

THE NATIONAL EXAMINATIONS COUNCIL OF TANZANIA



**CANDIDATES' ITEM RESPONSE ANALYSIS REPORT FOR
DIPLOMA IN SECONDARY EDUCATION EXAMINATION
(DSEE) 2018**

**762 EDUCATIONAL RESEARCH, MEASUREMENT
AND EVALUATION**

THE NATIONAL EXAMINATIONS COUNCIL OF TANZANIA



**CANDIDATES' ITEM RESPONSE ANALYSIS
REPORT FOR THE DIPLOMA IN SECONDARY
EDUCATION EXAMINATION (DSEE) 2018**

**762 EDUCATIONAL RESEARCH, MEASUREMENT
AND EVALUATION**

Published by

The National Examinations Council of Tanzania,

P.O. Box 2624

Dar es Salaam, Tanzania.

(C) The National Examinations Council of Tanzania, 2018

All rights reserved

TABLE OF CONTENTS

FOREWORD.....	iv
1.0 INTRODUCTION	1
2.0 Analysis of The Candidates' Performance in Each Question.....	4
2.1 SECTION A: Short Answers Questions.....	4
2.1.1 QUESTION 1: Educational Measurement.....	4
2.1.2 QUESTION 2: Educational Measurement.....	7
2.1.3 QUESTION 3: Educational Research.....	10
2.1.4 QUESTION 4: Assessing Achievement.	14
2.1.5 QUESTION 5: Analysis and Interpretations of the Test Results.....	17
2.1.6 QUESTION 6: Assessing Achievement.....	20
2.1.7 QUESTION 7: Assessing Achievement.....	23
2.1.8 QUESTION 8: Test Construction	26
2.1.9 QUESTION 9: Qualities of Tests.....	29
2.1.10 QUESTION 10: Analysis and Interpretations of the Test Results ..	31
2.2 SECTION B: Essay Questions	35
2.2.1 QUESTION 11: Analysis and Interpretations of the Test Results ..	35
2.2.2 QUESTION 12: Test Construction.....	40
2.2.3 QUESTION 13: Educational Assessment and Evaluation	44
2.2.4 QUESTION 14: Qualities of Tests	49
2.2.5 QUESTION 15: Analysis and Interpretations of the Test Results ..	53
2.2.6 QUESTION 16: Educational Measurement.....	58
3.0 PERFORMANCE OF CANDIDATES IN EACH TOPIC	62
4.0 CONCLUSION.....	62
5.0 RECOMMENDATIONS	62
APPENDIX A	64

FOREWORD

The National Examination Council of Tanzania is pleased to issue this report on Item Response Analysis for the 2018 Diploma in Secondary Education Examination (DSEE) in Educational Research, Measurement and Evaluation subject. The report provides feedback to students, 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 two years of diploma level in education. It is a summative evaluation which shows the effectiveness of the education system in general and education system in particular. Basically, the report indicates what the education system was able or unable to offer to the students in their two years of the Diploma in Secondary Education.

In this report, factors which enabled the candidates to answer questions correctly or incorrectly have been analysed. The analysis shows that the candidates with higher scores understood the demands of questions; and had basic knowledge of the subject matter. However, the candidates with lower scores depicted contrary attributes.

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

The National Examinations Council of Tanzania will highly appreciate comments and suggestions from tutors, students and the general public that can be useful for improving future items responses analysis reports. Finally, the Council is quite grateful to all stakeholders who provided valuable assistance during preparation of this report.



Dr. Charles E. Msonde
EXECUTIVE SECRETARY

1.0 INTRODUCTION

This report analyses the performance of the candidates who sat for the Diploma in Secondary Education Examination (DSEE) in Educational Research, Measurement and Evaluation subject in 2018. A total of 2,181 candidates sat for the examination, out of which 1,274 candidates were using University of Dodoma (UDOM) curriculum and 907 were using the Tanzania Institute of Education (TIE) curriculum. The examination tested the candidates' competences in using different scales of measurement in carrying out assessment and evaluation of performance, 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. The general performance of the candidates was good as Table 1 shows.

Table 1: Performance of Candidates in Educational Research Measurement and Evaluation

Candidates Type	Sat	Number of Candidates and Percentage					
		Passed	Grades				
			A	B	C	D	F
ALL (DSEE)	2,181	2,169 99.50	11 0.50	211 9.68	1,433 65.73	514 23.58	11 0.50
UDOM Curriculum	1,274	1,268 99.53	0 0.00	57 4.47	854 67.03	357 28.02	6 0.47
TIE Curriculum	907	901 99.45	11 1.21	154 17.00	579 63.91	157 17.33	5 0.55

Table 1 shows that 901 candidates (99.45%) under TIE curriculum passed the examination while 99.53% of the candidates under the UDOM curriculum passed. However, there was no candidate under UDOM curriculum who passed at 'A' grade.

Since the assessment for the candidates who are pursuing DSEE using UDOM curriculum is in transition in this report, the detailed analysis was done on the performance in individual examination questions and topics based on the candidates who sat for examination using TIE curriculum only.

In the TIE Curriculum, the examination consisted of two sections, A and B, with a total of sixteen (16) questions. Section A had ten (10) short answer questions, each question carried four (4) marks; making a total of 40 marks. Section B had six (6) essay questions, from which candidates were required to attempt any four (4) questions. Each question carried 15 marks; making a total of 60 marks. The candidates were instructed to attempt a total of 14 questions in all sections. The questions were set from the following topics: Educational Measurement, Assessing Achievement, Qualities of Test, Test Construction, Analysis and Interpretation of the Test Results, Educational Assessment and Evaluation, and Educational Research.

In this report, the analysis of the questions is based on the category of the question items i.e. Short answer items in section A and essay type items in section B. In short answer items, the performance of the candidates 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 essay items the performance of the candidates 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

The performance in the examination has been categorised into five grade ranges. The performance of the candidates is regarded as *fail (F)* if the scores range from 0 to 39 marks, satisfactory (D) if the scores range from 40-54 marks, Good (C) if the scores range from 55-69 marks, Very good (B) if the scores range from 70-79 marks and Excellent (A) if the scores range from 80-100 marks as shown in figure 1.

Figure 1 shows the performance of Candidates who were using TIE Curriculum.

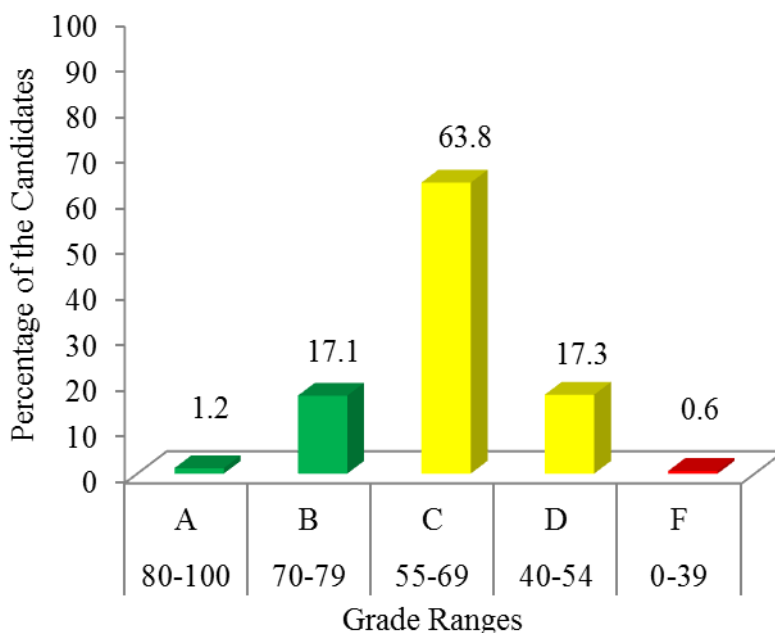


Figure 1 shows the performance of candidates with grades

The analysis shows that few candidates 1.2%, 17.1% passed with grade A and B respectively, while most of the candidates 63.8% passed with grade C, whereas 17.3% passed with grade D. The candidates who failed (grade F) 0.6%. The General Performance in this subject was good as 99.45% passed with grade A to D.

Samples of the candidates' answers have been attached to illustrate their responses. Similarly, Appendix A shows the candidates' performance in each topic. It is expected that the report will enable educational stakeholders, tutors and student teachers to improve the teaching and learning process in Educational Research, Measurement and Evaluation subject.

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 instructed to attempt all ten questions. The total scores allocated for the section was forty (40); each question carrying four (04) marks.

2.1.1 Question 1: Educational Measurement

This question was set from Educational Measurement topic. The question required the candidates to identify, with examples, the two types of Educational Measurement based on test interpretation. Total scores for the question were 4 marks.

The question was attempted by all 907 candidates (100%), of which, 298 candidates (32.9%) scored from 3 to 4 marks; indicating a good performance, 19 candidates (2.1%) scored 4 marks. Further more 539 candidates (59.4%) scored from 2 to 2.5 marks; indicating an average performance. However, 70 candidates (7.7%) scored from 0 to 1.5, of which, 59 candidates (6.5%) scored 0 mark; indicating a poor performance. Generally, the candidate performance in the question was good considering that 837 candidates (92.3%) scored from 2 to 4 marks. Figure 2 illustrates the performance of candidates in question 1.

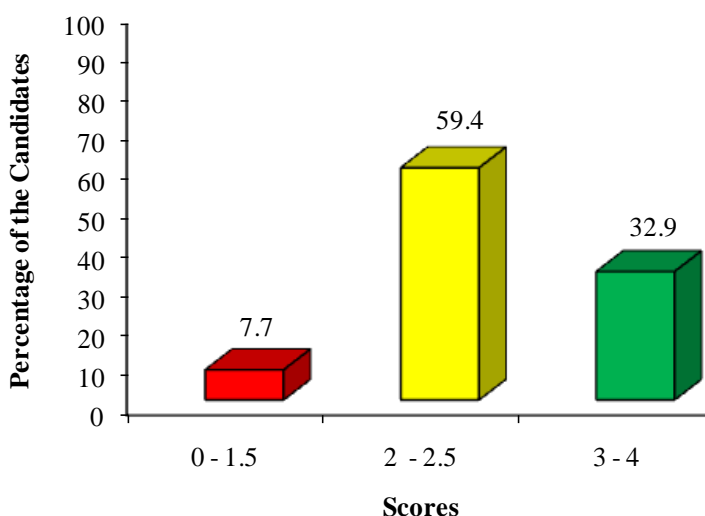


Figure 2: *The Candidates' Performance in Question 1*

The analysis shows that 298 candidates (32.9 %) who scored from 3 to 4 marks correctly identified the two types of educational measurements based on test interpretation and managed to meet the demand of the question as per the agreed standardized guide. They identified the types as *norm and criterion referenced measurement*. Also they managed to give correct examples of criterion referenced measurement like: *writing skills test, driving tests* while others wrote: *music tests, typing tests*. Other candidates identified and explained each type of Educational measurement based on test interpretation as: *norm-referenced measurement is used to indicate students' performance by comparing with other students* and gave the example like: *classroom tests such as weekly tests, terminal, annual examinations, and national examinations*. Moreover, some candidates defined criterion-referenced measurement as *the type of educational measurement used to indicate the individual performance without comparing to other students (used to assess what the student can do) thus, criterion referenced measurement is skills oriented*. This shows that these candidates had adequate knowledge and skills in the subject as it is shown in Extract 1.1.

Extract 1.1

1	The types of educational measurement based on test interpretation are	
	Norm referenced measurement, is the type of education measurement in which the performance of individual student are compared with the another students who took the same test. forexample	
	Mid term test	
	Criterion referenced measurement is the type of education measurement in which measure the achievement of learner to a particular programme without compare with other	
	Example performance of Vocational training.	

Extract 1.1 is a sample of responses of candidates who identified two types of educational measurement correctly as norm and criterion referenced measurement with an example for each type.

However, 539 candidates (59.4%) who scored from 2 to 2.5 marks identified the two types of educational measurement correctly, but failed to give correct examples for each. Some examples from such candidates' responses on norm referenced measurement were: *weekly tests, terminal tests, monthly test, mid-term test*, which are correct. However, many of them failed to give the correct examples of criterion-referenced measurement and they wrongly wrote examples of norm-referenced measurements such as *national examinations; quizzes*. This shows that they had insufficient knowledge and skills in the subject.

Further analysis revealed that 70 candidates (7.7%) who scored from 0 to 1.5 marks failed completely to respond correctly. Many of them wrote incorrect answers such as: *formative measurement and summative measurement*. Others wrote: *formative educational measurement and diagnostic educational measurement*. Some of them wrote types of

assessment instead of the types of educational measurement. Others defined what assessment is and what evaluation is. Some of them provided examples as: *science subject test* and *essay test* respectively implying that they had misconceived the types of educational measurement basing on test interpretation. Further more, candidates who scored from 1 to 1.5 marks (1.2%) failed to use correct examples after identifying and describing the correct types of Educational Measurement. Extract 1.2 is a sample of such a poor response from one of the candidates.

Extract.1.2

1.	1/. Objective test example short answers
	2/. Subjective test example essay questions

Extract 1.2 is a sample of poor responses from a candidate who failed to identify the types of educational measurement based on test interpretation as norm and criterion referenced measurement.

2.1.2 Question 2: Educational Measurement

This question consisted of two parts A and B; each carrying two (2) marks, composed from Educational Measurement topic. The question instructed the candidates to explain the concepts measurement in part (a) and educational measurement in part (b). Scores allocated for the question were 4 marks in total.

The data analysis shows that 907 candidates (100%) attempted the question. 639 candidates (70.5%) of them scored from 3 to 4 marks; indicating good performance, of which, 410 candidates (45.2%) scored 4 marks. 163 candidates (18%) scored from 2 to 2.5 marks indicating average performance and 105 candidates (11.6%) scored from 0 to 1.5 marks; indicating poor performance, of which, 45 candidates (5%) scored 0 mark. Generally, the performance in the question was good considering that 802 candidates (88.5%) scored from 2 to 4 marks as shown in figure 3.

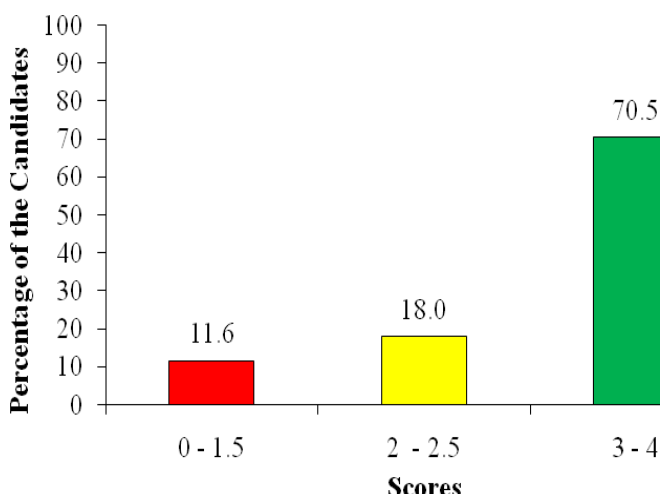


Figure 3: *The Candidates' Performance in Question 2*

Further, the analysis shows that 639 candidates (70.5%) who scored from 3 to 4 marks; good performance provided correct definition of measurement in part (a). For example, they defined measurement as: *the process of assigning numerical values to a particular phenomenon to describe the degree to which an individual possesses a particular attributes*. others wrote: *it is the process of obtaining or gathering numerical data to the extent that individual possesses a particular subject*; other candidates wrote that: *it is the process of finding the performance of something by using different measures such as watch, thermometer, ruler*. Others wrote: *it is the process of gathering numbers in order to determine the attributes possessed by an object or individuals*. In part (b) the candidates defined educational measurement as: *the process of assigning numerical values to educational attributes like interest, intelligence, skills and performance*. Others wrote that: *it is the process of finding the performance of education in learning activities, it may be individual or grouped and it should be measured through tests, examinations, quizzes and exercises*. The analysis of the candidates' performance shows that candidates had sufficient knowledge on Educational measurement. Extract 2.1 is a sample of a correct response from one of the candidates.

Extract 2.1

Q2.	(a) Measurement: Is the process of assigning number of attributes or characteristics of the person, object or events according to their rules.
	Example: Marks → % percentages Height → Meter (m) or Centimeter (cm) Weight → gram (g) or kilogram (kg)
	(b) Educational Measurement: Refers to the application of measurement knowledge of evaluating students achievement during teaching and learning process.

Extract 2.1 is a sample of a response from a candidate who was able to define the measurement and educational measurement concepts and gave examples correctly.

Further analysis shows that 163 candidates (18%) whose scores ranged from 2 to 2.5 marks performed averagely in the question. It appears that they had partial ideas on the meanings of measurement and educational measurement. In part (a), some of them provided incorrect responses such as: *Measurement it is the systematic of analysing data of certain phenomena/problems in useful means*; instead of writing: *it is the systematic way*. They similarly defined educational measurement as: *the process of correcting, analysing data in education system*. This shows that the candidate was poor in the English language, the medium of the examination.

Moreover, 105 candidates (11.6%) who scored from 0 to 1.5 marks; they defined measurement and educational measurement wrongly. Their responses were such as: *measurement is process of getting full total of something through different things*; and educational measurement was defined as: *the type of measurement where by experts doing for getting the learner who perform well and bad*. Other candidates defined measurement

as: *it refers to the science of collecting, analyzing, and recording a data for the aim of making evaluation and ability of an individual.* In part (b), some of the candidates defined educational measurement as: *the process of data collection which the education measurement during collect data.* Looking at these responses, it can be noted that the candidates were poor in English grammar; which was unfortunately the designated medium of answering the question.

Moreover, 45 candidates (5%) who scored 0 marks lacked knowledge of the subject matter. These provided irrelevant answers such as: *measurement is the process of observing and recording observation that are collected as part of research effort;* and educational measurement as: *the situation of using education assessment and analysis of data.* Others defined measurement: *as the process of obtaining unity of something while Educational measurement refers to the measurement which involve measurement of students performance so as to know achievement of their performance.* The responses show that candidates lacked sufficient knowledge of measurement and educational measurement. Extract 2.2 presents a sample of such a poor response.

Extract 2.2

2.	(a) Measurement is the instrument of data collection which involves measurement in data collection.	
	(b) Education measurement is the process of data collection which the education measurement during collect data.	

Extract 2.2 is a sample of a poor response from a candidate who failed to explain measurement and educational measurement.

2.1.3 Question 3: Educational Research

This question was set from a topic of Educational Research. The question had part (a) and (b). Part (a) instructed the candidates to define what action research is, while part (b) instructed them to give three areas in which action research can be used in education. Scores allocated to the question was 4 marks in total.

The question was attempted by 906 candidates (99.9%), of which, 425 candidates (46.9%) scored from 3 to 4 marks; indicating a good performance. of which 222 candidates (24.5%) scored 4 marks, and 150 candidates (16.6 %) scored from 2 to 2.5 marks; indicating an average performance. 331 candidates (36.5%) scored from 0 to 1.5 marks; indicating a poor performance out of which, 95 candidates (10.5%) scored 0 mark. The general performance in the question was average considering that 575 candidates (63.5%) scored from 2 to 4 marks. The data analysis is graphically illustrated by Figure 4.

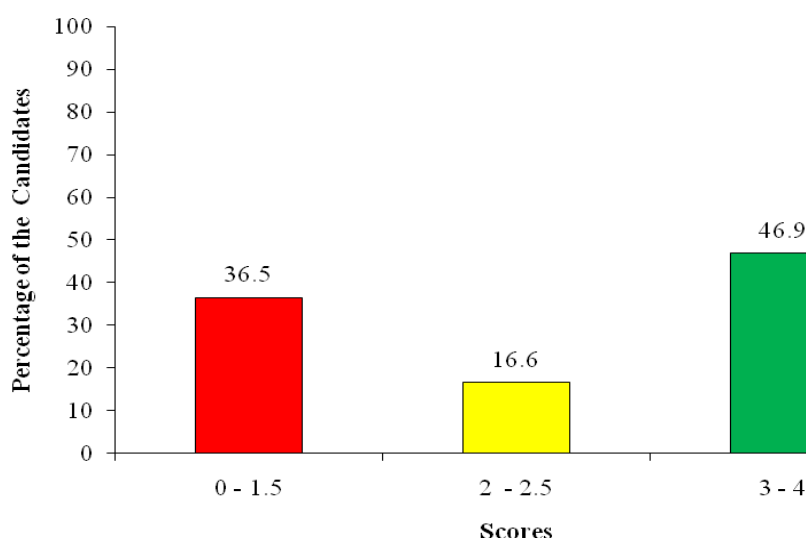


Figure 4: The Candidates' Performance in Question 3.

Further analysis indicates that 425 candidates (46.9%) who scored from 3 to 4 marks defined the concept of action research and provided the three areas in which action research can be used in education correctly. In part (a), the candidates' responses were such as: *action research is a practical approach to professional inquiry in any social situation and it is very important to teachers because it focuses on solving specific teaching and learning problems*. Others wrote: *an action research refers to the type of research which used to solve different problems in the educational issues*. Others defined action research as: *the type of an applied research which conducted with the aim of finding solution to the problem immediately*. In part (b), the candidates wrote areas in which action research can be used in education as: *in teaching methods; in area of teaching and learning materials; in curriculum development, during teaching and learning*

process, undertaken directly in the situation. Others wrote: in determining failure and success of curriculum materials; in learners learning difficulties, in academic issues. The responses imply that the candidates had sufficient knowledge and skills required by the question. Extract 3.1 is a sample of such a good response from the candidates.

Extract 3.1

3	(a) Action research is the form of applied research. It is used to solve the problems at the local levels. It concerns with day to day phenomena and problems. It is categorized into cyclic and spiral action research.
	(b) It can be used to investigate teaching and learning process
	• It is applied in determining the learning difficulties facing some students
	• It helps to know why some students perform well compared to others.

Extract 3.1 shows a correct response from a candidate who defined the concept of action research and showed different areas which action research can be used in education correctly.

Moreover, 150 candidates (16.6 %) who scored from 2 to 2.5 marks, an average performance, did not provide complete responses for example in part (a) they gave responses such as: *action research is the type of applied research in which data is planning, analysing, evaluating and reflecting and correcting in order to obtain real information.* Others wrote: *is the scientific and systematically way of information which used in action process and used to solve problem, is the type of research which formulate the result.* In part (b), the candidates gave areas in which action research can be used as: *in schools, in colleges, and in universities, in problem solving, in students' behavioural problems, in measuring learning difficulties, in academic performance, it can be used to treat problem of*

truancy amongst students, it can be used where there is bad/poor performance at school. Thus, the responses from the candidates denote insufficient knowledge and skills in action research and areas in which action research can be used.

Further analysis shows that 331 candidates (36.5%) with poor performance (who scored from 0 to 1.5 marks) lacked keenness in defining what action research is and identifying areas in which it can be used. Additionally, 95 candidates (10.5%) who scored 0 mark deviated from the demand of the question. Some of them wrote the overall definition of research, as others gave wrong areas of education in which action research can be used. Their responses were such as: *action research is the scientific procedures of solving, interpreting, observing and evaluating the certain problem immediately or by doing*. Others wrote: *action research is the proposal that guide research of things or topic that is going to take*. Similarly, the candidates cited different areas in which action research can be used in education as: *in mining industries; in hospitals; in agriculture; in engineering science*. Others wrote: *in identifying learning outcomes; in improving the performance of the learners*. Extract 3.1 presents a wrong response from one of such candidates.

Extract 3.2

3	a) What is an action research?	
	Action research refers to type of research which provide information and data to the researcher so as to make decision on different issues	
	b) Three areas in which action research can be used in education	
	i/ In Cognitive areas	
	ii/ In affective areas	
	iii/ In Psychomotor areas	

Extract 3.2 shows a response from a candidate who incorrectly defined action research and mentioned blooms taxonomy domains instead of areas in which action can be used in education.

2.1.4 Question 4: Assessing Achievement.

The question consisted of part (a) and (b) each from Assessing Achievement topic. Part (a) instructed the candidates to give two types of Interview in Educational Research, while, part (b) demanded the candidates to describe three limitations of using interview in educational research. Scores allocated to the question was 4 marks in total.

This question was attempted by 907 candidates (100%), of which, 456 candidates (50.3%) scored from 3 to 4 marks; indicating a good performance, of which, 168 candidates (18.5%) scored 4 marks. 226 candidates (24.9%) scored from 2 to 2.5 marks; indicating an average performance. 225 candidates (24.8%) scored from 0 to 1.5 marks; indicating a poor performance out, of which, 66 candidates (7.3%) scored 0 mark. The general performance in this question was good considering that 682 candidates (75.2%) scored from 2 to 4 marks. The analysis is graphically portrayed by Figure 5.

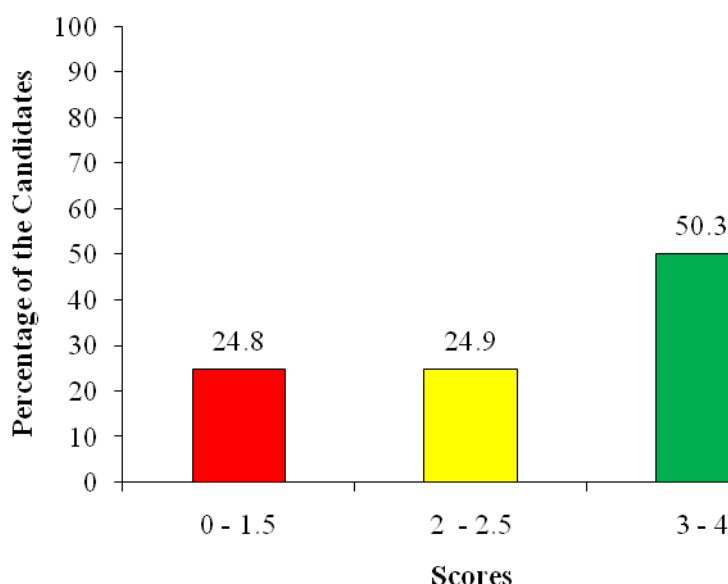


Figure 5: *The Candidates' performance in Question 4.*

Further analysis reveals that 456 candidates (50.3%) who scored marks ranging from 3 to 4 marks understood the demand of the question and had adequate knowledge of the concept of interview. For example, in part (a), many of them correctly mentioned the two types of interview in educational

research. They gave responses such as: *structured and semi-structured*. Others wrote: *unstructured interview and structured interview*. A few of the candidates wrote: *face-to-face interview and telephone interview*, and others wrote: *telecast interview and face-to-face interview*. In part (b), the candidates described three limitations of using interview in educational research correctly. For example, they wrote: *it consumes more time; it can face language barrier when two different languages are used; and it is difficult to be used in large population*. Other candidates wrote: *it creates biasness; it is not suitable to people with hearing problem; and it needs knowledge and skills in asking questions to the interviewee*. Others managed to describe the limitations of interview as: *sometimes interviewer can collect wrong data; sometimes it costs money when using telephone type of interview, and it may face technical problem if telephone interview is used when the network is down or insufficient power of the devices used*. All these varied responses on the limitations in using interview in educational research were correct and were in line with the demand of the question; showing the strengths of their responses. Extract 4.1 is a sample of a good response from one of the candidates.

Extract 4.1

4.	a/ Types of Interview in education research	
	① Structured Interview	
	② Non Structured Interview	
	b/ Limitations of using Interview in educational research.	
	① It discriminates hearing impaired for example deaf people can not be involved in Interview	
	② It can produce wrong information, some time interviewee can produce wrong information to the interviewer	
	③ It is costful in case of time and money because to conduct interview in research education it need money and more time.	

Extract 4.1 is a sample from a response of a candidate who gave two types of interview as structured and unstructured and described three limitations of using interview in educational research.

Furthermore, 226 candidates (24.9%), with average scores ranging from 2 to 2.5 marks, had partial knowledge of the types of interview and limitations of using interview in educational research. They mentioned the types of interview in part (a), though failed to describe three limitations of using interview in educational research. Their responses in part (b) were such as: *it creates fear to the learner, it can lead risk between respondent and an interviewer; it can lead to get wrong information, it needs high skills and knowledge to conduct an interview*. Moreover, a few of the candidates wrote: *sometimes interview bores if a lot of questions are to be asked*. Despite having valuable points, their responses were not clear and did not meet the demands of the question.

Additionally, 225 candidates (24.8%) who performed poorly (from 0 to 1.5), of which, 66 candidates (7.3%) scored 0 marks, failed completely to mention the types of interview and wrongly described three limitations of using interview in educational research. Their responses in part (a) were: *null interview, and action interview; open interview, and closed interview, questionnaire and physical interview*. Others wrote: *interview research, and conversational interview research; respondent interview and non-respondent interview*. In part (b), some wrote answers like: *you are required to keep secret, interview has no reference, and question may not be understood*. These responses were incorrect. Other wrote types of interview as: *job interview and mass media interview*. In part (b), others deviated from the question requirements, and wrote ethical issues to consider when researcher collecting data such as: *the researcher required keeping secret* instead of limitations of interview. This was not one of the limitations of using interview in educational research. Others wrote things such as: *lack of focus; stress and out of control emotion; exclude body language; helps to know the difficulties of the learners* which are not correct. Others wrote: *it involves all students in participation of question, and it improves language using*. These varied illogical responses show that the candidates were not certain about what interview is, and its limitations in educational research. Extract 4.2 illustrates poor responses from one of the candidates in this category.

Extract 4.2

4	(a) Formative Interview	
	(b) Diagnostic Interview	
	(b)	
	(i) It measures speaking skills only	
	(ii) The responses given by the students are not recorded for future educational use!	
	(iii) It need high questioning skills to obtain full information	

Extract 4.2 is a sample of poor responses from candidates who wrongly gave two types of interview and limitations of using interview in educational research.

2.1.5 Question 5: Analysis and Interpretations of the Test Results.

This question contained two parts: (a) and (b). These two parts were set from Analysis and Interpretations of the Test Results topic. Part (a) instructed the candidates to describe two measures of variability. Part (b) instructed the candidates to list down two techniques of standardizing test scores. The total marks allocated for this question was 4 marks.

The question was attempted by all 907 candidates (100%). 250 candidates (27.6%) scored from 3 to 4 marks; indicating a good performance, of which, 156 candidates (17.2%) scored 4 marks. Moreover, 329 candidates (36.3%) scored from 2 to 2.5 marks; indicating an average performance. The rest 328 candidates (36.2%) scored from 0 to 1.5 marks; of which, 278 candidates (30.7%) scored the 0 mark; indicating a poor performance. The overall performance in the question was average considering that 579 candidates (63.9%) scored from 2 to 4 as portrayed in Figure 6.

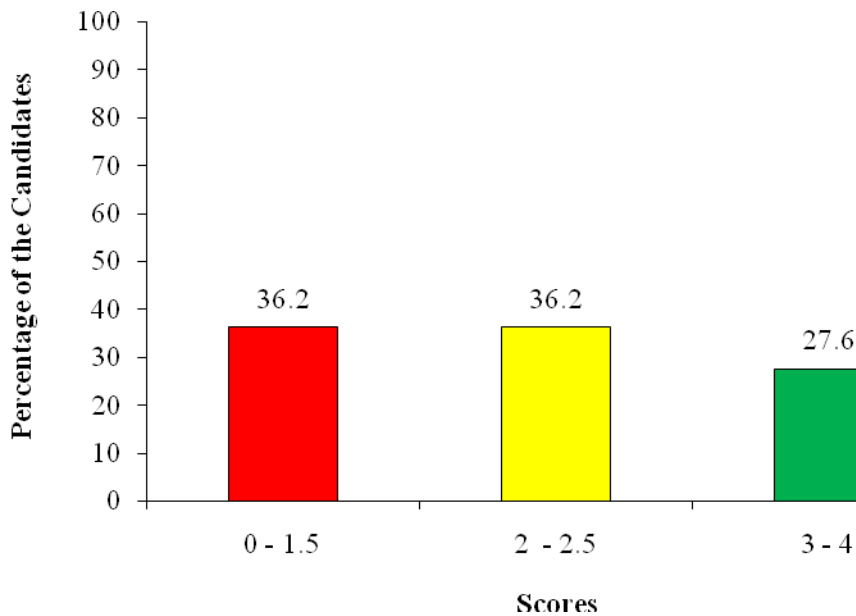


Figure 6: *Candidates' Performance in Question 5.*

The analysis reveals that 250 candidates (27.6%) who scored from 3 to 4 marks managed to briefly explain two measures of variability and presented their explanation in organized manner. They also cited the formula for each measure correctly. In part (a), many of the candidates demonstrated adequate knowledge of measures of variability. Their responses on measures of variability were such as: *range*; *variance*; and *standard deviation*. The candidates correctly defined range, variance, and standard deviation as: *range shows the difference obtained between the highest and the lowest score in the observation, variance is the squared deviation over the number of objects, and standard deviation is the square root of variance* respectively. In part (b) of the same question, many of the candidates correctly listed the two techniques of standardizing test scores, such as: *Z-score* and *T-score*. Some of the candidates went as far as by writing their formulae and examples for each technique. They listed and wrote formulae as:

$$Z - Score = \frac{(X - \bar{X})}{SD} \text{ and } T - Score = 10Z + 50. \text{ These formulae were}$$

correct. This demonstrates that the candidates were knowledgeable of measures of variability. Extract 5.1 presents a correct response from a candidate who scored full marks.

Extract 5.1

05	97	
	There following are the two measures of variability which are	
	i/ Standard deviation	
	Is the square root of variance	
	Standard deviation = $\sqrt{\frac{\sum (x - \bar{x})^2}{N}}$	
	ii/ Variance Is the ratio between square of deviation and the total number of score	
	$\sigma^2 = \frac{\sum (x - \bar{x})^2}{N}$	
	⇒	
	i/ Z-score	
	Is the ratio between deviation and standard deviation	
	Z score = $\frac{\text{deviation}}{\text{Standard deviation}}$	
	ii/ T-Score = $10Z + 50$	

Extract 5.1 is a sample from a candidate who correctly described measures of variability as standard deviation and variance and listed two techniques of standardising test scores as T-score and Z-score.

Further analysis indicates that 329 candidates (36.3%) who performed averagely (from 2 to 2.5 marks) had inadequate knowledge and skills on part (a) or part (b) of the question. Some of them confused measures of variability with measures of central tendency such as: *median and range, variance and mean, standard deviation and mode, deviation and standard deviation*. In part (b), many of the candidates listed the two techniques of standardizing the test scores as: Z-Score and T-Score but they failed to elaborate them by words or formulae which lowered their marks. Some of the candidates wrote Z- Score formulae as: $Z\text{-Score} = \frac{(X - X)}{SD}$ instead of $Z\text{-Score} = \frac{(X - \bar{X})}{SD}$.

In addition, 328 candidates (36.2%) who scored from 0 to 1.5 mixed different ideas their responses on part (a) of the question were such as: *content; range; concrete; criterion related; objective test and subjective test*. Although some of these responses were correct, some were incorrect. In part (b), the candidates' responses were such as: *rating; objective test and subjective test; Z - Score; decimal values; negative and positive values*. As only *range* was correct out of the six responses, they were awarded low marks. Another candidate instead of writing measures of variability wrote measures of central tendency such as: *mean, mode, and median*. In part (a), some of them wrote the types of variables such as: *independent and dependent variable* instead of measures of variability such as: *range, variance, and standard deviation*. In part (b), some of the candidates listed: *spearman's rank order method*; and *split half method*, these two are the methods of estimating tests reliability (coefficient correlation). Extract 5.2 shows such incorrect response from one of the candidates.

Extract 5.2

a)	Briefly explain two measures of variability
	Dependent variability (deals with teaching method)
	Independent variability (provide good performance)
b)	List down two techniques of standardising test scores
	- spearman's rank order method
	- split half method

Extract 5.2 shows a response from a candidate who failed to explain two measures of variability as standard deviation, variance and range and techniques of standardizing test scores as T-score and Z score.

2.1.6 Question 6: Assessing Achievement

The question required the candidates to identify the four functions of objective test. The question was set from Assessing Achievement topic. The total scores allocated to the question were four (4) marks, whereby, each correct response carried 1 mark.

All 907 candidates (100%) attempted this question. 832 candidates (91.7%) scored from 3 to 4 marks; indicating a good performance, of which, 564

candidates (62.2%) scored 4 marks. 49 candidates (5.4%) scored from 2 to 2.5 marks; indicating an average performance. However, 26 candidates (2.9%) scored from 0 to 1.5 marks; indicating a poor performance, of which, 7 candidates (0.8%) scored 0 mark. The general performance in the question was good considering that 613 candidates (97.1%) scored from 2 to 4 marks. Figure 7 illustrates the candidates' performance in question 6.

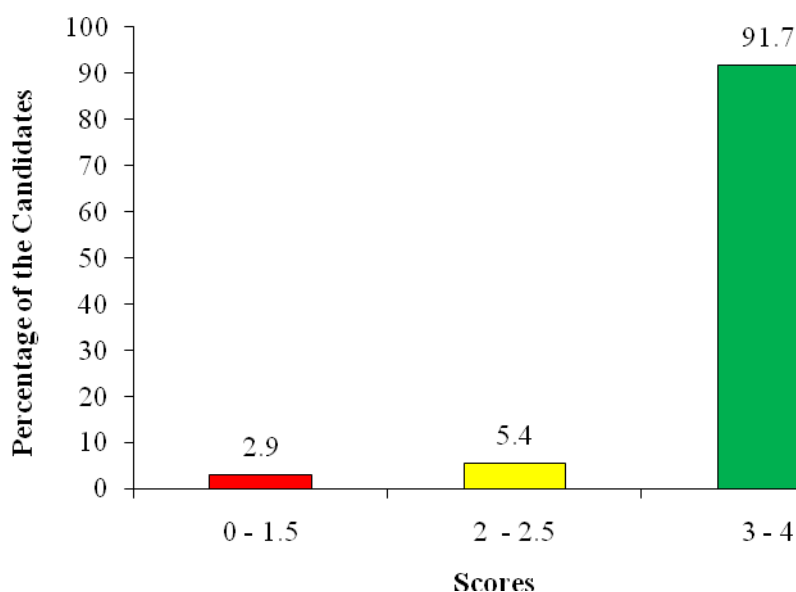


Figure 7: *The Candidates' Performance in Question 6*

Detailed analysis shows that 832 candidates (91.7%) who scored from 3 to 4 marks, high performance, had sufficient knowledge of the four functions of objective tests. They correctly identified four functions of objective test such as: *it is used to measure ability of recalling for the student, objective tests enable students to be attentive in their course of study, helps to give out short answers; easy to mark it*. Extract 6.1 presents a sample of such good responses in question 6.

Extract 6.1

6.	Function of objective test	
	(i) It make student in larger area of coverage	
	e. To pass different area of content in a certain subject	
	(ii) Reduce time for scoring the test, A markers can mark the test within short period of time compare with objective test.	
	(iii) Objective test involve logical question which makes learner to become critical thinker.	
	(iv) High reliability compare to subjective test.	

Extract 6.1 shows correct responses from a candidate who correctly identified four functions of objective test

Moreover, 49 candidates (5.4%) scored from 2 to 2.5 marks; which is average. The candidates identified near correct responses such as: *objective tests take few times to mark and they enable students to respond more quickly to the item, it cover small domain range, they encourage high degree of guessing due to presence of pre-determined answers, especially multiple choice questions are easy to construct objectives tests*. These candidates had partial knowledge on the types of tests.

Additionally, 26 candidates (2.9%) who scored from 0 to 1.5 marks, a poor performance lacked knowledge and skills in functions of objective questions. They gave wrong responses such as: *high maximum rate of guessing; can be scored by using machines or man power; ability to grasp meaning; to measure learners thinking capacity*. Most of those who scored 0 mark skipped the question. Extract 6.2 is a sample from a candidate who performed poorly in the question.

Extract 6.2

6.	To motivate recalling and remembering of student	
	ii) for selecting team under Multi under Selection Item)	
	iii) selecting similar answers under Matching team.	
	iv) differentiate unlike answers.	

Extract 6.2 is a sample of incorrect responses from one of the candidates who was unable to identify functions of objective test as intensive coverage of content, can scored quickly, student are able to respond quickly and encourage student to pay attention to what they are studying.

2.1.7 Question 7: Assessing Achievement

This question was derived from *Assessing Achievement* topic. The question required the candidates to briefly explain four effective uses of anecdotal records in assessing student achievements. Scores allocated to the question was 4 marks in total.

This question was the most poorly attempted question since 866 candidates (95.5%) scored from 0 to 1.5 marks; indicating a poor performance, of which, 474 candidates (52.3%) scored 0 mark. 32 candidates (3.5%) scored averagely, from 2 to 2.5 marks. The rest 9 candidates (1%) scored from 3 to 4 marks, a good performance. Generally, the performance in the question was poor considering that 866 candidates (95.5%) scored from 0 to 1.5 marks. Figure 8 presents the students' performance in question 7.

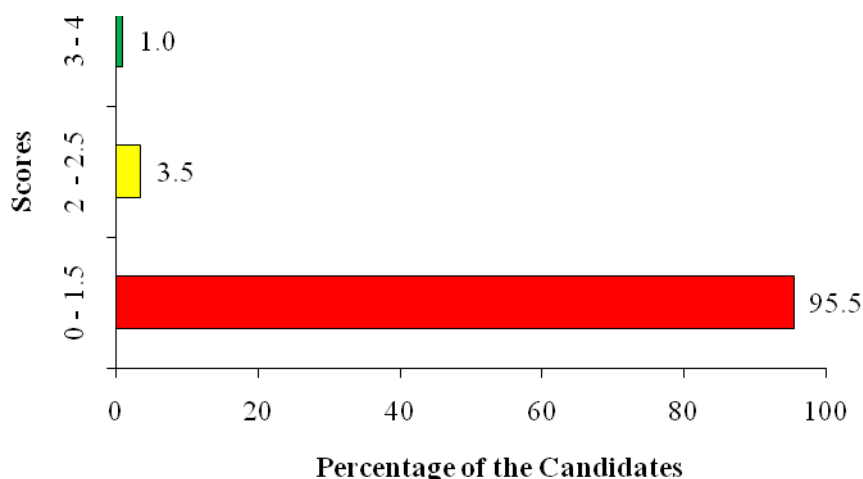


Figure 8: *The Candidates' Performance in Question 7*

Detailed analysis shows that 866 candidates (95.5%) who scored from 0 to 1.5 marks lacked knowledge and skills in the effective uses of anecdotal records in assessing students' achievement. Some of the candidates, in this category, did not answer the question at all. Others incorrectly provided effective uses of anecdotal records. For example, one defined it as: *anecdotal records is used to determine performance of the learners*. For its effective uses, the candidate wrote: *a teacher used records to solve learners' problem, helps to know why some students performed well, it can*

be used to evaluate learning achievement of learners. Furthermore, some of the candidates defined anecdotal records correctly, but failed to give its effective uses. For example one wrote: *anecdote is the factual description of behaviour observed soon after happening which include time, place, day, and comments on the event observed.* Some of the candidates however mixed correct and incorrect use of anecdotal records. The exemplary uses are such as : *help to make a record on incidence as soon as researcher observe and used to determine cognitive ability of the learners, to assess students' achievement to a course, and it helps in continuous assessment, respectively.*

Extract 7.1

7.	It uses the accurate of information of which is achieved, Anecdotal records is very important for student achievement due to the accurate of information which are being used to assess the student.	
	It involves investigation in searching the detail for assessing the achievement of student, the anecdotal use more investigation about the achievement of student performance.	
	It uses the recorded data or information - which has passed for a long time, Anecdotal records use the information which were recorded for long time in order to achieve the student achievement of every students through those informations.	
	It uses a past things in measuring those - achievement of a student by looking on both things which has identified the achievement of the students.	

Extract 7.1 is a sample of poor responses from candidates who failed to explain the uses of anecdotal records as determine what to observe, observe and record situation, record the incidence as soon as you observe, record both positive and negative behaviour incident.

Moreover, the analysis shows that 32 candidates (3.5%) who scored averagely (from 2 to 2.5 marks) gave responses such as: *used to show behaviour of the students, Help to collect number of incidences (anecdotes) before drawing, used to solve the educational problem, limits each anecdote to a brief description* while the second and third response are correct, the first and fourth are incorrect. These indicate that, the candidate had partial knowledge on the tool of assessing student achievement.

In addition, the analysis shows that 4 candidates (1%) scored from 3 to 4 marks, which is good score. These gave the effective uses of anecdotal records such as: *used to make a record of events as soon as they appear, help to record all positive and negative behaviour observed and draw conclusion on them, it should be documented in various way by considering day, time and event itself. Others wrote that: it is used in grading or identifying students with different behaviour; help observer in recording learner's changes of behaviour by considering when, where, and time the incident occurred and finally interpret such behaviour observed.* Extracts 7.2 illustrates such a good response from the candidates.

Extract 7.2

7.	A anecdotal records is the factual description of behaviour which observed soon after happens. This include time, name, age, and place where events occurred or happens. Its effective uses is that.
-	It used to know real behaviour of students. Anecdotal record observe real situation soon after happening.
	It used to express real time when event happen, in anecdotal record, its major element is time which explains when event happens.
	It identifies where the situation are occurred. soon after event happen and where, anecdotal record explains.
	Used to explain age and name of participant who shows some level of behaviour. either his/her name and ages.

Extract 7.2 presents responses from a candidate who satisfactorily defined and explained the four uses of anecdotal records.

2.1.8 Question 8: Test Construction

The question required the candidates to briefly explain four advantages of using table of specifications in test construction. The question was set from Test Construction topic, and carried 4 marks in total.

The question was attempted by all 907 candidates (100%). The analysis shows that 328 candidates (36.2%) scored from 3 to 4 marks; indicating a good performance, of which, 93 candidates (10.3%) scored 4 marks. 368 candidates (40.5%) scored from 2 to 2.5 marks; indicating an average performance and 211 candidates (23.3%) scored from 0 to 1.5 marks; indicating a poor performance, of which, 36 candidates (4%) scored 0 mark. The general performance in the question was good considering that 796 candidates (79.7%) scored from 2 to 4. Figure 9 illustrates the performance in question 8.

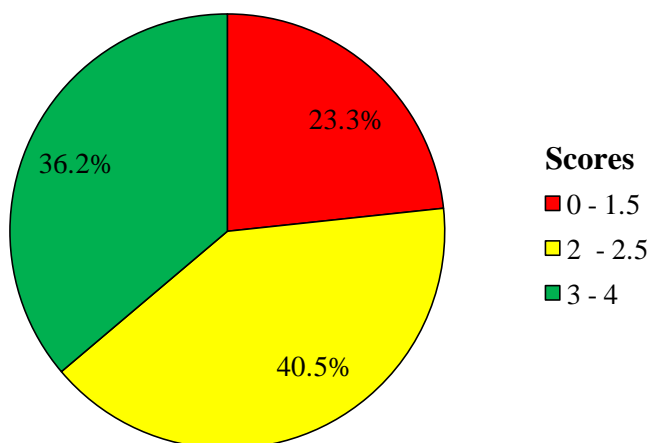


Figure 9: *The Candidates' Performance in Question 8*

Detailed analysis shows that 328 candidates (36.2%) who scored from 3 to 4 marks briefly explained four advantages of using table of specifications they started with definition of table of specifications and proceeded to explaining its four advantages. An exemplary response is: *table of specification (blue print) is a two way chart showing the relationships between the content/topic area in one hand and the domain levels on the other hand. In explaining its advantages, the candidate wrote: it helps to indicate the content to be tested; it helps to show the cognitive domain to be measured by a teacher, it helps to show number of questions appear in each topic so as to determine the length of the test; it helps to show the*

topic in which a teacher give more weight in the examination. These responses were correct. The responses imply that the candidate had sufficient knowledge of the concept of table of specification. Extract 8.1 is a sample of such good responses.

Extract 8.1

8	Four advantages of using table of specification in test construction,	
	Table of specification ÷ This is the two way chart which showing both domain skill and the course content intended to be measured.	
	Advantages of table of specification in test construction.	
	Show what is intended to be measured ÷ Table of specification show all items and domains which is intended to measured	
	Large coverage of domain skills ÷ Table of specification also can help in large coverage of domain skills performance.	
	Specify the items used in specific test ÷ Also table of specification may specify the items used in performing the test.	
	Good arrangement of the course content ÷ Table of specification also may show the good arrangement of the content course. Also may show the items in specified order.	

Extract 8.1 shows responses from a candidate who correctly explained four advantages of using table of specification.

The analysis equally shows that 368 candidates who scored from 2 to 2.5 marks had partial knowledge and skills in the advantages of using table of specifications, or misconceived it. For example, some of their responses were such as: *it helps to identify the total number of items to be constructed thus can enable the teacher to understand the number of test items to be constructed. It helps to understand the percentage that carried out on each item cross ward and down ward, can guide during test construction, it helps to identify the content to be tested: this can enable the teacher to stick*

and concentrate to topics in the table of specifications; helps determine the purpose of the test – all these responses were correct but not clear.

Further, analysis shows that 211 candidates (23.3%) who scored from 0 to 1.5 marks gave incorrect or incomplete responses like: *to determine the strengths and weaknesses of teaching methods; helps teacher determine the techniques of preparing the test.* Others wrote: *it helps teacher to pay attention about the ability of each student; helps to understand the objective of the test; it is used to measure the whole content on the specific topic; it is used in evaluation.* The responses demonstrate the candidates lack of knowledge of table of specifications. Extract 8.2 is a sample of such poor responses.

Extract 8.2

8.	four advantage of using table of specification	
	tion	
	(i) It is easy scoring test: Due to the good arrangement of items in the table.	
	(ii) It save time! Means during the test construction it help the teacher to construct fast	
	(iii) It avoid content discrimination: Due to the use of table of specification all content is considered	
	(iv) It avoid confusion: Due to when the teacher select the items in order for	

Extract 8.2 presents responses from a candidate who failed to explain the uses of table of specifications as; Specify systematically the domain of knowledge to be covered in a test, indicate the weighting of different content areas, determine the number of items to be included and Ensures that the test has a proper balance

2.1.9 Question 9: Qualities of Tests

This question required the candidates to provide a brief explanation on the four qualities of a good test. Scores allocated to the question was 4 marks in total.

All 907 candidates (100%) attempted the question, of which, 273 candidates (30%) scored from 3 to 4 marks, of which, 24 candidates (2.6%) scored 4 marks; indicating a good performance, 500 candidates (55.1%) scored from 2 to 2.5 marks; indicating an average performance. However, 134 candidates (14.8%) scored from 0 to 1.5 marks, of which, 29 candidates (3.2%) scored 0 mark; indicating a poor performance. Generally, the performance in this question was good considering that 723 candidates (85.2%) scored from 2 to 4 marks. Figure 10 presents the candidates' performance in question 9.

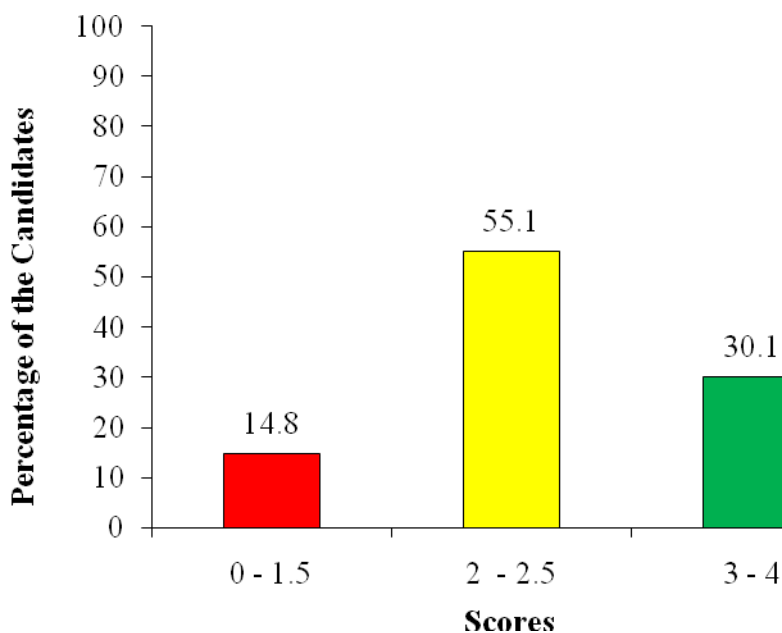


Figure 10: *The Candidates' Performance in Question 9*

The detailed analysis shows that 273 candidates (30%) who scored from 3 to 4 marks correctly explained the four major qualities of a good test and elaborated each quality of a good test. For example, one candidate wrote: *test should be valid to test what the test intended to test, test should be reliable by providing identical results when it is administered in different*

time, test should be clear. Another one wrote: test should indicate the distribution of test items by regarding the table of specifications; test should show the domain to be assessed. Extract 9.1 is a sample of such good responses.

Extract 9.1

9.	Test is the instrument used to measure the ability of the learner or students.
	The qualities of good test are as follows:
	There are-
	Validity: This measures the content that is claim to measure.
	Reliability: This is the type of test that measure same samples of behaviour.
	Relevance: Means that test must correlate with subject content taught so as make student to understand well.
	Specificity: Means that the test constructed must specific to the level of learner and their age.

Extract 9.1 is a sample from a candidate who correctly explained the four qualities of a good test.

Moreover, data analysis shows that 500 candidates (55.1%) who scored averagely (from 2 to 2.5 marks) wrote partially correct responses. For example, they wrote: *validity which refers to consistency in measuring student achievement*, instead of, validity measures what the test intended to measure, *a good test must be specific; a good test should be relevant to the level of the learner*. Others wrote: *it should be reliable in providing similar results*. Some of the candidates wrote: *clarity; applicability; practicability; specificity* with no explanation. Though all of these are correct, they could not explain each quality correctly.

Likewise, 134 candidates (14.7%) who performed poorly (from 0 to 1.5 marks) lacked knowledge and skills needed. They gave wrong responses such as: *capability, should show time and content, simplicity, should consider learning ability*. Extract 9.2 is a sample of such poor responses from the candidates.

Extract 9.2

9	i) Testable	
	ii) Measurable	
	iii) Impirical	
	iv) Logical	

Extract 9.2 is a response from a candidate who failed to explain the four qualities of a good test correctly.

2.1.10 Question 10: Analysis and Interpretations of the Test Results

This question had part (a) and (b). The question required the candidates to distinguish between the measures of dispersion. The measures asked in part (a) included range and standard deviation, and in part (b), variance and standard deviation. Total scores allocated to this question was 4 marks in total.

The question was attempted by all 907 candidates (100%), of which, 461 candidates (50.8%) performed well (from 3 to 4 marks). Moreover, 124 candidates (13.7%) performed averagely (from 2 to 2.5 marks). 322 candidates (35.5%) performed poorly (with the scores ranging from 0 to 1.5 marks). The overall performance in the question was average considering that 585 candidates (64.5%) scored from 2 to 4 marks as shown in

Figure 11.

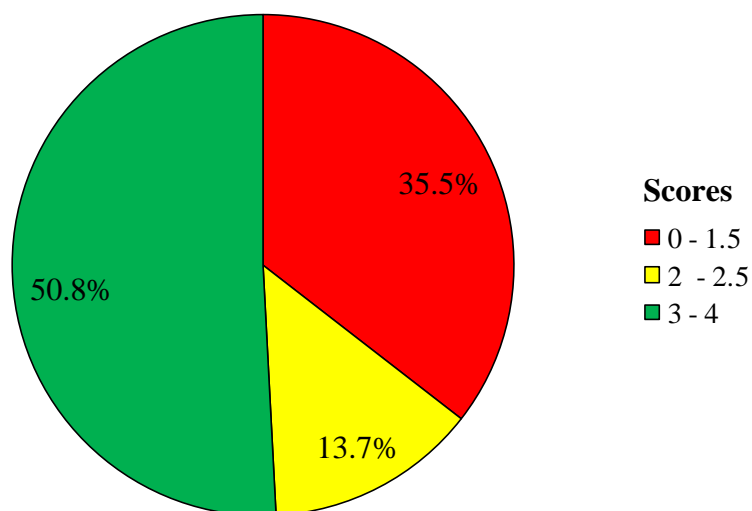


Figure 11: *The Candidates' Performance in Question 10*

The analysis shows that 461 candidates (50.8%) who attempted the question and scored well (from 3 to 4 marks). This proves sufficient knowledge and required skills. In part (a), they, for instance, wrote: *range as the different between highest and lowest number in the distribution while standard deviation is the under root of variance*. Others wrote: *Range refers to different between the highest and lowest scores and provided an example of series of scores to determine the range as 40, 45 68, 70, and 80. Thus, range = 80-40 =40*. Other candidates defined and write the formula for each measure of variability as: *range refers to highest score minus lowest =h - l while standard deviation is the value after under rooting the variance value $SD=\sqrt{V}$* . In part (b) of the question, the candidates with good correctly provide correct distinguished between variance and standard deviation and provided example for each measure. They wrote: *variance is the average of the squared deviations from the mean $V= \frac{\sum(x-\bar{x})^2}{N}$ while standard deviation is the square root of the variance. it is the measure of degree in which scores tend to deviate from the mean $SD= \sqrt{\frac{\sum(x-\bar{x})^2}{N}}$* . Extract 10.1 shows a sample of good responses from the candidates.

Extract 10.1

SECTION "A"

10. @ Range: Is the arrangement of numbers (attributes) from either highest number or lowest number in a given attributes, where by can be calculated by the formula of $\text{Range} = \text{Highest score} - \text{lowest score}$.

while standard deviation: Is the square root of the Variance that are consistent in the attributes of a test, where can be calculated by the formula

$$SD = \sqrt{\frac{\sum (x - \bar{x})^2}{N}}$$

where by SD = standard deviation

\sum = summation of

x = The Number of scores

\bar{x} = The mean score

N = Number of items

(b) Variance: Is the measures of Variability that obtained after the consideration of the deviation score of the items. It can be calculated by using the formula of

$$V = \frac{\sum (x - \bar{x})^2}{N}$$

where by V = Variance

\sum = summation of

x = The number of score

\bar{x} = The mean score

N = The number frequency

while

10.	Standard deviation; is the square root of the Variance that may be constant in a given test scores. It written by the symbol (s.d) and calculated by the formula
	$SD = \sqrt{\frac{\sum (x - \bar{x})^2}{N}}$
	Where by SD = Standard deviation
	\sum = summation of
	x = The score number
	\bar{x} = The Mean score
	N = The number of frequency

Extract 10.1 is good responses from a candidate who correctly distinguished measures of dispersion.

Further, the analysis shows that 124 candidates (13.7%) who performed averagely (from 2 to 2.5 marks) misconceived on measures of dispersion in either part (a) or part (b) of the question. They wrote things such as: *range is high item minus low item while standard deviation is the under root of variance*. Others wrote: *the range is the differences which occur from the higher score to lower score in the examination while standard deviation is the under root of the number which multiply with themselves*. In part (b) of the question some of the candidates distinguished between variance and standard deviation as: Variance = $\frac{\sum (x - \bar{x})^2}{N}$ and Standard deviation as: $SD = \sqrt{\frac{\sum (x - \bar{x})^2}{N}}$ without any explanation. The responses though correctly lacked elaborations.

Furthermore, 322 candidates (35.5%) who performed poorly, with the scores ranging from 0 to 1.5 marks, failed to distinguish between the measures of variability in both parts or in one part of the question. For instance in part (a), one distinguished range and standard deviation as: *range refers to the measure of dispersion in which involve the highest score to minus or substitute the lowest score in a given population while*

Standard deviation refers to measures of dispersion in which it is calculated by $\sqrt{\frac{\sum(x-\bar{x})^2}{N}}$. The responses which are only partially correct.

Others wrote: range is difference between highest and lowest number while standard deviation is the sum of scores divide by their number. In part (b), the candidates distinguished between variance and standard deviation as: Variance is the variable value from mean while standard deviation is dispersion deviated from variance. Extract 10.2 presents a sample of poor responses from the candidates.

Extract 10.2

10	a) range - if not used in stan	
	- range formula is not the same	
	with standard deviation : due to	
	that the two terms are differ	
	b) standard deviation - use to measure	
	degree which deviated in the test	
	score while variance do not	
	measure the degree in the test score	

Extract 10.2 is from a candidate who failed to distinguish between the measures of dispersion.

2.2 SECTION B: Essay Questions

This section had six (6) questions. Question eleven (11) was compulsory to candidates while other five (5) questions were optional. The total scores allocated to this section was sixty (60) marks; each carrying 15 marks.

2.2.1 Question 11: Analysis And Interpretations of the Test Results

This question had two parts: (a) and (b). Part (a) required the candidates to calculate the correlation coefficient between students' performance of the two tests (A and B) by using Spearman's Rank Order Method. Part (b)

required the candidates to give reasons on whether the two subjects indicate the same students' ability or not. The allocated scores to this question were 15 marks.

The question was compulsory and was attempted by all 907 candidates (100%). 143 candidates (16.6%) scored from 10.5 to 15 marks; indicating a good performance, of which, 1 candidate (0.1%) scored the highest 4 marks. 206 candidates (22.9%) scores from 6 to 10 marks, indicating an average performance. The other 546 candidates (60.5%) scored from 0 to 5.4 marks; indicating a poor performance, of which, 242 candidates (26.7%) scored 0 mark, the lowest. Generally, the performance in the question was average considering that 349 candidates (39.5%) scored from 6 to 15 marks. Figure 12 illustrates the candidates' performance in the question.

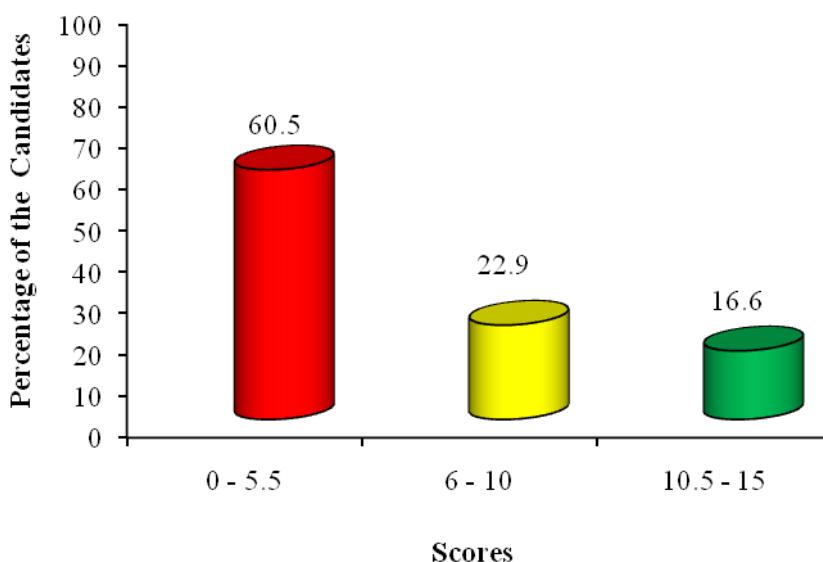


Figure 12: *The Candidates' Performance in Question 11*

The analysis of the candidates' performance shows that 143 candidates (16.6%) who scored from 10.5 to 15 marks, a good performance, managed to use the Spearman's Rank Order Method effectively to calculate the correlation coefficient between scores in test A and scores in test B according to the procedures. In part (a), they managed to draw a table consisting of seven (7) columns and labelled them as: Students' Name (SN), Scores in Test A (TA), Scores in Test B (TB), Rank in Test A (RTA), Rank

in Test B (RTB), Rank Difference (D), and Squared Difference (D^2). The candidate also wrote the Spearman's Rank Order Method formula as:

$$r = 1 - \frac{6 \sum D^2}{N(N^2 - 1)}$$
 and used this formula to get the correct correlation

coefficient of the Test A and Test B = 0.9. In part (b), they managed to interpret the correlation coefficient. They wrote: *the test A and B indicate that students in test A and B had the same ability because the value of correlation coefficient was positive.* Other wrote that: *The two tests indicate students' same ability.* Extract 11.1 is a sample from a candidate with good responses.

Extract 11.1

Students Name	Score in Test A	Score in Test B	D	D ²
Simbomboy	2	3	1	1
Mugo	4	5	1	1
Kinani	10	10	0	0
Magu	9	9	0	0
Muturubhani	1	2	1	1
Mapoke	3	1	-2	4
Kinye	6	4	-2	4
Dude	5	7	2	4
Mwike	7	6	-1	1
Kine	8	8	0	0
				$\sum D^2 = 16$

$$r = 1 - \frac{6 \sum D^2}{N(N^2 - 1)}$$

where

r = Correlation coefficient

\sum = Summation symbol

D = Difference of scores of Test A and Test B

N = Total number of students

$$r = 1 - \frac{6 \times 16}{10(10^2 - 1)}$$

$$r = 1 - \frac{96}{990}$$

$$r = 1 - 0.09697$$

$$r = 0.903$$

$$r \approx 0.9$$

$$\therefore \text{Correlation coefficient} = 0.9$$

(b) Two test indicates the same students ability because the correlation coefficient got of 0.9 shows that the test is stable.

Extract 11.1 is good responses from a candidate who successfully calculated correlation coefficient and interpreted the correlation value between test A and test B.

Further analysis reveals that 206 candidates (22.9%) who scored from 6 to 10 marks, average, draw a table and labelled the heading of each column according to the need of formula. However, they were correct in part (a) but incorrect in part (b). Their responses in part (a) were such as: *the correlation coefficient is calculated by using the formula as $r = 1 - \frac{6\sum D^2}{N(N^2 - 1)}$ and managed to get $\sum D^2 = 16$ correctly and use the*

Spearman's Rank Order Method to get $r = 0.9$. Further, they were unable to interpret whether the test A and B showed the same students' ability or not. For instance in part (b), many of the candidates wrote: *there were no same students' ability in test A and test B*. This response was incorrect. Many other candidates wrote: *yes, the two tests indicate the same students' ability* (correct response) while others wrote: *the two tests indicate no students' ability* (incorrect response). These two varied answers lacked elaboration.

Furthermore, the data shows that 562 candidates (60.5%) who scored poorly in part (a) were unable to write the formula of Spearman's Rank Order Method appropriately. They also failed to follow the procedures of calculating the correlation coefficient of test A and B correctly. In part (b), they failed to explain whether the two tests showed the same student ability or not. The reasons for part (b)'s responses depended much on the correctness of the value of correlation coefficient computed in part (a). The candidates wrote: *the test indicates the same ability because some students in the two tests administered obtained the same results for example Kime, Magu and Kinani*. Others wrote: *students scored the same in test A and B*. The following extract 11.2 presents a sample of such poor responses from the candidates.

Extract 11.2

11	Names	X	Y	XY	X ²	Y ²
	Simbambaye	2	3	6	4	9
	mugo	4	5	20	16	25
	kinani	10	10	100	100	100
	magu	9	9	81	81	81
	mutumbani	1	2	2	1	4
	maake	3	1	3	9	1
	kinye	6	4	24	36	16
	Dade	5	7	35	25	49
	mwike	7	6	42	49	36
	kinu	8	8	64	64	64
	ΣX	55	ΣY	55	ΣX^2	385
Solution: $r = \frac{n(\Sigma XY) - (\Sigma X)(\Sigma Y)}{\sqrt{[n \Sigma X^2 - (\Sigma X)^2][n \Sigma Y^2 - (\Sigma Y)^2]}}$						
$r = \frac{10(377) - (55)(55)}{\sqrt{[10(385) - 3025][10(385) - 3025]}}$						
$r = \frac{3770 - 3025}{\sqrt{[10(385) - 3025][10(385) - 3025]}}$						
$r = \frac{3770 - 3025}{\sqrt{[148250 - 148225][148250 - 148225]}}$						
$r = \frac{114025}{\sqrt{134025 \times 134025}}$						
$r = \frac{114025}{15480898}$						

Extract 11.2 is from a candidate who failed to calculate the correlation coefficient of two tests by use of the spearman's ranks order method formula and reasons as to whether the two tests indicated the same student ability.

2.2.2 Question 12: Test Construction

In this question, the candidates were instructed to analyze five factors affecting subjectivity of a test. Total marks allocated for this question was 15 marks.

The question being optional, was attempted by only 401 candidates (44.2%), of which, 30 candidates (7.5%) scored from 10.5 to 12.5 marks; indicating a good performance. 291 candidates (72.5%) scored from 6 to 10 marks; indicating an average performance. Moreover, the rest 80 candidates (20%), scored from 0 to 5.5 marks, of which, 1 candidate (0.1%) scored the lowest 0 mark; indicating a poor performance. The general performance in the question was good – considering 321 candidates (80.1%) scored from 6 to 15 marks. Figure 13 below gives summary of the performance in the question.

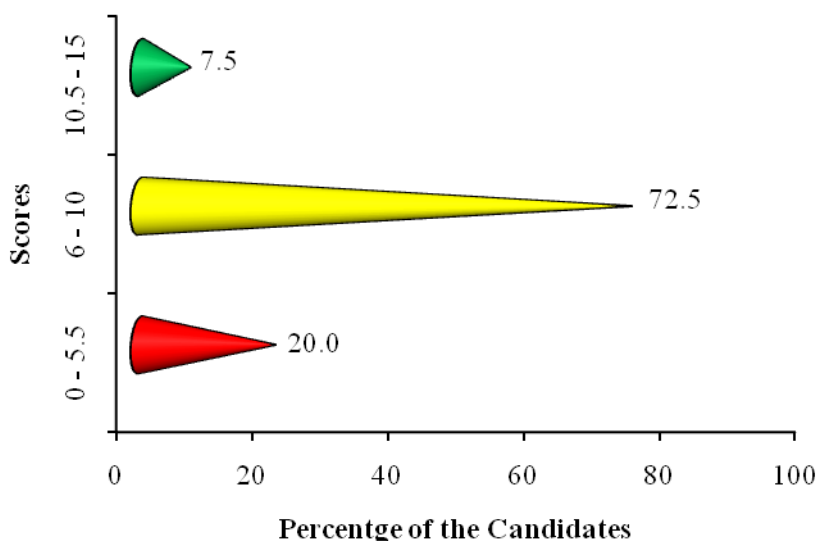


Figure 13: *The Candidates' Performance in Question 12*

The detailed analysis shows that 30 candidates (7.5%) who scored from 10.5 to 15 marks provided relevant introduction and clearly analyzed five factors affecting subjectivity of the test such as: *unclear instructions that the test provide, there is no pre-determined answer, differences in words and grammar may affect subjectivity, marking criteria may affect subjectivity*. Some of them wrote: *unclear marking scheme, improper panel system during scoring exercise when one marker marks more than one*

essay question, use of candidates' number instead of candidates' name, the use of difficult language. Other candidates wrote: poor construction of test items, absence of marking scheme, the use of ambiguous words may lead to confuse the markers. Extract 12.1 illustrates a correct response from the candidates.

Extract 12.1

12:	Subjectivity of the Test. means testing method which lack moderated scoring method. This has got several limitation like difficult of distribution of score to each test item and also the influence of intellects/ factor during scoring like good hand writing.
	The factor which can affect subjectivity of a test is as follow:
	The use of model answer on the marking scheme. The subjectivity of the test can be affected by using of marking scheme. The marking scheme consist of model answer and also distribution of score to each test items. so this will help to affect the subjectivity of the test.
	Marking without looking at the student's name. The subjectivity of the test can also be affected by marking without looking at the student's name. this is because when someone looks at the name of student and when he/she noticed that it is her relative can be influenced to provide score which not required. so the marking without looking at the student's names can also help to affect subjectivity of the test.

12.	<p>Marking one question to all paper before moving to the next question. This means the marker should mark one question to all question paper provided before he/she to mark the next question. This will help he/she mark accurately rather than marking the whole paper. This will also help to affect the Subjectivity of a test.</p> <p>Avoid the influence of irrelevant factors: Irrelevant factor are like good hand writing which influence the marker so when scoring the score test the marker should avoid the influence of irrelevant factors. This may help in affecting the subjectivity of a test.</p> <p>Avoid Marking when you're tired. When the marker is tired is unable to mark accurately. So the marker are supposed to stop marking when are tired in order to minimize the subjectivity. So this will also help in affecting the subjectivity of a test.</p> <p>Generally, Subjectivity of a test can also be affected by adapting scoring method. During marking and also the marker should be in good mood when marking.</p>
-----	---

Extract 12.1 is from a candidate who correctly analysed five factors affecting subjectivity of a test.

Furthermore, the candidates who scored from 6 to 10 marks had some weaknesses in their responses. The samples of their responses were such as: *language difficulties; repetition of points; lack of knowledge and skills; ambiguity of the test; lack of enough time; inappropriate level of difficult; unclear instruction or direction; the long of the test*. Some of them wrote: *lack of knowledge and skills, ambiguity of the test*, instead of writing, the use of ambiguous words without elaboration showing that candidates had insufficient knowledge of the factors affecting subjectivity.

Detailed analysis shows that 80 candidates (20%) who scored from 0 to 5.5 marks had insufficient knowledge of the factors affecting subjectivity of a test. As the result, they gave wrong responses such as: *maturity; atmosphere; gender balance; secure the preparation; age of the learners; lack of table of specifications*. Some confused the concept of subjective test with other concepts. For example they wrote things such as: *atmosphere, gender balance lack of table of specification and generally*. Extract 12.2 is a sample of such poor responses from the candidates.

Extract 12.2

18.	Test is the series of question prepared by a teacher to measure the knowledge obtained after complete a course.
	The following are the factors affecting subjectivity of the test.
	Cheating opportunities: During examination construction the teacher ensure to have more than one test constructed when the first is cheated the other test is there to be done.
	Cost in production of copies, Administration and Transportation to the required places where is done at time.
	Learners disabilities: These where the test prepared does not consider on the nature of the learners to those who have disability and those who haven't.
	Construction of the test: The test constructed when it does not follow the good ways of test construction it led to confusion to the learners.
	Length of the test: The test length affect the subjectivity of the test because when is too long lead tiredness and also not to short must be average to ensure its subjectivity.

Extract 12.2 is from a candidate, who failed to analyse factors affecting subjectivity of a test as nature of scoring criteria, biased coverage of content, timing, Unclear test instructions, there are no pre-determined correct answers, The examinees are required to respond using their own words.

2.2.3 Question 13: Educational Assessment and Evaluation

This question instructed the candidates to explain five objectives of examinations in Tanzania. The total marks allocated to the question were 15 marks.

The question was optional and was thus attempted by only 852 candidates (93.9%). The analysis shows that 176 candidates (20.7%) scored from 10.5 to 14 marks; indicating a good performance. 661 candidates (77.6%) scored from 6 to 10 marks; indicating an average performance. The rest, 15 candidates (1.8%) scored from 0 to 5.5 marks, of which, 1 candidate (0.1%) scored the lowest 0 mark; indicating a poor performance. The general performance in the question was good considering that 837 candidates (98.3%) scored from 6 to 14 marks. Figure 14 presents the performance in the question.

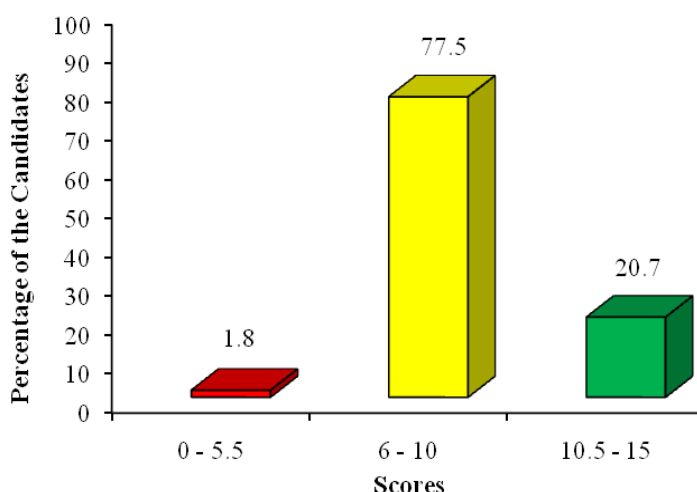


Figure 14: *The Candidates' Performance in Question 13*

The data analysis shows that 176 candidates (20.7%) who scored from 10.5 to 14, a good performance, provided relevant introduction and explained the five objectives of examinations in Tanzania satisfactorily. Example of their responses are such as: *assessing individual achievement in different education levels, used to select students for placement and further learning, examination used for certification, for informing policy makers and education stakeholders on education matters, helps to identify and address challenges encountering education sector*. Others wrote: *examinations help provide feedback to parents, help to improve education sector, increase status of education in the country,.* Others wrote: *examinations help to upgrade students, help to determine difficulties in students learning and plan for correct measures, help to encourage competition from the rank the candidate got*. Generally, the question was well attempted and was well

performed. Extract 13.1 portrays a correct sample from one of the candidates.

Extract 13.1

13.	<p>Examination is the instrument which is used in assessment in education. This examination is used in Tanzania because it has very correct feedback. So the following are the objectives of examination in Tanzania</p> <p>It used to motivate learners, in Tanzania examination is used as a way of motivating learners about what they learn, for example when the student studying the subject and there is no any examination can not study - hard rather than when the examination used.</p> <p>To provide certificate, also the examination in Tanzania is to reward the certificate for those who perform well in subject of a certain level of education such as the student of secondary school are certified after completing their subject and to perform well.</p> <p>To provide feedback to educational authorities, like teachers and ministry of education about the quality of teacher, in his/her position and the relevance of curriculum for the student so if can be some problem the improvement can be done.</p> <p>To group and to make selection, also education in Tanzania its objective is to group the learners according to the level of understanding such as talented students to be in their special schools like Tabora boys and also to make selection for those who perform well to proceed with further level of education.</p> <p>To identify learning difficulties, also examination if provided to students so as to</p>	
-----	--	--

13	to identify the problems and challenges which learners faced during the teaching and learning processes as to solve it in education.	
	There fore in other way education examination in Tanzania has several importance like to create competent learners, to get different professionalist like teachers and doctors.	

Extract 13.1 is from a candidate who satisfactorily explained the objectives of examinations in Tanzania.

Detailed analysis shows that 661 candidates (77.6%) who scored from 6 to 10 marks provided relevant introduction and relevant points but failed to elaborate them and conclude them well. Exemplary responses from the candidates in this category are such as: *examinations provide value of education, help to select learners for further studies, examinations help to motivate learners, to grade the students' performance and examinations help to evaluate education system*. Other candidates wrote: *to determine the level of country's education, to get competent expertises, to develop curriculum*.

Further, 15 candidates (1.8%) who scored poorly (from 0 to 5.5 marks) demonstrated several weaknesses in their responses. Some of them provided irrelevant introduction or mixed relevant and irrelevant points. Some of their irrelevant points were such as: *examinations must have a time limit, must have a section, must have instruction, and must have arranged in number*. These responses were incorrect and irrelevant because the candidate gave steps to consider when preparing/constructing the test, instead of explaining the objectives of examination in Tanzania. Further, some of their relevant points were: *examinations help to provide quality education, to identify and determine learning outcomes, it aimed at providing quality certificate, examination aim to give students employment, aimed at producing student who will serve in different sector in a country*. Extract 13.2 portrays such wrong responses from the candidates.

Extract 13.2

13.	Examination, Is the things which are in form of written where by this are used to test the performance or to see the performance of the learner. The following are the objectives of examinations in Tanzania which must have a time limite, the examining
	tion in Tanzania must have time where the examiners well should know how time or for what time may do that question which may help for them in arrange their time and answering without any problem.
	It must have a section, this is the another objectives of examinations in Tanzania where by this is due to the fact that this in this section may know the question which are what more explain and which they want the analyse only of the point. So that is the another objectives of examination in Tanzania.
	It must have instruction, this is the another objectives of examination in Tanzania where by this is due to the fact that the all exams in Tanzania must have the instruction which may help to know the things which are needed and which are not needed in examination system or process.
	It is arranged in number, this is the another examination objection where by the question are arranged in number which may create the examiners to know the question without any problem in the the examination.

Extract 13.2 is from a candidate who poorly explained the objectives of examinations in Tanzania.

2.2.4 Question 14: Qualities of Tests

This question instructed the candidates to explain five strategies to consider in ensuring test validity. Total marks allocated to the question were 15 marks.

The question was opted for by 609 candidates (67.1%), of which, 59 candidates (9.7%) scored from 10.5 to 13 marks; indicating a good performance, and 454 candidates (74.5%) scored from 6 to 10 marks; indicating an average performance. The rest 96 candidates (15.8%) scored from 1 to 5.5 marks; indicating a poor performance. The general performance in the question was good considering that 572 candidates (84.2%) scored from 6 to 13 marks. Figure 15 presents the candidates' performance in graphical form.

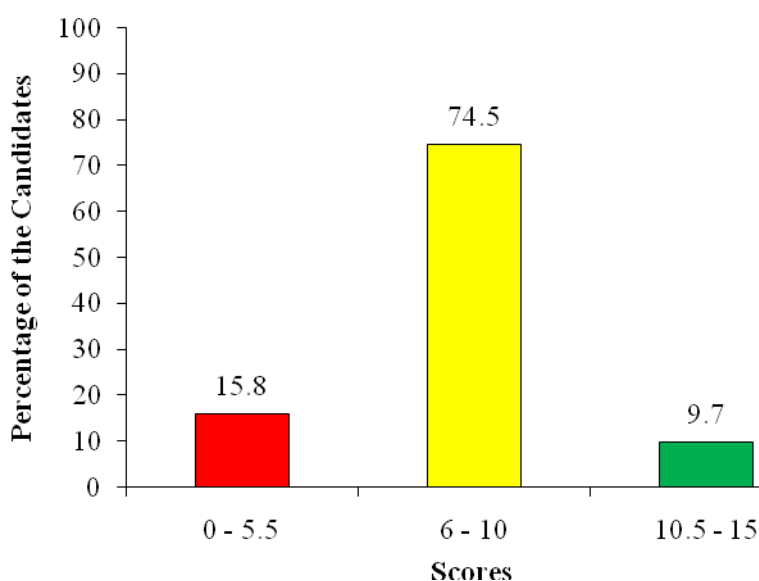


Figure 15: *The Candidates' Performance in Question 14*

The analysis indicates that 59 candidates (9.7%) who scored from 10.5 to 13 marks, a good performance, answered the question correctly and some of their points were relatively well presented. This is probably because of adequate knowledge of the topic from which the question was derived. These candidates satisfactorily explained five strategies to consider in ensuring test validity some with supportive examples. For example, one candidate pointed out: *language used should be well understood to most of*

the learners, to ensure all test items are of moderate or appropriate level of difficulty, the sentence should be out of ambiguity, the constructed questions should reflect the table of specification – the responses which are correct. Other candidates wrote: it enables the teacher to understand if all alternatives functioned as what the test intend, the test should be relevant according to level of understanding, test should use clear language, test should not be too short or too long but it should be of appropriate length according to the time set. Extract 14.1 illustrates good responses from the candidates.

Extract 14.1

14.	Test validity – This refers to the ability of the
	test to measure what was intended/supposed to be test
	ed for example face related validity, content related
	Validity and Concurrent related Validity. The following
	are the five strategies to consider in ensuring test validi-
	ty as follows:-
	To ensure test have clear instructions and directions
	this means that instrument on its face must resemble
	the content for example if the test contain Section A and B
	and 16 questions at the cover page also inside should
	have those content, hence test is valid.
	To ensure that test have the appropriate level
	of difficulties, they may be considered, so as to construct

the test that is challenging or moderate which match with the level of learners for example the content of the test should match with level of learners for example form one test should contain all concept covered in form one hence that test is likely to be valid.

To consider that the test is free from ambiguous words or terms that may confuse the learners, therefore it's advised to use simple and straightforward language and avoiding use of ambiguous because ambiguous statements have more than one meaning, hence test become valid.

To consider proper arrangement of the test items, this means that it should start from simple to complex questions always section A should start with Multiple choice, True and False, Matching items and Section B contain short answer and lastly section C contain essay questions therefore questions or test items should be well arranged.

To consider test length, test should not be too short or too length, means should be moderate in terms of length, because too short test will present few number of test items, therefore teachers are supposed to construct a test that is somehow length so as to ensure test validity.

Generally the teacher should be very carefully during constructing the test items, to avoid all factors which lead to lower validity of the test, so as to simplify easy answering of the questions during the test and hence maintaining test validity.

Extract 14.1 is from a candidate who satisfactorily explained strategies to consider in ensuring test validity.

Further analysis indicates that 454 candidates (74.5%) who scored from 6 to 10 marks, average, had inadequate knowledge of strategies to consider in ensuring test validity. The candidates also failed to focus on the demands of the question. Consequently, they provided fewer points or skipped some parts of the question. Some of these explained three or four correct points out of five needed points. The exemplary answers are: *language used must be clear, test must be conducted in conducive environment, students should be motivated, and time indicated must reflect maturity level of the learners.* Others wrote: *to ensure test administration, test should be specific, test should use simple language, the constructed items must relate with the taught content.* The points though partially correct are not supported with good explanations.

Moreover, 96 candidates (15.8%) who scored poorly (from 1 to 5.5 marks) explained the needed five strategies to consider in ensuring test validity correctly. For example, one candidate wrote: *clarity, fairness, validity, specificity, and applicability* which are not correct responses to the question. Another candidate wrote: *culture belief, attitude, absenteeism, maturity, and language.* The responses show that many of the candidates in this category lacked skills and knowledge of the subject matter. The general performance in the question was good (considering that 609 candidates (67.1%) scored from 6 to 13 marks). Extract 14.2 presents a satisfactory response from the candidates.

Extract 14.2

14.	Test, refer to a series of designed question in a particular subject, a test should be validity, after following the strategies that following bellow, To avoid the length of a question, a question should not be too long, in order to have a validity test the length of test should be considered. To avoid the complex language that can confuse the learners, simple language should be considerable when you construct a validity test. It should be relevant to the respondent, the level of learners should be considered when constructing the test hence, to be validity. It should have a key term as a model of position where the objective of question sticked, for example explain the factors for the Meiji war, the key term is "Factors".	
-----	--	--

Extract 14.2 is from a candidate who failed to explain strategies to consider in ensuring test validity as test items are clear and free from ambiguity, clear directions are given on what and how to respond, Language used is well understood, time allocated for the test is sufficient etc.

2.2.5 Question 15: Analysis and Interpretations of the Test Results

This question instructed the candidates to evaluate five roles of item analysis in curriculum development. Scores allocated to the question was 15 marks in total.

This was the most skipped question [since 728 candidates (80.3%) did not attempt the question]. It was attempted by only 179 candidates (100%), of which, 32 candidates (17.9%) scored from 10.5 to 13 marks; indicating a good performance. 127 candidates (70.9%) scored from 6 to 10 marks; indicating an average performance. The rest 20 candidates (11.2%) scored from 0 to 5.5 marks, of which, 1 candidate (0.1%) scored 0 mark; indicating a poor performance. The general performance in the question was good considering that 159 candidates (88.8%) scored from 6 to 13 marks. Figure 16 presents the candidates' performance in question 15.

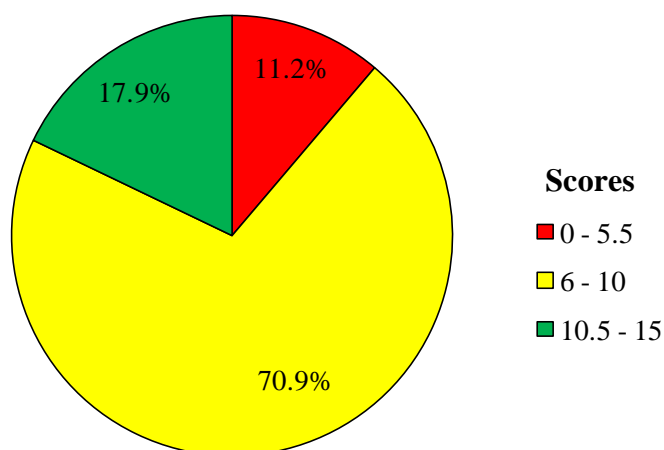


Figure 16: *The Candidates' Performance in Question 15*

The analysis shows that 32 candidates (17.9%) who scored from 10.5 to 13 marks, a good performance, had good organization of points and evaluated the five roles of item analysis in curriculum development quite well. For example, one candidate wrote: *item analysis provides feedback developers on level of difficulties for some contents, the results from item analysis help to determine some specific areas that require support to improve teaching and learning, item analysis helps to know higher and lower achievers on the given test, it helps the analyst to know the way different items function in the test.* These responses were correct and well elaborated. Others wrote: *helps to develop new knowledge of students, provide facts upon students' scores, it is used in educational evaluation, the analysis helps to enable curriculum developers to address the teaching and learning difficulties, helps to improve learners' achievement in the future.* Extract 15.1 presents a satisfactory from one of the candidates.

Extract 15.1

15: Item analysis is the process of evaluating a quality of the test. The teacher make analysis during scoring and grade the tasks of the learner in order to determine the too easy and too hard of the test. The following are the roles of item analysis in curriculum development.

Help to judge norm-reference if item test are too easy or too hard: Item analysis evaluate the norm-reference in order to understand if the item test are too hard or too easy, which emphasize to change the curriculum, if the item test too easy.

Help to discriminate the high scores and low scores of item test: Item analysis help to evaluate the high score and low score in the test or examination provided. Determine the low scores and high score after analysis take place, the curriculum development start to be active.

Help to determine the test item for improve teaching and learning: Item analysis to evaluating the test item if too hard or too easy, so make choose to provide the appropriate item test, make changes in curriculum to make normal item test.

Help to determine if the alternative item test function: When the test provided or examination there are alternative item test provided in the course outline. So the item analysis help to evaluate if alternatives are function, in order to make clearly.

	Help to evaluate test item: Item analysis	
	will make evaluation of test item. example	
	when test provided to the learner after mark,	
	scoring and grade, the teacher make item	
	analysis to the item test or test/examination	
	Item analysis: Help to determine the	
	reliability of the test item, stability of the	
	test item, quality of the test item. We have	
	to analysis the task or examination or test in	
	order to qualify the item test	

Extract 15.1 is from a candidate who satisfactorily evaluated roles of item analysis in curriculum development

Further analysis show that 127 candidates (70.9%) who scored from 6 to 10 marks had average understanding of the roles of item analysis in curriculum development. They would point out the roles of item analysis but fail to provide a connection between it and the curriculum development. As a result, they could not score above 10 marks. Others provided only a few correct roles. For example, one candidate wrote: *it promotes the foundation for curriculum development, direct sources of curriculum content, it leads of picture of curriculum, enhancing effectiveness of curriculum, it suggests source of curriculum*. All these points were shallow and had some elements of repetition.

In addition, 20 candidates (11.2%) who scored poorly (from 0 to 5.5 marks) had insufficient knowledge of roles of item analysis in curriculum development. Most of them started well but ended poorly. Others mixed relevant with irrelevant points, for instance one candidate gave the following responses: *it helps a teacher to set a very difficult examination; it is easy to put marks; the students cannot cheat properly, to identify difficulties in learning, identify higher and lower achievers*. That is, the last two points are correct but the first three points are not correct. Extract 15.2 a sample of a poor responses from one of the candidates.

Extract 15.2

15	<p>The Curriculum development supported by the presence Presence of Ideology of the Country. Things to Consider in the Process of Curriculum of developing are most important because it gives the how to drive our economy Example in Tanzania there is ideology of Education and self reliance, so it is better to Consider.</p> <p>The item analysis helps to know the psychology of the of the learners and individual, the the teacher they use this types of to know the learners needs, the Psychology of the learners is very important in analysis of the Curriculum.</p> <p>Also the item analysis helps to know Cultural needs of the society. In analysis the Curriculum developers consider the Cultural needs because it may become the Causative agent for some areas.</p> <p>Therefore the Curriculum developers are the very important figure in the education system because they show and direct the student or learner what to do and what to Consider.</p>
----	---

Extract 15.2 is from a candidate who unsatisfactorily evaluated roles of item analysis in curriculum development.

2.2.6 Question 16: Educational Measurement

This question instructed the candidates to identify five characteristics of norm-referenced measurement. Scores allocated to the question was 15 marks in total.

The question was attempted by 672 candidates (74.1%), of which, 245 candidates (36.5%) managed to score from 10.5 to 14 marks; indicating a good performance. 390 candidates (58.0%) scored from 6 to 10 marks; indicating an average performance. The rest 37 candidates (5.5%) from 0 to 5.5 marks, of which, 1 candidate (0.1%) scored 0 mark; indicating poor performance. Generally, the performance in the question was good – considering that 635 candidates (94.5%) scored from 6 to 14 marks. Figure 17 presents the candidates' performance in the question.

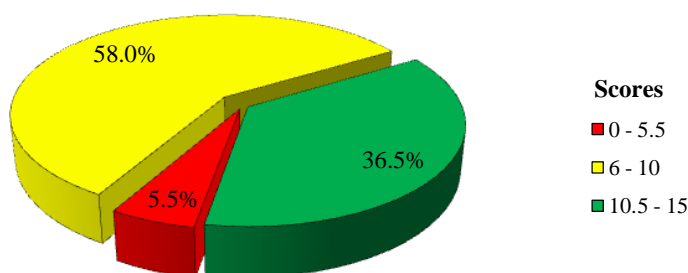


Figure 17: *The Candidates' Performance in Question 16*

Detailed analysis reveals that 245 candidates (36.5%) who scored from 10.5 to 14 marks, a good performance, had high understanding of the question. As a result, they satisfactorily identified five characteristics of norm-referenced measurement. They also showed high ability of associating points asked on question one of the paper. Further, they ensured coherence from introductory part of the question to the end point. What is more, they clearly identified the five characteristics of norm-referenced measurement are such as: *table of specifications is used when constructing the items, sometimes normal distribution curve is used in assessing learners' achievement, students' scores are the bases of judging position between individuals..* Some of the candidates wrote: *it leads to discriminations among students; it is characterized by competitions; it is used in predicting future performance; it characterized by human biases; it is characterized by the teacher to know the problem existing among the*

learners. Another candidate wrote: it is used to motivate learners; it involves large domain of measurements; it favours average item of difficulties. Extract 16.1 is a sample of good responses from the candidates.

Extract 16.1

16	<p>Norm-referenced Measurement is the type of measurement where by the student performance are being compared with another student results who took the same test. The following are the characteristics of norm referenced Measurement.</p> <p>Table of specification is commonly used, in norm reference measurement when you compare student results a table of specification is used to show the relationship of the objective and content found in the test item.</p> <p>Interpretation require clear defined group, this is another characteristic of norm referenced measurement when you compare student performance interpretation of the test result must be clear stated.</p> <p>Favour item difficulty, this is another characteristic of norm referenced measurement when comparing student result it favour the percentage of items found in the test.</p> <p>Typically cover large domain, norm referenced measurement when comparing student performance it cover large cognitive domain of the content of items in the test.</p> <p>It discriminate item analysis, this is another characteristic of norm referenced measurement where by it does not look on the effectiveness of the test result on how the student have been respond.</p> <p>Therefore, norm-referenced measurement has been used to place student for further studies.</p>	
----	---	--

Extract 16.1 is from a candidate who satisfactorily identified characteristics of norm-referenced measurement.

Further, analysis reveals that 390 candidates (58.0%) who scored averagely, from 6 to 10 marks, provided good introduction, conclusion and correct points, though with partial explanations. Others fragmented one points into several points unnecessarily. The following extract is a case in point: *compare the performance of an individual, characterized by relative decisions made to different students, norm referenced measurement used to select students to next level of education, it is used to screen students for further studies*. This candidate identified the characteristics of norm referenced measurement correctly in point one and two but, confused it with the advantages of norm-referenced measurement in point three and four.

In the analysis, 37 candidates (5.5%) who scored poorly (from 0 to 5.5 marks) identified five the characteristics of criterion - referenced measurements instead of those of norm-referenced measurement. For instance, one wrote: *norm referenced measurement is skills oriented, it covers small domain range, it avoids discrimination amongst students, and it deals with affective rather than cognitive and psychomotor domains*. The varied responses from this candidate denote uncertainty about the concept in question. Some of them failed to provide a good introduction on norm-referenced measurement. For example, one candidate wrote: *norm-referenced measurement is the type of educational measurement where by performance of students cannot be compared, it is used for outlying students' mastering of contents, It is not good for comparison of students and schools, is used to punish students who do not study hard*. Extract 16.2 is a sample of unsatisfactory response to the question.

Extract 16.2

16	Norm-referenced Measurement is the type of Measurement which used to measure the performance of the students who took the same test without compare them. The following are the characteristics of Norm referenced Measurement.
	It have prescribed data where by a performance of an individual are not compared with another person.
	Dealing with measuring the performance of students without compare them, This means that it deals with individual assessment.
	It provide the general view about the performance of students. This because it deals with measuring performance of student without compare them.
	It find to know the performance of an individual rather than a group of people which can be compared.
	Also it does not compare the performance of the students even if they were took a same test. This mean that Norm referenced Measure

Extract 16.2 is from of a candidate who unsatisfactorily identified characteristics of norm-referenced measurement.

3.0 PERFORMANCE OF CANDIDATES IN EACH TOPIC

The Educational Research, Measurement and Evaluation examination comprised of seven topics into which the examination questions were organised. The analysis of candidates performance in each topic shows that, candidates had good performance in the following topics; Educational Assessment and Evaluation (98.2%), Educational Measurement (91.7%), Qualities of Test (82.4%) and Test Construction (75%). On the other hand, the candidates had an average performance in the topics of Educational Research (69.4%), Analysis and Interpretation of Test Result (63.8%) and Assessing Achievement (50.8%). For detailed information see Appendix A.

4.0 CONCLUSION

Statistical data analysis for each question shows that the candidates' overall performance in Educational Research Measurement and Evaluation subject for Diploma in Secondary Education Examination (DSEE) in 2018 was good. The analysis shows that the candidates' good performance was caused by their ability to identify the demands of questions, sufficient knowledge of the subject matter and computation skills. Only a few candidates revealed low ability in these areas; which earned them low marks.

5.0 RECOMMENDATIONS

Based on the observations made in this Candidates' Item Response Analysis (CIRA) report, the following are recommended to improve the performance of prospective candidates in this subject:

- (a) Student and tutors should be encouraged to use English so as to improve their language skills. This can be done through practicing speaking English inside and outside the classroom, during discussion as well as essay writing competitions.
- (b) More effort is needed in the topics of Analysis and Interpretation of Test Result, and Assessing Achievement. A topic such as Analysis and Interpretation of Test Results should be taught through many exercises and tests because it requires mastery of mathematical operations.
- (c) Tutors should guide students in studying all topics in the Educational Measurement and Evaluation syllabus to equip them

with knowledge and skills needed for answering examination questions.

- (d) Candidates should be guided on how to identify the demands of questions. This could be done through giving them enough homework, assignment, tests, and inter-college examinations. Equally, tutors should provide timely feedback to teacher students after such tests.

Appendix A

THE PERFORMANCE OF THE CANDIDATES IN EACH TOPIC DSEE 2018

S/N	Topics Examined	DSEE 2018		Average Performance per Topic in %	Remarks
		Questions Number	Percentage of the Candidates who Scored an Average of 40 Percent and Above.		
1	Educational assessment and Evaluation	13 3 4	98.2 63.5 84.2	98.2	Good
2	Educational Measurement	1 2 16	92.3 88.4 94.5	91.7	Good
3	Qualities of Test	9 14	85.2 79.5	82.4	Good
4	Test Construction	8 12	76.7 80.0	75	Good
5	Educational Research	3 4	63.5 84.2	69.4	Average
6	Analysis and interpretation of test result	5 10 11 15	63.8 64.5 38.2 88.8	63.8	Average
7	Assessing Achievement	6 7	97.1 4.5	50.8	Average
General performance				75.9	Good

From the Appendix A, the overall performance in all topics was good with the average of 75.9%. The topic of *Educational assessment and Evaluation* was the highly performed followed by topics of *Educational Measurement*, *Qualities of Test*, *Test Construction*, *Educational Research*, *Analysis and interpretation of test result* and the last with an average performance was *Assessing Achievement*.

