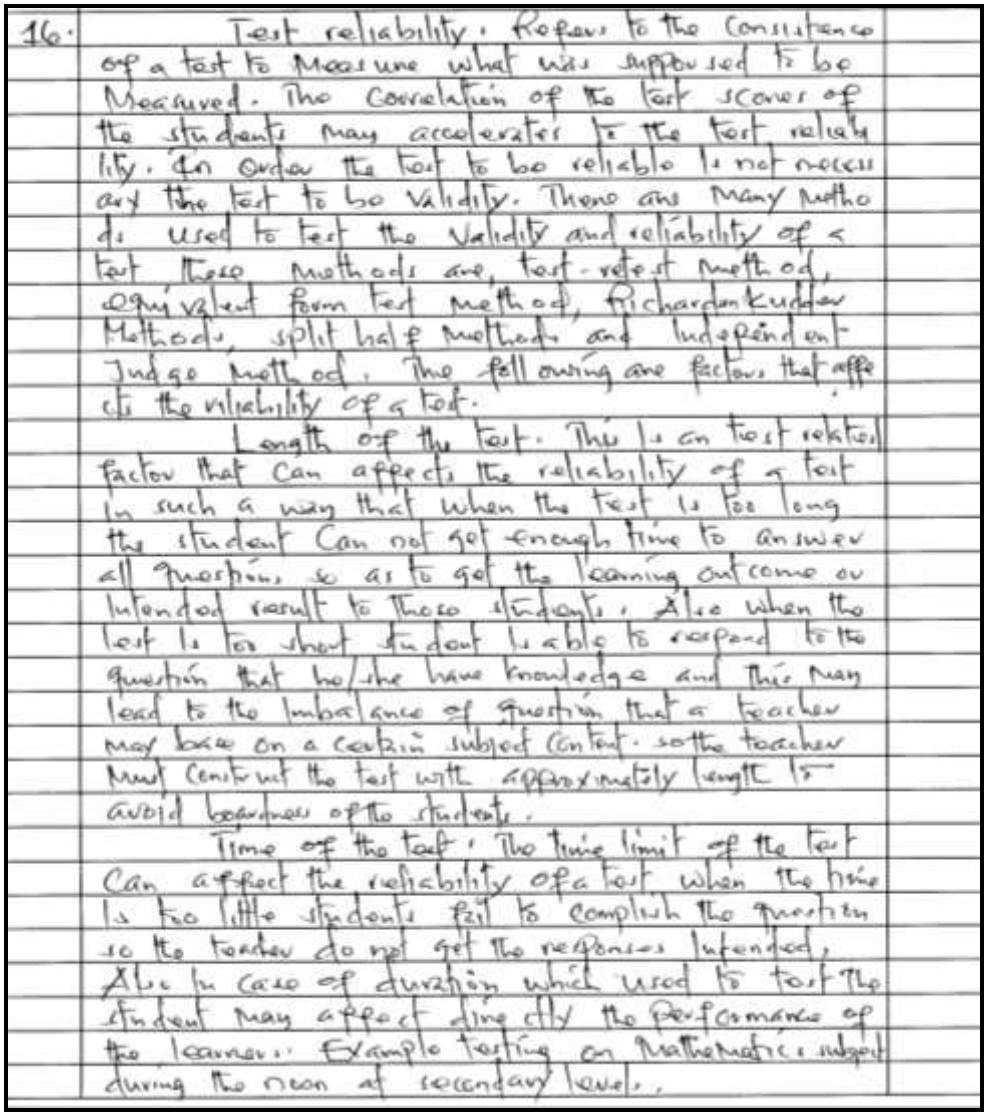
Figure 16: *The Candidates' Performance in Question 16*

Figure 13 indicates that among the candidates who attempted this question; 887 (55.1%) candidates scored from 6 to 10 marks, 565 (35%) candidates from scored 10.5 to 15 marks and 159 (9.9%) candidates from scored 0 to 5.5 marks.

The analysis reveals that 565 (35%) candidates, who scored high marks (10.5 - 14), had knowledge of the factors that affect test reliability. They provided coherence organization of introductory part of the question to the end point of conclusion. They clearly defined the concept of reliability as: *the consistency in measurement in which the test managed to give the same results over time*. In the main body part, candidates gave the factors which affect test reliability as: *(i) time allocated for the test- if time is too short compared to number of items many students will find it difficult to finish the test, therefore the reliability will be affected; (ii) nature of the students tested- low reliability of the results from the test can be detected of there is a big range between higher and lower achievers; (iii) difficult level of each item in the test may affect test reliability where each item should be of appropriate level of difficulty; (iv) length of the test- if test has huge number of items contrary to the time set, the results will not be reliable simply because some students will not manage to finish them.*

Another candidate wrote that: *(i) familiarity of the testee with the test- this is the extent to which the examinee has been acquainted with that particular test; (ii) spread of scores- this is the way scores are divided*

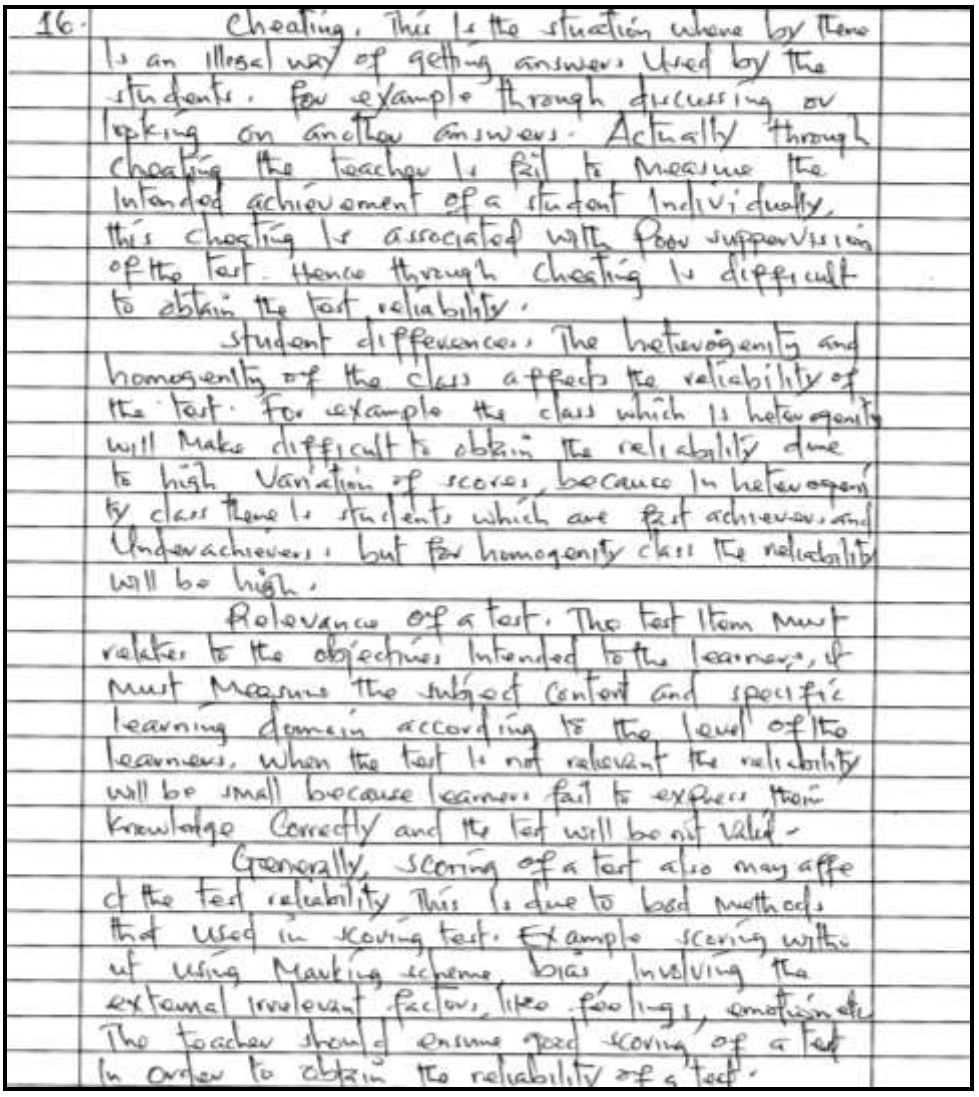
throughout the test where by a good spread of score result into high consistency and the vice versa; (iii) language difficult-the grammatical words used by the examiners sometimes confuse the examinees by not understanding what real the question require; (iv) irrelevancy / invalidity of test items the test will be irrelevant if it does not measure the learning outcomes according to the instructional objectives of the course, hence the student will end up failing. Provision of these relevant factors that may affect test reliability justify that candidates understood the demands of the question. Extract 16.1 is a sample of good responses from one of the candidates.



16. Test reliability: Refers to the consistence of a test to Measure what was supposed to be Measured. The Correlation of the test scores of the students may accelerates the test reliability. In order the test to be reliable is not necessary the test to be validity. There are many methods used to test the validity and reliability of a test these methods are, test-retest method, equivalent form test method, Richardson Kudder Methods, split half methods and independent-Judge method. The following are factors that affects the reliability of a test.

Length of the test. This is an test related factor that can affects the reliability of a test in such a way that when the test is too long the student can not get enough time to answer all questions so as to get the learning outcome or intended result to these students. Also when the test is too short student is able to respond to the question that he/she have knowledge and this may lead to the imbalance of question that a teacher may base on a certain subject content. so the teacher must construct the test with approximately length to avoid boardness of the students.

Time of the test: The time limit of the test can affect the reliability of a test when the time is too little students fail to complish the question so the teacher do not get the responses intended, Also in case of duration which used to test the student may affect directly the performance of the learners. Example testing on Mathematics subject during the mean of secondary level.



16. Cheating. This is the situation where by there is an illegal way of getting answers. Used by the students. For example through discussing or looking on another answers. Actually through cheating the teacher is fail to measure the intended achievement of a student individually, this cheating is associated with poor supervision of the test. Hence through cheating is difficult to obtain the test reliability.

Student differences. The heterogenity and homogenity of the class affects the reliability of the test. For example the class which is heterogenity will make difficult to obtain the reliability due to high variation of scores, because in heterogenity class there is students which are fast achievers and underachievers, but for homogenity class the reliability will be high.

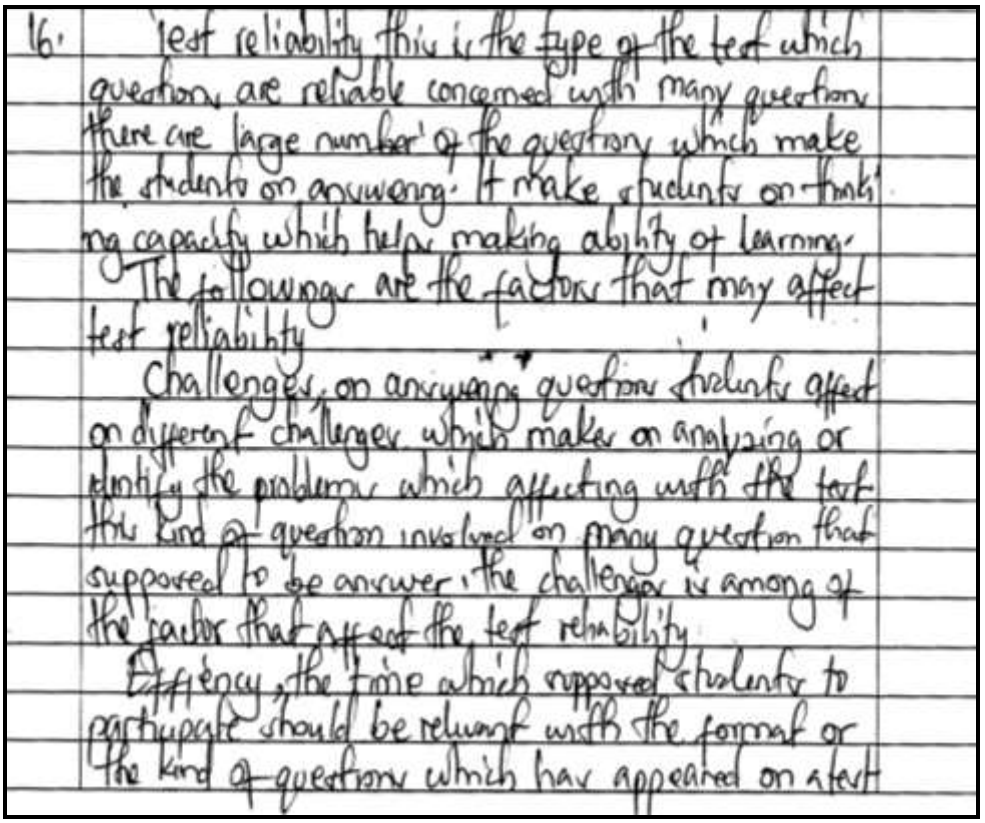
Relevance of a test. The test item must relates to the objectives intended to the learners, it must measure the subject content and specific learning domain according to the level of the learners. When the test is not relevant the reliability will be small because learners fail to express their knowledge correctly and the test will be not valid.

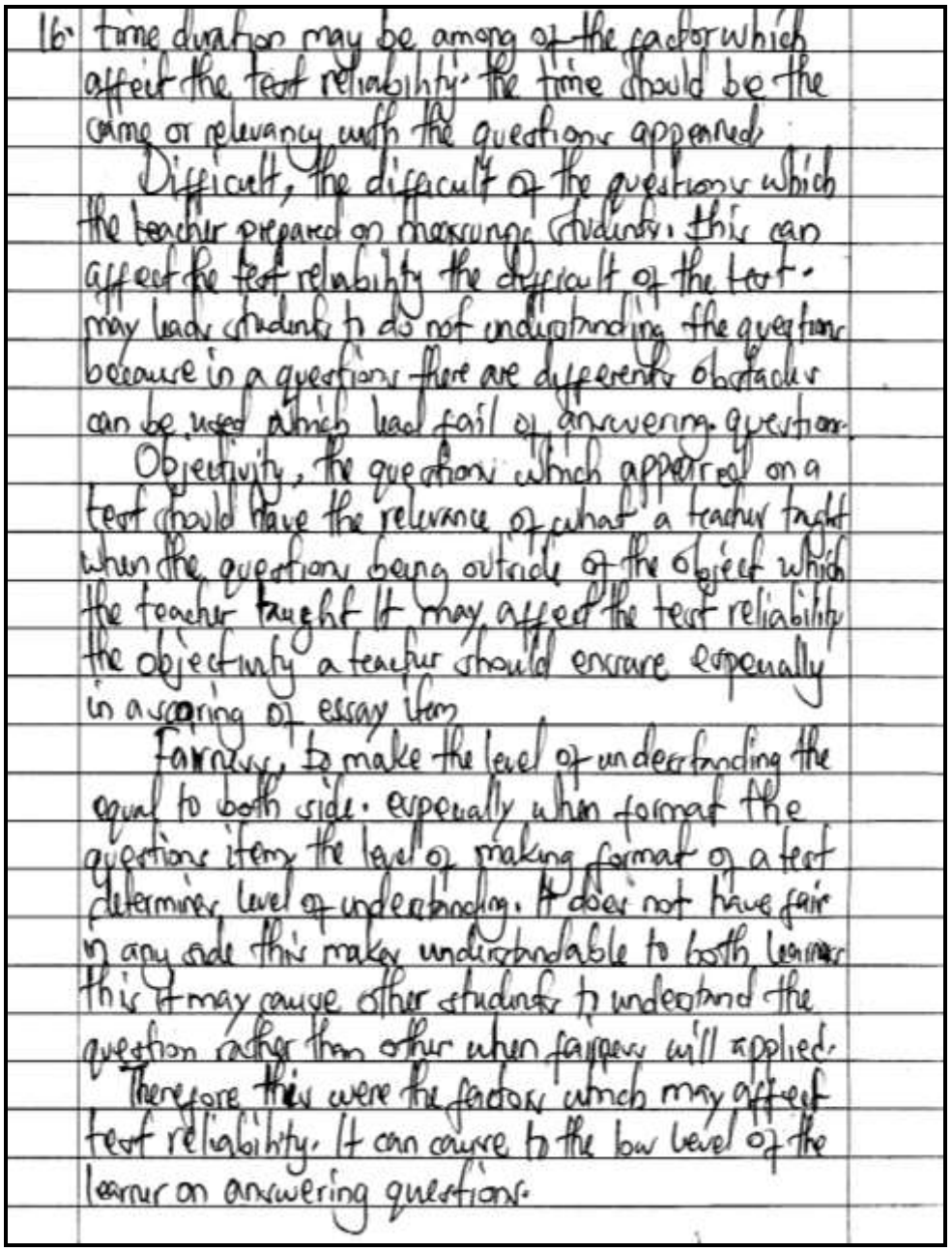
Generally, scoring of a test also may affect the test reliability this is due to bad methods that used in scoring test. Example scoring without using marking scheme, bias involving the external irrelevant factors, like feelings, emotions etc. The teacher should ensure good scoring of a test in order to obtain the reliability of a test.

Extract 16.1: A sample of response by a candidate who correctly identified factors that affect test reliability in Question 16.

Further analysis reveals that 887 (55.1%) candidates who scored averagely, from 6 to 10 marks provided good introduction, conclusion and correct points but their responses were characterized by partial explanations. Others presented several points mixing up relevant and irrelevant points. Some candidates highlighted the points without thorough explanations as they wrote: *(i) length of the test; (ii) the use of ambiguous statements; (iii) the chance of cheating; and (iv) the difficultness of test items,* while others wrote fewer points contrarily to the question requirements. This failure to explain the points justifies that they had insufficient knowledge of the subject matter of the question.

Moreover, the analysis of candidates' responses shows that 159(9.9%) candidates who scored poorly, from 0 to 5.5 marks, examined incorrect factors that affect test reliability though some were able to provide relevant introduction on the concept of reliability. For instance, one candidate wrote the factors as: *(i) high scoring: - students may score high marks because have already discussed the examination; (ii) motivation: - motivation is low due to students to lose confidence of doing the exams; (iii) test validity: - also if you make test to be reliability there is no test validity because the students have already done; (iv) similar answers: - the answer that we get after the first test are the same to the answers that provided in the second test.* Other candidates also identified wrong factors as: *(i) school time table: - this can affect test reliability in that, if at school there is different time tables, i.e sports and games, cleanliness it may cause students to fail; (ii) through objective of tests; (iii) different levels of the learners in learning; (iv) unplanned contents: - if the test involves the topics which have been omitted from the syllabus may affect the reliability.* These are irrelevant factors from the candidates in this group. Extract 16.2 shows a sample of weak responses from one of the candidates in the question.


16. Test reliability this is the type of the test which
questions are reliable concerned with many questions
there are large number of the questions which make
the students on answering. It make students on think-
ing capacity which help making ability of learning.
The followings are the factors that may affect
test reliability
Challenges, on answering questions students affect
on different challenges which make on analysing or
identify the problems which affecting with the test
this kind of question involved on many question that
supposed to be answer. The challenges is among of
the factor that affect the test reliability
Efficiency, the time which supposed students to
participate should be relevant with the format or
the kind of questions which have appeared on a test


16. Time duration may be among of the factor which affect the test reliability. The time should be the same or relevancy with the questions appeared.

Difficult, the difficult of the questions which the teacher prepared on measuring students. this can affect the test reliability the difficult of the test. may leads students to do not understanding the questions because in a questions there are different obstacles can be used which lead fail of answering questions.

Objectivity, the questions which appeared on a test should have the relevance of what a teacher taught when the questions being outside of the object which the teacher taught it may affect the test reliability the objectivity a teacher should ensure especially in a scoring of essay item

Fairness, to make the level of understanding the equal to both side. especially when format the questions items the level of making format of a test determines level of understanding. It does not have fair in any side this makes understandable to both learners this it may cause other students to understand the question rather than other when fairness will applied.

Therefore, there were the factors which may affect test reliability. It can cause to the low level of the learner on answering questions.

Extract 16.2: A sample of response by a candidate who failed to identify factors that affect test reliability in Question 16.

3.0 ANALYSIS OF THE CANDIDATES' PERFOMANCE IN EACH TOPIC

The 2021 DSEE Educational Research, Measurement and Evaluation examination had seven topics from which the examination questions were set. The analysis of the candidates' performance in each topic shows that the candidates had good performance in three (3) topics; *Educational Measurement* (98.4%), *Qualities of Tests* (90.1%) and *Test Construction* (70.65%). The topic, *Analysis and Interpretation of Test Results* (61.3%) had average performance. The reasons for good performance were sufficient knowledge and skills on items that involved detailed explanation and competency on numerical manipulations.

Further analysis shows that weak performance was attained in three topics; *Educational Assessment and Evaluation* (39.2%), *Educational Research* (38.2%) and *Assessing Achievement* (10.6%). It was noted that, the main reasons were inadequate knowledge and failure to understand the requirements of the question. Appendix I summarizes the candidates' performance in each topic.

The comparison of performance for 2020 and 2021 reveals that there were decline of performance for three topics: *Educational Research* (47.99%), *Educational Assessment and Evaluation* (59.68%), and *Assessing Achievement* (76.9%) in 2020, where in 2021 the performance was 38.2%, 39.2% and 10.6% respectively as shown in Appendix II.

4.0 CONCLUSION

The performance in Educational Research Measurement and Evaluation subject for the Diploma in Secondary Education Examination (DSEE) in 2021 was good, as 97.97 per cent of candidates passed. The analysis shows that the candidate's good performance was due to their good abilities to identify the demands of questions, sufficient knowledge of the subject matter, proficiency in the English Language, as well as computational skills. Only a few candidates showed lack of such qualities; which earned them low marks.

However, it was evidently observed from the analysis of candidates' item response that the performance in questions which involve numeric (Question 11 and 12) for calculating central tendency and computing

difficult index respectively were still a challenge, as majority of candidates had weak performance.

5.0 RECOMMENDATIONS

In order to improve the performance of the prospective candidates in this subject, the following are recommended:

(a) Deliberate initiatives need to be taken in the topics of Assessing Achievement, Educational Research, and Educational Assessment and Evaluation. A topic such as Analysis and Interpretation of Test Results should be taught through demonstration, group discussion, gallery walk and brainstorming. This topic requires mastery of mathematical operations while the topics, Assessing Achievement and Educational Research should be taught through group discussion, classroom discussion, jigsaw and brainstorming.

(b) Candidates must be encouraged to read the questions carefully before attempting them so as to understand the requirements of the questions.

Appendix I

THE 2020 AND 2021 TOPIC-WISE SUMMARY OF THE CANDIDATES' PERFORMANCE IN 762 EDUCATIONAL RESEARCH, MEASUREMENT AND EVALUATION

S/N	Topic	Question number	Performance in each question (%)	Average performance per topic (%)	Remarks
1.	Educational Measurement	13	98.4	98.4	Good
2.	Qualities of Tests	16	90.1	90.1	Good
3.	Test Construction	1	85.7	70.65	Good
		2	78.3		
		3	25.8		
		8	83.3		
		10	54.8		
		15	96		
4.	Analysis and Interpretation of Test Results	11	48.6	61.3	Average
		12	74		
5.	Educational Assessment and Evaluation	7	39.2	39.2	Weak
6.	Educational Research	5	19.8	38.2	Weak
		6	1.6		
		9	71.8		
		14	59.6		
7.	Assessing Achievement	4	10.6	10.6	Weak

Appendix II

COMPARISON OF THE CANDIDATES' PERFORMANCE IN 762 EDUCATIONAL RESEARCH MEASUREMENT AND EVALUATION SUBJECT (DSEE 2021)

		2020			2021		
S/N	**Topic**	**Question number**	**Average performance of the Candidates per topic (%)**	**Remarks**	**Question number**	**Average performance of the Candidates per topic (%)**	**Remarks**
1.	Educational Measurement	1	93.31	Good	13	98.4	Good
2.	Qualities of Tests	13	90.00	Good	16	90.1	Good
3.	Test Construction	8, 12 &15	69.67	Average	1, 2, 3, 8, 10& 15	70.65	Good
4.	Analysis and Interpretation of Test Results	6, 9, 11 &16	50.41	Average	11 & 12	61.3	Average
5.	Educational Research	2, 4 &5	47.99	Average	7	38.2	Weak
6.	Educational Assessment and Evaluation	3, 10 & 14	59.68	Average	5, 6, 9 &14	39.2	Weak
7.	Assessing Achievement	7	76.90	Good	4	10.6	Weak