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

CANDIDATES' ITEM RESPONSE ANALYSIS REPORT ON THE DIPLOMA IN SECONDARY EDUCATION EXAMINATION (DSEE) 2022

EDUCATIONAL RESEARCH, MEASUREMENT AND EVALUATION

THE UNITED REPUBLIC OF TANZANIA

MINISTRY OF EDUCATION, SCIENCE AND TECHNOLOGY

NATIONAL EXAMINATIONS COUNCIL OF TANZANIA

CANDIDATES' ITEM RESPONSE ANALYSIS REPORT ON THE DIPLOMA IN SECONDARY EDUCATION EXAMINATION (DSEE) 2022

762 EDUCATIONAL RESEARCH, MEASUREMENT AND EVALUATION

TABLE OF CONTENTS

FOREWORD

The National Examinations Council of Tanzania (NECTA) is pleased to issue the Candidates' Item Response Analysis Report on the Diploma in Secondary Education Examination (DSEE) 2022 in Educational Research, Measurement and Evaluation subject. This report provides feedback to student-teachers, tutors, policy makers and the public in general about the performance of the candidates. Basically, the candidates' response to the examination questions indicate what the education system was able/ unable to offer to student-teachers in their Diploma in Secondary Education course.

The general performance of the candidates in Educational Research, Measurement and Evaluation subject was good. The report shows the factors which contributed to the majority of candidates' ability to answer the examination questions correctly and score high marks. The factors included; ability to understand the requirement of the questions, sufficient knowledge of the subject matter, good mathematical skills and correct application of the principles of Educational Research, Measurement and Evaluation. However, few candidates with low marks showed lack of such qualities.

It is hoped that the suggestions and recommendations provided in this report will enable various education stakeholders to take proper measures which will strengthen teaching and learning process. The interventions will enable the student-teachers to master the required skills and knowledge hence improve academic performance in the future examinations administered by the Council.

Lastly, the Council is grateful to all examination officers, examiners and all other staff members who participated in the preparation of this report.
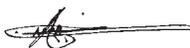
Athumani S. Amasi
EXECUTIVE SECRETARY

1.0   INTRODUCTION

This report analyses the candidates' item responses in Educational Research, Measurement and Evaluation examination for the year 2022. It pinpoints the strength and weaknesses of the candidates in responding to each question of this examination.

The Educational Research, Measurement and Evaluation examination paper for DSEE 2022 covered the 2009 syllabus and was set based on the 2021 Examinations Format. The examination paper consisted of sections A and B with a total of fourteen (14) questions. Section A consisted of ten (10) short-answer questions, each carrying 4 marks. Sections B had four (4) essay/structured questions, each carrying 15 marks. The candidates were required to answer all questions from both sections.

The report highlights what candidates were required to do in each question and evaluates the strengths and weaknesses of candidates' responses. Furthermore, it provides a statistical analysis of candidates' performance and sample of extracts which illustrate both candidates' good and weak responses.

A total of 4,423 candidates sat for the Examination in which 99.11 per cent passed while the remaining 0.89 per cent failed. The performance in 2022 has increased by 1.14 per cent as compared to the performance in 2021, whereby only 97.97 per cent passed. The analysis of the candidates' performance in 2022 with different grades as compared to the year 2021 is summarized in Table 1.

Table 1: Comparison of Candidates' Performance in Grades and Percentages for the year 2021 and 2022

Year	Sat	Total Passed	Grade				
			A	**B**	**C**	**D**	**F**
2022	4423	4326	6	427	2789	1104	39
		99.11%	0.14%	9.78%	63.89%	25.29%	0.89%
2021	2095	2029	2	215	1291	521	42
		97.97%	0.09%	10.4%	62.3%	25.2%	2.0%

Table 1 indicates that majority of the candidates scored grade C and D for the two years consecutively. On the other hand, the candidates who scored grade F in 2022 decreased by 1.11 per cent as compared to the year 2021. The next part analyses the performance of the candidates on each question:

2.0 ANALYSIS OF THE CANDIDATES' PERFORMANCE ON EACH QUESTION

The candidates' performance on each question is analysed by indicating the competencies tested and the requirement of each question. The analysis shows the percentage of the candidates who attempted each question, those with good, average and weak performance based on their responses. The performance classifications is as follows: 70 - 100 per cent is good represented in this report by the green colour; 40 - 69 per cent is average denoted by yellow; and 0 - 39 per cent is weak performance and is marked by red.

2.1 SECTION A: SHORT ANSWER QUESTIONS

Section A comprised of 10 short answer questions which carried four (4) marks each, making a total of forty (40) marks. The questions were composed from six (6) topics; *Test Construction, Educational Research, Assessing Achievement, Educational Measurement, Analysis and Interpretation of Test Results* and *Qualities of Tests*. The candidates' response analysis for each question is as follows:

2.1.1 Question 1: Test Construction

The question measured the candidates' knowledge about scoring procedures of objective tests. The candidates were required to explain four procedures of scoring objective test items. The analysis of candidates' performance reveals that most of the candidates (86.8%) scored low marks (0.0 to 1.5), while a few (13.2%) scored 2.0 marks and above. Table 2 illustrates:

Table 2: Candidates' Scores Range and Percentage for Question 1

Scores Range	Description	Percentage of Candidates
0.0 - 1.5	Weak	86.8
2.0 - 2.5	Average	10.4
3.0 - 4.0	Good	2.8

The analysis of the candidates' responses for question 1 reveals that those who scored low marks (0.0 -1.5) had inadequate knowledge on the concept of scoring procedures. This was evident through their responses provided. For instance, some of the candidates confused the concept of scoring procedures of objective tests items with either the rules used in the construction of test items or procedures of test administration. For example,

one candidate outlined the responses related to test administration instead of scoring procedures as: *(i) to ensure the minimum administration of test (ii) to discourage cheating (iii) to avoid generation of test item.* Another candidate wrote the rules used to construct test items instead of scoring procedures as to: *(i) determine the purpose of the test (ii) determine the learning outcomes (iii) construct the table of specification (iv) construct the relevant test items.* Extract 1.1 shows incorrect responses from one of the candidates.

Extract 1.1: A sample of incorrect responses to question 1.

As clearly illustrated in Extract 1.1, the candidate outlined the procedures for processing test scores instead of scoring procedures of objective test items.

In other instances, some of the candidates scored only one mark because they provided only one correct procedure of scoring objective such as: *to make a scoring key which contains correct answer*, *in this case each correct answer was counted as one point*. Other candidates gave incomplete responses such as: *(i) to mark all test items*, *(ii) to read carefully question and answer*.

Despite the weak performance of most candidates on this question, 13.2 per cent of the candidates' scores ranged from 2.0 to 4.0 marks. These candidates demonstrated good knowledge on the concept of scoring procedures of objective test items. Out of these, 10.4 per cent scored average marks (2.0 to 2.5) because they explained correctly two out of four procedures and missed the other two. A few candidates (0.8%) were able to explain correctly all four procedures of scoring objective tests and scored

3.0 to 4.0 marks. Extract 1.2 shows responses by a candidate who scored high marks

Four procedure for scoring objective questions
(i) Using scoring key/guide which contain all relevant answers. This will guide him/her to save time when scoring.

(ii) Each correct answer contain one mark. This also is the procedure for objective whereby it must have one mark in each answer given.

(iii) Mark each correct answer and draw a line within a answer. This will help him/her to avoid student who will lies him/her about an answer.

(iv) Correction for guessing, This will ensure reliability of the test when ho/she will correct guessing.

Extract 1.2: A sample of good responses to question 1.

2.1.2 Question 2: Educational Research

The question measured the candidates' knowledge in the concept of basic research. They were required to describe four characteristics of basic research. The performance of the candidates in this question was good since 70.2 per cent scored from 2.0 to 4.0 marks. These data are summarised in Figure 1.
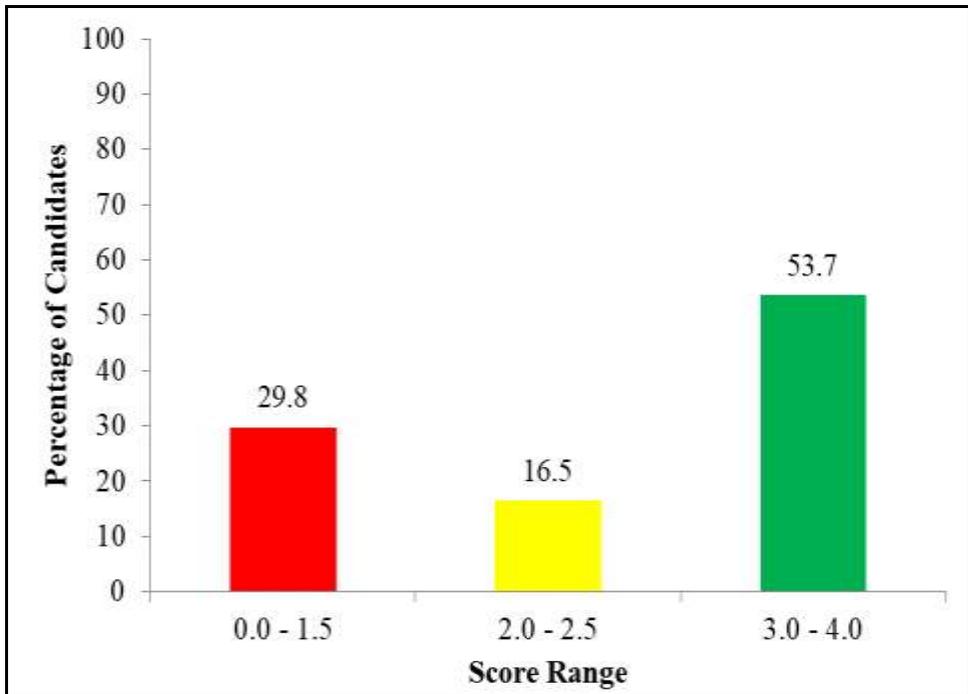
Figure 1: *Candidates' Performance on Question 2*

Figure 1 shows that 53.7 per cent of the candidates scored from 3.0 to 4.0 marks, 16.5 per cent scored from 2.0 to 2.5 marks and 29.8 scored from 0.0 to 1.5 marks.

The 53.7 per cent of the candidates who scored high marks (3.0 to 4.0) were aware of the characteristics of basic research which include; *it is analytical in nature, systematic* and *primarily concerned with the expansion of knowledge.* In addition, they knew that psychologists prefer basic research *because it develops theories and principles which may be used in further research*. Therefore, the candidates provided most of the correct responses to the question. Extract 2.1 is a sample of correct responses to question 2 from one of the candidates:

Extract 2.1: A sample of correct responses to question 2.

In extract 2.1, the candidate correctly described 3 out of 4 characteristics of basic research and scored 3 marks accordingly.

The candidates who scored from 2.0 to 2.5 marks, correctly described only two out of four characteristics of basic research while attempting to guess the other characteristics. Other candidates gave two characteristic correct and skipped the other two. This was likely caused by candidates' partial knowledge of the concept of basic research.

Most of the candidates who scored from 1.0 to 1.5 marks mixed up some characteristics of the basic research with action research. However, they gave one to two correct responses out of four as per question demand. The candidates who scored a zero mark described all four characteristics which did not relate to basic research. For example, one candidate described general features of a quantitative research approach such as *use of stable and controlled settings; it is a key role of numerical data and present finding in tables and graphs*. Another candidate described the features of a research topic which are: *"should consider ethical issues when stated, should be manageable,* and *it has theoretical or practical significance"* instead of characteristics of basic research which are: *"employs careful sampling procedures, deals with development of theories* and *leads to development of knowledge."*

In addition, some of these candidates gave the components of the research such as *research problem*, *literature review* and *methodology* instead of the characteristics of basic research. Other candidates pointed out the characteristics of qualitative research approach instead of the characteristics

of basic research. Extract 1.2 is an example of incorrect responses from one of the candidates:

Extract 2.2: A sample of candidates' incorrect responses to question 2.

In Extract 2.2, the candidate outlined the components of essay writing report instead of characteristics of basic research.

2.1.3 Question 3: Assessing Achievement

The question measured the candidates' knowledge about behavioural assessment techniques. The question required the candidates to explain four learning assessment techniques. The analysis of their performance shows that the general performance was weak since majority (70.3%) of candidates scored below 1.5 marks while only 29.7 per cent scored above 1.5 marks. The overall performance is summarized in Figure 2:
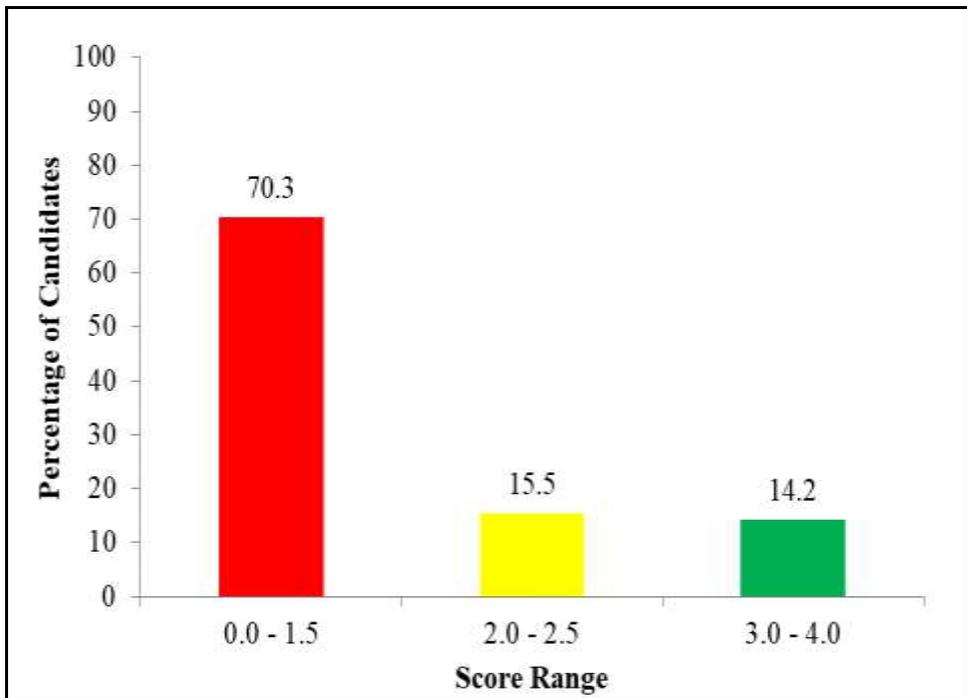
Figure 2: *Candidates' Performance on Question 3*

The statistical data in Figure 2 indicates that 70.3 per cent of the candidates scored from 0.0 to 1.5 marks (weak), 15.5 per cent scored 2.0 to 2.5 (average), only and 14.2 per cent scored from 3.0 to 4.0 marks (good performance).

The candidates with weak performance provided incorrect responses about behavioral assessment techniques. Most of them provided responses related to either categories of educational tests or types of educational measurement such as: *objective test and subjective test* and *criterion referenced measurement and norm-referenced measurement* respectively. Moreover, other candidates gave one correct response and other irrelevant responses to the question. For example, one candidate explained four behavioral assessment techniques as; *group work*, *think pair and share*, *brainstorming and role play*. The candidate's responses allied to teaching techniques but not behavioral assessment techniques. The other candidates described types of evaluation such as; *formative*, *summative*, *diagnostic and placement* instead of behavioural assessment techniques. The major reason for their weak performance was lack of knowledge on the topic of assessing achievement particularly on behavioral assessment techniques. Extract 3.1 is a sample of incorrect responses from one of the candidates.

9

> 3
>
> i - Affective - this deal with the feeling of a person
>
> ii - Cognitive - this deals with the behaviour of brain of a person
>
> iii - Psychomotor - This deal with sense organs So a learner learn through it sense

Extract 3.1: A sample of the candidates' incorrect responses to question 3.

Extract 3.1 shows that, the candidate wrote three Bloom's Taxonomy of learning domain instead of explaining behavioral assessment techniques as per question demand.

The analysis of the candidates' responses reveals that those who scored average marks (15.5%) had partial knowledge about the concept tested. Their responses had mixture of correct and incorrect explanations on behavioral assessment techniques. Furthermore, 14.2 per cent of the candidates who scored high (3.0 to 4.0) marks were competent about behavioral assessment techniques and adhered to question requirement. They provided about three to four appropriate learning behavioral assessment techniques which are *checklist*, *rating scale*, *anecdotal records* and *contract*. Extract 3.2 is a sample of the correct responses from one of the candidates.

Extract 3.2: A sample of good responses to question 3.

2.1.4 Question 4: Educational Research

The question measured the candidates' knowledge in the data collection methods. The candidates were required to explain the reasons for choosing questionnaire method as a data collection tool. The overall performance in this question was average as 68.6 per cent of the candidates scored from 2.0 to 4.0 marks. Figure 3 illustrates the candidates' performance.

11
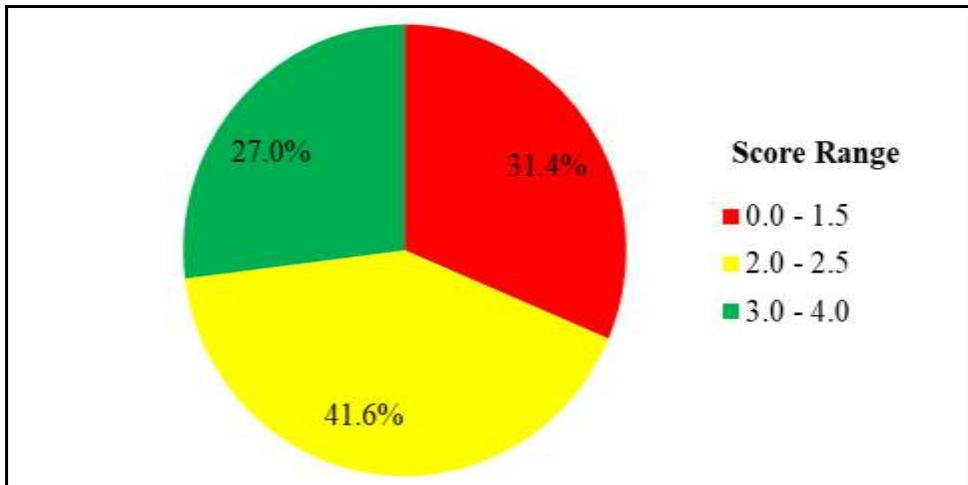
Figure 3: *Candidates' Performance for Question 4*

Figure 3 shows that 27.0 per cent of the candidates scored 3.0 to 4.0 marks, 41.6 per cent scored from 2.0 to 2.5 marks and the rest (31.4%) scored from 0.0 to 1.5 marks.

The candidates who scored above 2.5 marks were knowledgeable about the concept of questionnaire as data collection tool. They were aware that questionnaire is preferred by researchers because: *it enables the respondents to express their feeling in their own words*; *it is less cost and relatively economical*; and *widely spread geographical ranges*. Extract 4.1 is a sample of responses from one of the competent candidates in the tested concept.

Extract 4.1: A sample of the candidates' correct responses to question 4.

Despite the average performance for Question 4, a further analysis indicates that 31.4 per cent of the candidates scored below 2.0 marks. These candidates demonstrated low knowledge in the concept of questionnaire as data collection tool. For example, one candidate wrote advantages of data collection tools as: *It helps to develop the understanding and it increases the critical thinking of learners*. This response indicates that, the candidate was unaware that data collection tool does not relate with teaching methods which are used in teaching and learning process. Another candidate wrote: *questionnaire is the method which is based on face to face between a researchers and respondents*. The candidate was not aware that an interview collection tool involves oral interaction between interviewer and interviewee while a questionnaire method is based on written information and not verbal. Extract 4.2 is incorrect responses from one of the candidates.'

4	
(ii)	It used to Meansure the learners ability
(iii)	It help to understand the dificult face learers
(iii)	It help to Meansure learners Performance
(iv)	It help to determine the learnig outcome to the learners

Extract 4.2: A sample of the candidates' incorrect responses to question 4.

In Extract 4.2, the candidate incorrectly listed some importance of educational measurement and assessment instead of the reasons for researcher to choose questionnaire as a method of data collection.

2.1.5 Question 5: Educational Measurement

The question assessed the candidates' competencies in categorising the given statements into respective types of educational measurement. The statements were:
(i)   Maria is the first student in the class of 80 students.
(ii)   Ali can define correctly all measurement terms.
(iii)   A student has a better understanding of questions five but poor understanding in question two.
(iv)   Kulwa performed above the average while Doto performed below the average in the classroom test.

The overall candidates' performance was good since 72.0 per cent scored 2.0 marks and above. The overall candidates' performance is illustrated in Figure 4.
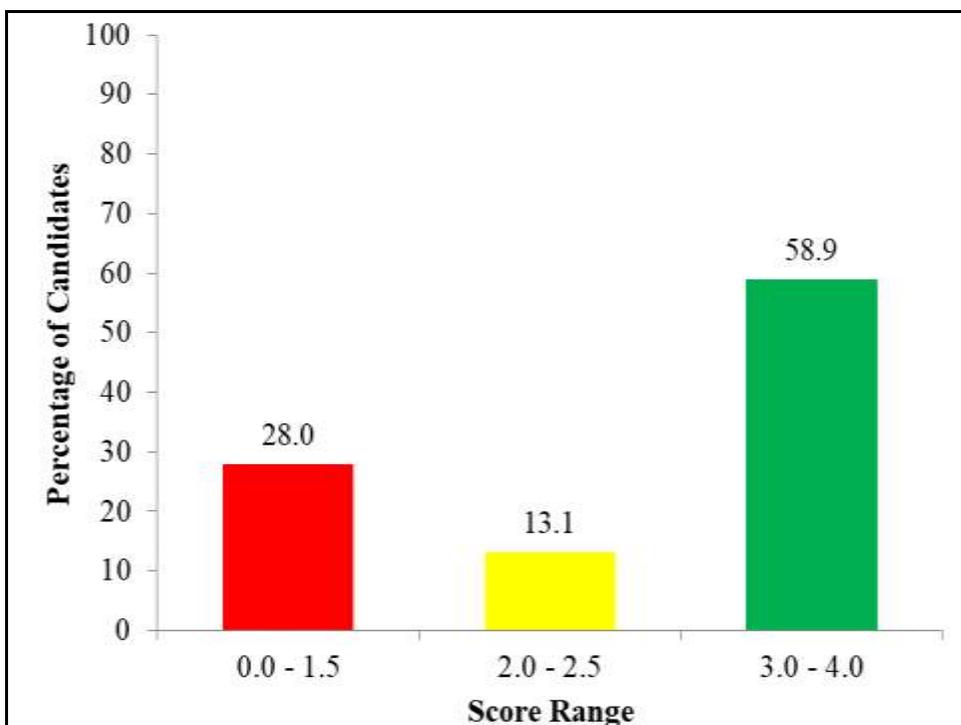
Figure 4: *Candidates' Performance on Question 5*

Figure 4 indicates that, 58.9 per cent scored from 3.0 to 4.0 marks, 13.1 from 2.0 to 2.5 marks and 28.0 from 0.0 to 1.5 marks.

On the one hand, the candidates (58.9%) who scored high marks (from 3.0 to 4.0) were able to categorise the given statements into their perspective types of educational measurement. These candidates classified correctly all or most of the given statements into their respective types of educational measurement as either *norm referenced measurement* or *criterion referenced measurement*. Extract 5.1 is a sample of the correct responses from one of the candidates.

Extract 5.1: A sample of the candidates' correct responses to question 5.

On the other hand, the candidates who scored low marks (0.0 to 1.5), lacked competence/ ability to categorize the given statement on norm and criterion referenced measurement. For example, one candidate wrote: *formative measurement, assessment measurement, summative measurement and diagnostic measurement*. These candidate's responses reflected the types of evaluation and not categories of measurement. Another candidate provided types of objective tests such as; *true-false, completion items, matching items and multiple-choice items* instead types of educational measurement which are *norm* and *criterion referenced measurements*. Extract 5.2 is a sample of the incorrect responses from one of the candidates.

Extract 5.2: A sample of the candidates' incorrect responses to question 5.

In Extract 5.2 the candidate outlined four scales of measurement instead of categorasing the educational measurement as required.

2.1.6 Question 6: Analysis and Interpretation of Test Results

The question measured the candidates' competence on item analysis. The candidates where required to differentiate the main features of the concept "discrimination index" and "difficult index." The performance was generally weak as 64.4 per cent of the candidates scored below 2.0 marks. The analysis of their performance is shown in Figure 5:
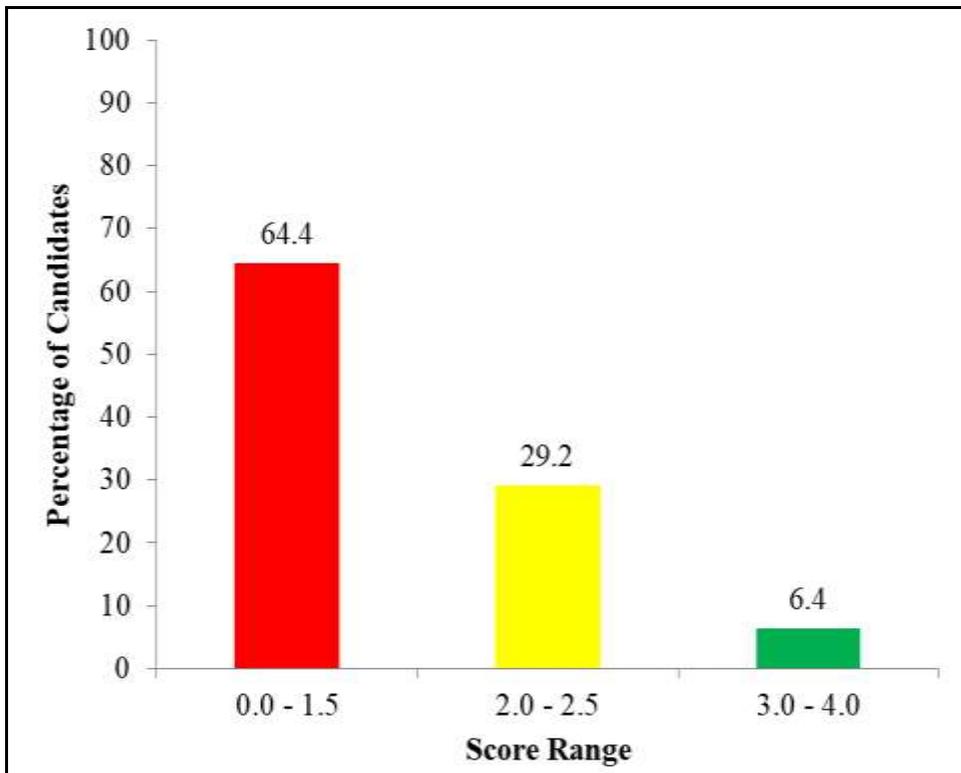
Figure 5: *Candidates' Performance on Question 6*

Figure 5 indicates that 29. 2 per cent of the candidates scored from 2.0 to 2.5 marks, 64.4 per cent scored from 0.0 to 1.5 marks and only 6.4 per cent scored 3.0 to 4.0 marks. Thus, the analysis of the candidates' responses showed that the candidates who scored below average (0.0 - 1.5) failed to understand the key concepts of item analysis. Hence failed to state the difference between discrimination index and difficult index. For example, one candidate wrote; *"discrimination index is the one which has no numerical value while difficult index as the one with numerical value."* The other candidate suggested that "*difficult index is expressed by mean of items while discrimination index is expressed by standard deviation*." In addition, some of the candidates used interchangeably the features of difficult index with those of discrimination index. For example, one of them stated that "*difficult index distinguishes high scores achievers from low score achievers while discrimination index determines the effectiveness of an item in a particular test*." Their responses show that they lacked knowledge about discrimination index and difficult index. These candidates failed to realise that, if the *discrimination index* is $d \geq 0.4$ *means good item* and $d < 0.19$ *means bad item (the higher the value of d, the more effective the item*

17

is). They were supposed to understand that, if percentage of the value of *difficult index* is in the range of 0 - 21; 30 - 39; 40 - 60 and 61 - 100 *means the item is very difficult, difficult, average difficult and easy item respectively*. Extract 6.1 represents one of the weak responses in this question.

6	
	i) Discrimination Index are those answer Which are In Correctly but difficult Index are those answer Which are the Same Within the one questions
	ii) Discrimination Index are Distracter While difficult Index Can be the Key or distracter.

Extract 6.1: A sample of incorrect responses to question 6.

In Extract 6.1, the candidate provided incorrect features that differentiate discrimination index from difficult index. She/he wrote about distractors and answers which are related to multiple choice items and not the item analysis.

However, 29.2 per cent of the candidates who scored average marks gave correctly a half of the required responses of tested concepts. Some of these candidates either stated correctly one feature that differentiate discrimination index from difficult index.

Further analysis reveals that some of the candidates with high marks (6.4%) were able to differentiate discrimination index and difficult index based on features such as: *distinguishing high score doers and low score doers* and *effectiveness of an item in a particular test* respectively. Generally, their responses imply that they had mastered the features of both discrimination and difficult indices. Extract 6.2 shows a sample of good responses from one of the candidates.

6		
	Discrimination Index	difficult index
	(i) It is Used to identify the highest and lowest scores of the learner in a given test Items	- Used to show how the difficult or easier of the the last item is.
	(ii) Used to identify at Which group of scores the learners Were performed better and Worse good and bad	- It enable us to know Which Item test were performed better good and bad in a given test Items.

Extract 6.2: A sample of correct responses to question 6.

2.1.7   Question 7: Educational Research

The question measured the candidates' competence in explaining the data collection tools based on educational research. The question required the candidates to give brief explanation on (a) Interview and (b) Focus group discussion as a research data collection tool.

Data show that the performance of candidates to the question was good as 98.8 per cent scored 2.0 marks and above. The analysis of their performance is shown in Table 3.

Table 3: Candidates' Score Range and Percentage on Question 7

Score Range	Description	Percentage of Candidates
0.0 - 1.5	Weak	1.2
2.0 - 2.5	Average	3.6
3.0 - 4.0	Good	95.2

Table 3 indicates that, 95.2 per cent of candidates scored from 3.0 to 4.0 marks, 3.6 per cent scored 2.0 to 2.5 marks and the rest 1.2 per cent scored from 0.0 to 1.5 marks. Most of the candidates (95.2%) were adequately knowledgeable about data collection tools used in educational research context. They provided the correct explanations of interview and focus group discussion as per demands of the question. They had clear understanding that an interview is a verbal interaction between interviewee

and interviewer while focused group discussion is based on a specific group of respondents who have enough information about a particular problem and represent the targeted population. Extract 7.1 is a sample of the correct responses from one of the candidates.

Extract 7.1: A sample candidates' correct responses to question 7.

Despite the candidates' good performance on this question, a few candidates (3.6%) scored average marks that ranged from 2.0 - 2.5. These candidates partially explained how interview and focus group discussion are used as data collection tools. Their explanations did not deserve full marks because some had mixed the concept of portfolio, checklist and normal class discussion group. The candidates who scored below two marks gave incorrect explanations to the asked tools of data collection. For instance, one candidate gave incorrect description about focus group discussion as; *experimental research and is the tools of research which hide the weakness of assessor and make organization that improve the learners to cooperate*. The other candidate wrote; *focus group a discussion is the instrument used by teacher for students and helps them to discuss themselves without feeling shame*. Specifically, most of the candidates had problems in their explanations about focus group discussion. Extract 7.2 is a sample of candidates' incorrect responses:

Extract 7.2: A sample of candidates' incorrect responses to question 7.

In Extract 7.2, the candidate outlined concepts related to teaching techniques instead of data collection tools specifically for interview and focused group discussion.

2.1.8 Question 8: Educational Research

The question required the candidates to describe the features to consider in construction of a research topic. The overall candidates' performance was weak as 61.3 per cent scored from 0.0 to1.5 marks and the rest 38.7 per cent scored above 1.5 marks. Figure 6 illustrates the candidates' performance.
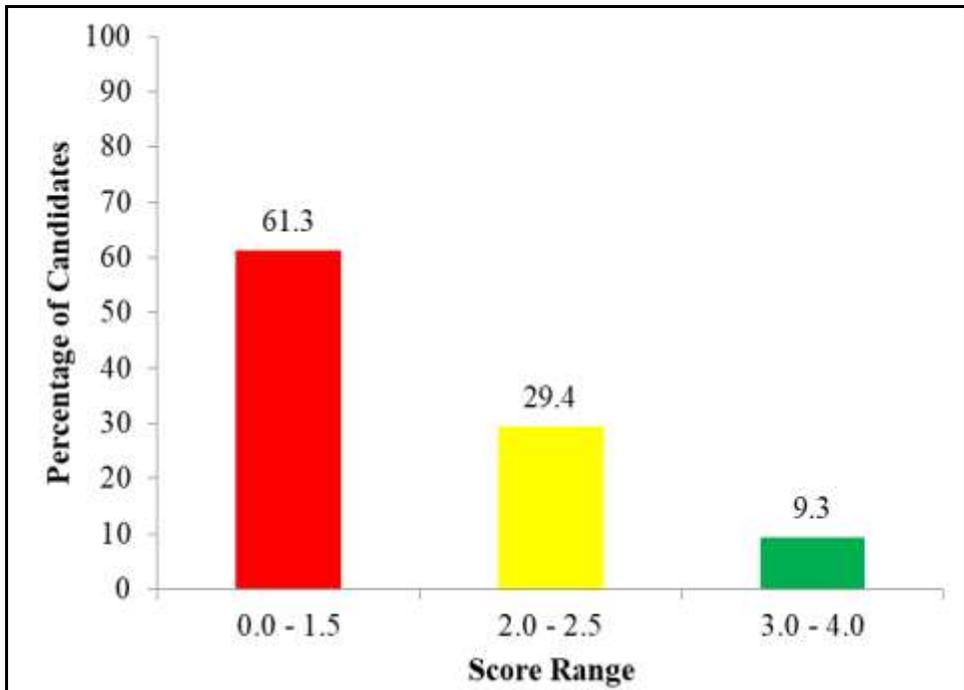
Figure 6: *Candidates' Performance on Question 8.*

Figure 6 shows that: 61.3 per cent of candidates scored from 0.0 - 1.5, 29.4 per cent scored from 2.0 to 2.5 marks, and only 9.3 per cent score 3.0 to 4.0 marks.

The candidates with low performance lacked competence on the tested concepts. This was revealed in their responses whereby they stated incorrect features considered in the formulation of good research topic. For example, one candidate wrote: *it gives good image of a research, it simplify work on when doing literature review and hypothesis.* Another candidate wrote*: generate knowledge, formulate education policies and determine the behavior of the learners*. Yet another candidate wrote: *it provides time or date for the research*. Besides, some candidates explained the functions of educational research which was contrary to the requirement of the question. Extract 8.1 is a sample of the incorrect responses to this question.

8	
	?) Literature revtew
	search different revrew example
	journals news and magazines and others.
	??) Methodology
	Collect different data about the topic
	given.
	???) Analysts
	analyse the collect data
	IV) Draw the conclussron or recommendatto
	ns. you draw the conclussron exam.
	ple by provrdrng suggestron and others

Extract 8.1 A sample of the incorrect responses to question 8.

In Extract 8.1, the candidate wrote the features of a research report instead of features of research topic.

Nonetheless, the candidates who scored high marks (3.0 to 4.0) had good knowledge about a research topic were able to provide proper features of a research topic. However, the candidates who had average performance provided two out of four correct features and therefore, scored marks ranging from 2.0 to 2.5. For example, one of the candidates wrote incorrect features as: "*must know the basic need such as food and researcher must know the transport problems*" and the other mentioned *"should have note book and should have a pen*." Extract 8.2 is a sample of good responses from one of the candidates.

i/ Research problem should be researable /testable

ii/ Research problem should be well stated clear.

iii/ Research problem should be ethical, which means that can not damage of affects people phychologizal or do not touch feelings and emotions of people or Population.

iv/ Research problem should, be aimed to find the solutions, Causes, measures and any positive change of a particular population /solving a problems!

Extract 8.2: A sample of good responses for question 8.

2.1.9   Question 9: Analysis and Interpretation of Test Results

The question tested the candidates' competence in using different approaches to process test scores. Particularly, it required the candidates to describe the application of ranking and histogram in processing test scores. The performance of the candidates to this question was generally weak since 72 per cent scored below 2.0 marks. Figure 7 illustrate further:
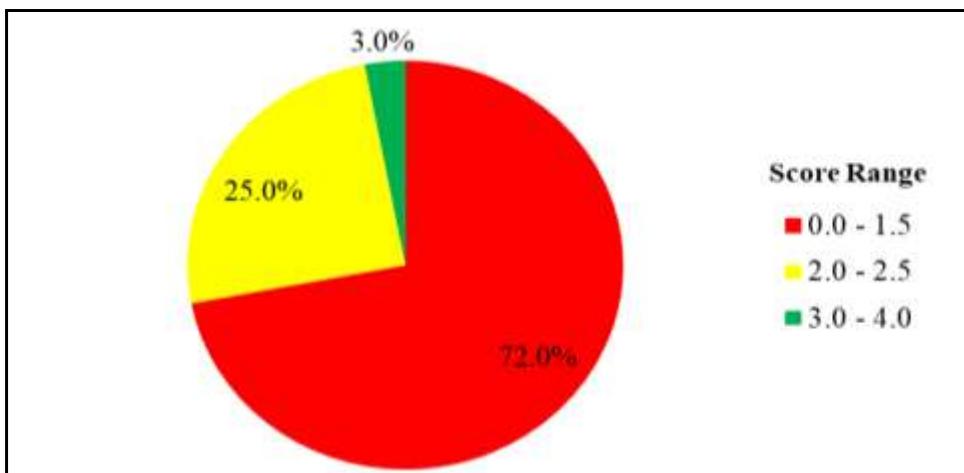

Figure 7: *Candidates' Performance on Question 9*

24

Figure 7 shows that the majority of the candidates (72.0%) scored from 0.0 to 1.5 marks, 25.0 per cent scored from 2.0 to 2.5 marks and only 3.0 per cent scored 3.0 to 4.0 marks.

Some of the candidates with low marks failed to describe the ranking as an approach associated with raw scores arranged in ascending orders. For example, one of the candidates wrote: "*ranking as a process that ensures set of characteristics associated with measurement of observable behavior of learners.*" Some of these candidates confused histogram with pie-chat data presentation. For instance, one candidate wrote: *"histogram is a circle which represents students' scores in terms percentage.*" This signified that the candidate had inadequate knowledge of processing test scores by ranking and histogram approaches. Extract 9.1 illustrates incorrect responses from one of the candidates:

Extract 9.1: A sample of incorrect responses to question 9.

In Extract 9.1, the candidate explained raking and histogram using the concept of correlation coefficient. She/he wrote equation of Spearman's rank correlation coefficient for histogram instead of the application of ranking and histogram.

The analysis of the candidates' responses reveals that those who scored average marks had adequate knowledge about the concept of ranking. These candidates demonstrated their skills by describing the concept of

ranking as used to process scores. They had clear understanding that the largest score is ranked number one, the second is ranked number two and so on until all scores are ranked. However, they had little knowledge on how histogram is used to present stores.

Furthermore, 3.0 per cent of the candidates who scored high marks gave appropriate application of ranking and histogram. They described that ranking indicates the relative position of each score in group or class and histogram presents scores in form of rectangular form. Extract 9.2 is a sample of correct responses to this question.

Extract 9.2: A sample of correct responses to question 9.

2.1.10 Question 10: Qualities of Tests

The question measured the candidates' knowledge on the concepts of validity and reliability of the tests. The question required the candidates to support the statement that "*A valid test is always reliable but a reliable test is not always valid.*" The general candidates' performance to this question was weak since 84.1 per cent scored below 2.0 marks. The data are summarised in Table 4.

Table 4: Candidates' Score Range and Percentage on Question 10

Score Range	Description	Percentage of Candidates
0.0 - 1.5	Weak	84.1
2.0 - 2.5	Average	12.4
3.0 - 4.0	Good	3.5

The data in Table 4 indicates that 84.1 per cent of the candidates scored 0.0 to 1.5 marks, 12. 4 per cent scored from 2.0 to 2.5 marks and 3.5 per cent scored 3.0 to 4.0 marks.

On the one hand, 84.1 per cent of the candidates who had weak performance faced difficulties in explaining the concepts of validity and reliability of test. They failed to explain how a test can be considered valid and reliable. Similarly, they applied the concept of coefficient of reliability instead of giving description why reliable test is not always valid when responding to the question. In addition, most of them lacked knowledge of the application of validity when used to qualify good test. For example, one of the candidates responded that: "*validity of a test is always usability and practicability, it always shows improvement of learners' performance.*" The other candidate explained test validity as: "*it can have different answers, and practicable and very realistic.*" They failed to understand that, test is reliable when it produces the same results for the same group of learners when tested on different occasion. The test is valid when it measures the contents that were supposed to be measured. Extract 10.1 is a sample of weak responses from one of the candidates:

Extract 10.1: A sample of the candidates' weak responses to question 10

Extract 10.1 shows that the candidate wrote some of the factors affecting reliability and validity instead of descriptions about reliability and validity of a test.

On the other hand, few (3.5%) candidates who scored high marks were fully aware that reliability of a test occurs when the same test results are maintained to the same students when administered twice or more. Also, they correctly agued on the given statement that, a test can consistently test the same contents twice but what was tested may not be what was intended by the test. Therefore, the candidates gave all or most of the correct points to the tested concepts. Extract 10.2 shows responses of a candidate who scored high marks.

10	In educational Measurement and Evaluation A Valid test is always reliable but a reliable test is not always valid
i)	test validity is ability to test what supposed to test.
	but
	Reliable test : is ability g test provide same result when administration provide same procedure are followed.
(ii)	Reliable test have method for estimate such test - retest method, independent judge method, split half method, equivalent form method while validity test have not estimate method.

Extract 10.2: A sample of the candidates' correct responses to question 10.

28

2.2 SECTION B: ESSAY/STRUCTURED QUESTIONS

Section B comprised of four structure questions which carried 15 marks each, making a total of sixty (60) marks. The questions were set from three (3) topics which are *Educational Assessment and Evaluation*, *Analysis and Interpretation of* and *Educational Research*. The candidates' response analysis for each question is as follows:

2.2.1 Question 11: Analysis and Interpretation of Test Results

The candidates were provided with the scores obtained from mathematics test administered twice to the same group of students within a short interval of time. They were required to study the scores shown in the table and answer the questions that followed. Table 5 shows the statistics provided:

Table 5: Mathematics Test Scores Provided

Names of Students	First Administration	Second Administration
A	15	14
B	10	12
C	11	11
D	13	13
E	14	15
F	12	13
G	12	16

Questions asked:

(a) Determine the correlation co-efficient of reliability of the two administered tests using the Pearson's product correlation coefficient.

$$r = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{\left(N \sum X^2 - (\sum X)^2\right)\left(N \sum Y^2 - (\sum Y)^2\right)}}.$$

(b) Comment on the value obtained in 11 (a).

(c) How could you interpret the correction of two variables when the value of correlation coefficient is high, moderate and zero?

Generally, the candidates' performance was weak as 72.3 per cent scored less than 6.0 marks. Their performance is summarised in Figure 8.
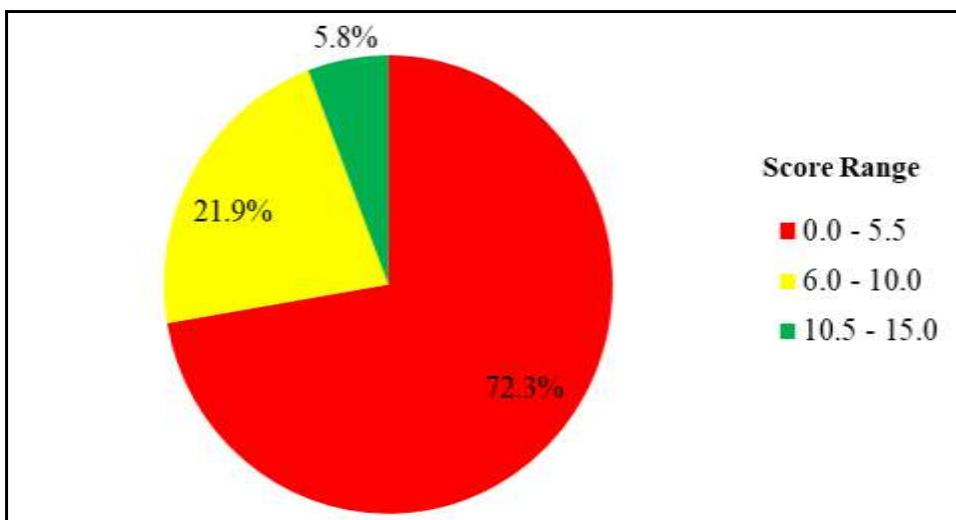
Figure 8: *Candidates' Performance on Question 11*

Figure 8 indicates that 72.3 per cent scored 0.0 to 5.5 marks, 21.9 per cent of the candidates scored 6.0 to 10.0 marks and 5.8 per cent scored 10.5 to 15.0 marks. In this case, majority of the candidates (72.3%) scored low marks since they were incompetent in mathematical analysis to determine the correlation coefficient of reliability of the two tests. They provided incorrect responses in most parts of this question. The candidates drew inappropriate columns of the table which led to incorrect values of columns XY, X^2 and Y^2 hence incorrect values of $\sum XY$, $\sum X$ and $\sum Y$. As result, they deduced wrong value of correlation coefficient. Some of them failed to determine the correlation co-efficient because they wrongly applied Pearson's product correlation coefficient formula in part (a). For example, one candidate used the formula; $r = \dfrac{2r}{1+r}$ and another candidate applied Spearman's rank correlation coefficient $r = \dfrac{1 - 6\sum D^2}{N(N^2 - 1)}$ instead of Pearson's product correlation coefficient $r = \dfrac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{\left(N \sum X^2 - (\sum X)^2\right)\left(N \sum Y^2 - (\sum Y)^2\right)}}$.

These candidates also failed to comment on the value of correlation co-efficient in part (b), because of the wrong value of *r* obtained in part (a) of the question. Regarding the value of *r* obtained, some of candidates claimed that there was negative correlation while others argued on zero correlation of the two administered tests. They commented that when the value of

30

correlation coefficient is negative means opposite correlation and when is zero means correlation is achieved for two tests. Their comments reveal the lack of knowledge to interpret the meaning correlation coefficient value. In part (c) of the question, the candidates gave incorrect interpretation of correlation of two variables when the value of correlation coefficient is high, moderate and zero. For example, one candidate stated: "*correlation coefficient is high where the student performance is homogeneity* and *correlation coefficient is zero means the test is accurate valid*." The other candidate responded that: "*the correlation coefficient is interpreted through the discrimination index, difficult index, Z-scores and T-scores*." They failed to recognise that the correlation coefficient is based on the relationship between two variables of scores (variable X and Y). Correlation is high when scores in variable X increase scores in Y while a negative correlation occurs when low scores in X increase scores in Y. Extract 11.1 is a sample of the incorrect responses from one of the candidates.

name of s.	1st test	2nd test	X	Y	x Y
A	15	18	1	3	3
B	10	12	7	6	42
C	11	11	6	7	42
D	13	13	3	4.5	13.5
E	14	15	2	2	4
f	12	13	4.5	4.5	20.25
G	12	16	4.5	1	4.5
			28	28	129.5

now.

$N = 7.$

$X = 28$

$Y = 28$

$xy = 129.5.$

from: $\gamma = \dfrac{N\Sigma xy - (\Sigma x)(\Sigma Y)}{\sqrt{(N\Sigma x^2 - (\Sigma Y^2 - (\Sigma Y)^4}}$

$\gamma = \dfrac{7 \times 129.5 - (28)(28).}{\sqrt{7(28^2) - (28^2 - (28^2).}}$

$\gamma = \dfrac{122.5}{\sqrt{28005264}}$

$\gamma = 0.9.$

@. The correlation co-efficient of reliability $= 0.9.$

Extract 11.1: A sample of incorrect responses to question 11.

In Extract 11.1, the candidate gave incorrect value of correlation coefficient due to inappropriate manipulation of numerator and denominator and skipped other parts of the question.

The candidates (21.9%) who scored average marks (6.0 to 10.0) were knowledgeable about the determination of correlation coefficient of two tests through Pearson's correlation coefficient. They correctly manipulated values of XY, X^2, and Y^2 which led them to obtain the value of correlation coefficient in part (a). However, they gave incorrect interpretation of coefficient in part (b) of the question.

The candidates (5.8%) who scored high marks (from 10.5 to 15.0) were aware of the Pearson's product correlation coefficient formula in part (a). They correctly solved value of columns of the table which led them to obtain value $\sum X^2$ and $\sum y^2$. These candidates explained correctly the meaning of the value of correlation coefficient in part (b). They commented that value of coefficient signify positive moderate correlation between two tests. In part (c), they correctly explained the significance correlation of two variables when the coefficient of correlation is high, moderate or zero. Extract 11.2 is a sample of the correct responses from one of the candidates:

(a)

Name	First(x)	Second (y)	x^2	y^2	xy
A	15	14	225	196	210
B	10	12	100	144	120
C	11	11	121	121	121
D	13	13	169	169	169
E	14	15	196	225	210
F	12	13	144	169	156
G	12	16	144	256	192
	$\Sigma x = 87$	$\Sigma y = 94$	$\Sigma x^2 = 1099$	$\Sigma y^2 = 1280$	$\Sigma = 1178$

$N = 7$,

from

$$r = \frac{N\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{(N\Sigma x^2 - (\Sigma x)^2)(N\Sigma y^2 - (\Sigma y)^2)}}$$

$$r = \frac{7(1178) - (87)(94)}{\sqrt{(7(1099) - (87)^2)(7(1280) - (94)^2)}}$$

$$r = \frac{8246 - 8178}{\sqrt{(7693 - 7569)(8960 - 8836)}}$$

$$r = \frac{68}{\sqrt{(124)(124)}} = \frac{68}{124}$$

$$r = 0.548$$

∴ The value of r ∞ is 0.548

(b) To comment on the value obtained

There are Moderate correlation co-efficient between first administration and second administration.

(c)(i) If there is high correlation coefficient this means there is strong correlation co-efficient between first and second administration. (homogenous students)

(ii) If there is Moderate correlation co-efficient this means the scores of the learners are equal there is no *t high different or adjacent score

(iii) If there is zero this means there is no relationship between test A and Test B.

Extract 11.2: A sample of the correct responses to question 11.

2.2.2 Question 12: Educational Assessment and Evaluation

The question tested the candidates' knowledge of the instruments of assessing students. Specifically, the question required the candidates to explain five instruments recommended for assessing students. The performance of the candidates was generally good because 79.1 per cent scored above 10.0 marks. The overall performance of the candidates is illustrated in Figure 9.
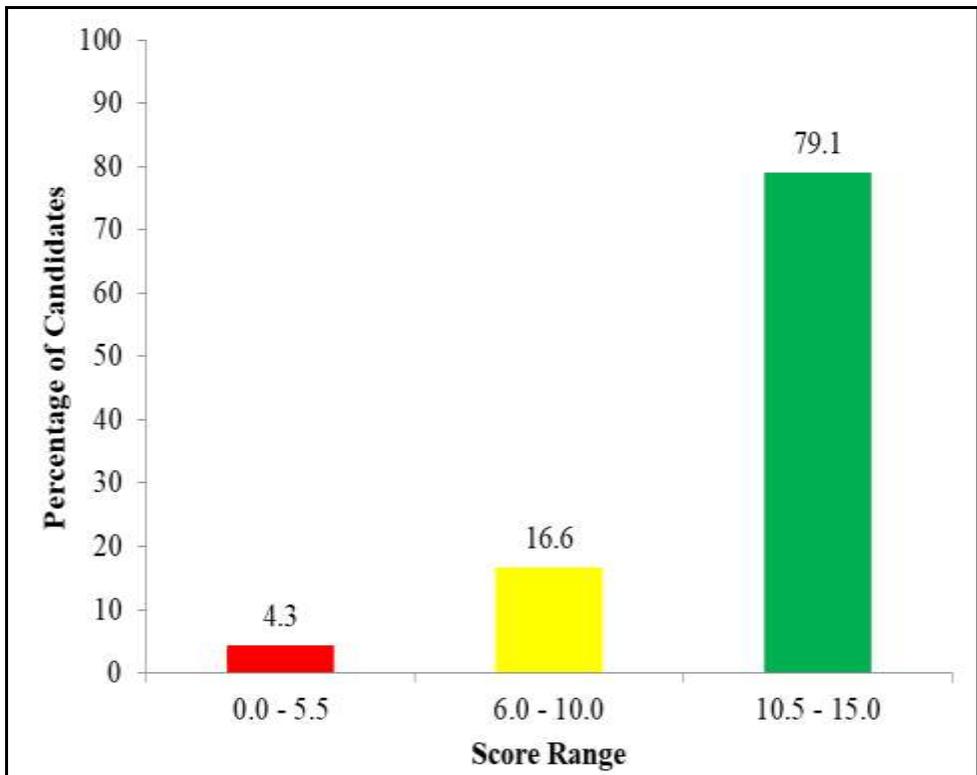
Figure 9: *Candidates' Performance on Question 12*

Figure 9 shows that the majority of the candidates (79.1%) scored from 10.5 to 15.0 marks, 16.6 per cent scored from 6.0 to 10.0 and only 4.3 per cent scored from 0.0 to 5.5 marks.

The candidates who scored high marks (6.0 and above) correctly explained five instruments of assessing students. They described the instruments which are mostly used in assessing students' achievement such as examinations, portfolio, questionnaires, interviews and observation checklists. The candidates' responses signify that they were knowledgeable about the concepts of instruments of assessing students. Extract 12.1 illustrates good responses from one of the candidates:

Assessment is the process of analysing and translating data into interpretable form so that the decision can be made.

The following are the instruments would recommend to be used in carrying out such responsibility.

Observation instrument, this is the instrument of assessing student which can employ the sence of organs such as hearing, seeing, smell and touching, This used to assess student progress by making observation of the students behavior in the school. This is the instrument which would to be recommend to be used in carrying out such responsibility.

Interview instrume, this instrument of assessment imploy the conversation between teacher and student which teacher is interviewer and students is interviewee, the interviewer ask the interviewee questions. then the interviewee respond the answer to that questions, which the teacher can measure, the skills, knowledge and attribute perpomed by the learners, this is the instrument would recommedd to be used to assess students.

Questionneire instrument, this is the instrument of assessment which the respondent written a piece of questions on a paper and then give to the respondents to be answer and the respondent answer the question through written way to be back to the student who ask a question to be

answered, the questionnaire instrument, it
is a good method because it can reduce
biasness and save time during assessing the
learners, this is an instrument would recommen
nd to use to assessed student accordingly.
portfolio instrument, this is the
scientific and systematic collection of all
student work in order to determine the
performance to be achieved goals, the portfoli
have two work part which is reflective
work and work sample, this can be used
to assess student achievement, the portfolio
help to make improvement of learners, help to
make diagnosis of learning problems and
also help to determine the difficulties of the
learners, this is the instrument would
recommend to use to assed student accordingly.
Test and examination instrument, thi
is the instrument of assessing student which
involve the using set of questions and task
in order to determine the attributes possessed
by the learners during the process of teaching
and learning, this is the instrument would
recommend to use to assess students accordingly.
Generally these are the instruments of
would recommend to be t used in carrying out
students accordingly, but the following are the
advantages of assessment which are motivation,
diagnosing learners problem, resharp the
method of teaching and learning.

Extract 12.1: A sample of correct responses to question 12.

Despite the good performance of the candidates on this question, few (4.3%) candidates had weak performance. It was noted that, most of these candidates explained correctly only one instrument out of five as per question requirement. Some of the candidates gave explanations basing of

school administration and management instead of assessment instruments. For example, one candidate explained: *create good organization with staff members and school policy; promote corporation among students; availability of teachers, teaching and learning materials*. In addition, some of them provided incorrect responses which reflect the concepts of curriculum teaching materials such as books and preparation for teaching such as teaching aids and models. This indicated the lack of knowledge in the concept of instruments for assessing students. Extract 12.2 shows one of the incorrect responses from one of the candidates.

12 Cont.	by the topic and sub topic. Cook A during to class room
	hours
	long book ; This is the book which show the
	works done by the teacher during to period of to
	years or term when a teacher wanted to conducted
	a test to study should have long books when,
	which show the work done by to teacher
	during to Class room hour. It is very importance
	for to teacher to be followed to long book.
	Those are the academic instrument should
	have when 'conductors test of the Student
	have to explaining about are to examination
	to the study when he appointed of man,

Extract 12.2: A sample of incorrect responses to question 12.

In Extract 12.2, the candidate wrote curriculum materials such as text books, scheme of work, lesson notes and syllabus instead of students' assessment instruments. This may suggest misinterpretation of the question or little knowledge possessed by the candidate with regard to students' assessment instruments.

2.2.3 Question 13: Educational Research

The question tested the candidates' competence in writing a research report. The candidates were required to describe major steps which are needed in writing a research report. The candidates' performance was average as 65.5 per cent of candidates scored from 6.0 to 15.0 marks. Their performance is illustrated in Figure 10:
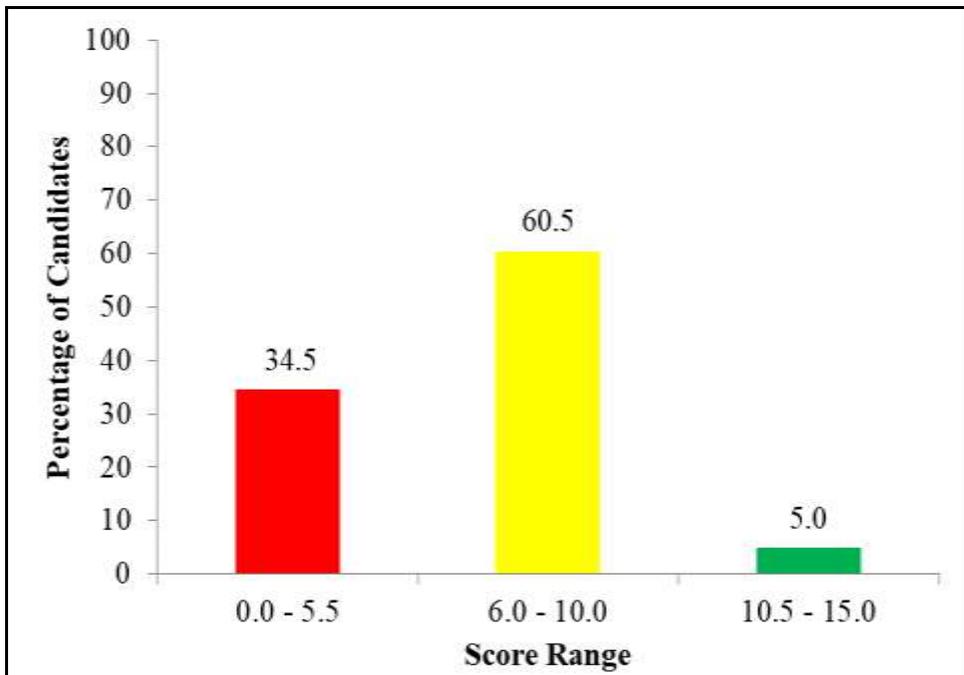
Figure 10: *Candidates' Performance on Question 13*

Figure 10 shows that 34.5 per cent of the candidates scored from 0.0 to 5.5 marks, 60.5 per cent scored from 6.0 to 10.0 marks and only 5.0 per cent scored from 10.5 to 15.0 marks.

The candidates (65.5%) with high marks provided appropriate description of six major steps of writing a research report in a logical order. Some of them supported their explanation by giving different examples. Similarly, they knew that a research report is a product of thoughtful and accurate inductive work that is written after conducting the research. This was an indication that these candidates had enough knowledge about the concept of research and the appropriate steps of writing its report. Extract 13.1 represents one of the correct responses for this question:

Research report refers to the document which is written to explain what was done during the research. Writing the report in research process is the last stage where what was done is explained systematically. Research report help the reader to understand how data revolve around the problem and also acquaint the reader on how the research was all about. The following are the major steps that will be taken in carrying out the activity of writing the research report!

The Tittle: The first step in writing a research report is writing the title of the research report. That means it shows what the report is about. For example, title may be "Reason for dropout of girls in science subject". The reader will be sure on what the report is all about. The title should be clear and specific.

The Introduction: In this part while writing a research report you give an introduction of your research. It include background of the study, hypothesis, preface, acknowledgements and other parts which make up the introduction of your research. Therefore this will be another very important step while writing a research report.

Methodology:- After the introduction is well stated and clearly written then, the following step is writing all the methodologies which was used to collect data for the research to take place. This methods will be accurately shown and the data must be well shown according to the data that was obtained from the field or through reviewing different literature.

Data analysis and Interpretation:- Then after writing the methodologies it's now the time to analyze the data and interpret them clearly. While writing a research report the data obtained must be well analysed and interpreted so that it becomes easier for a reader to understand and

13 Cont.

know what was meant on those data. So this is another crucial step in writing a research report.

Conclusion: Then after writing how the data was analysed and interpreted, now while writing a report the conclusion must be shown and clearly stated. It shows what is the conclusion drawn after all the activities which has been in conducting the research. Does our conclusion corresponds to the hypothesis which was made?, then it is in conclusion when all this will be answered and the reader will be aware on how the research was and the conclusion made.

Lastly, to write the references, Appendix glossary and other related information which mark the end of writing a research topic. At this step it help a reader to know the meaning of words which were used, the references and other which make the report more complete.

Consequently ; those are some of the steps to be followed when writing a research report. A research report should be systematically and well arranged so that the reader to understand well about the research problem and the conclusion that was drawn.

Extract 13.1: A sample of correct responses to question 13.

In contrast, the candidates who scored low marks (1.0 to 5.5) described only two steps of writing research report. Some of these candidates either ended up defining the research and research report in the introductory part or skipped some of the important stages of writing research report.

Some of the candidates who scored zero described the research instruments and research proposal. Other candidates described the components of research process. For example, one candidate wrote: *research report involves steps which are; identification of a problem, background to the study through literature review, methodology and summary of the research*.

43

Another candidate regarded steps of writing research report as research instruments. The candidate wrote: *The research report includes preparation of questionnaires, interviews and observation checklist*. Extract 13.2 is a sample of the incorrect responses to the question.

Reasearch report refer to a document that contain a summary of what is being researched by a researcher. It is writen after conducting the research as a last staps in any research, It aiming at providing the important hints that helpe to make evalution after a research. The following are the steps to be followed when starting writing research report.

Prepare tools essential for writing report, this is the first step during preparation of writing report to where by a researcher should observe the tools like pen, pepars, rular penlial as well as mathematical set this will help the researcher to write a good report and showing the illustration of different phenomenon if there is.

Prepare the cost that the researcher may use during conducting his/her researcher, this is another steps that the researcher must be aware of the cost that he or she should gooner to use during the activity of report writting.

Knowing exactly the area or field of study for writing your report, this will help to reduce the nsr that the research meet with during report writin It is the stage that can help the researcher to identify weakness and strength of his/her research that is being conducted.

Extract 13.2: A sample of incorrect responses to question 13.

In Extract 13.2; the candidate explained about cost, tools like pen, papers, rulers, pencil and mathematical set which were irrelevant to the asked question. These responses suggest that this category of candidates had little knowledge and competence in writing research report.

2.2.4   Question 14: Educational Research

The question measured the candidates' knowledge about action research and their competence in carrying it out. The candidates were required to use their skills in research design to advice the school academic committee on the five major steps of carrying action research. The overall performance of the candidates was average as 69.6 per cent passed the question by scoring 6.0 to 15.0 marks. The performance is summarised in Figure 11.
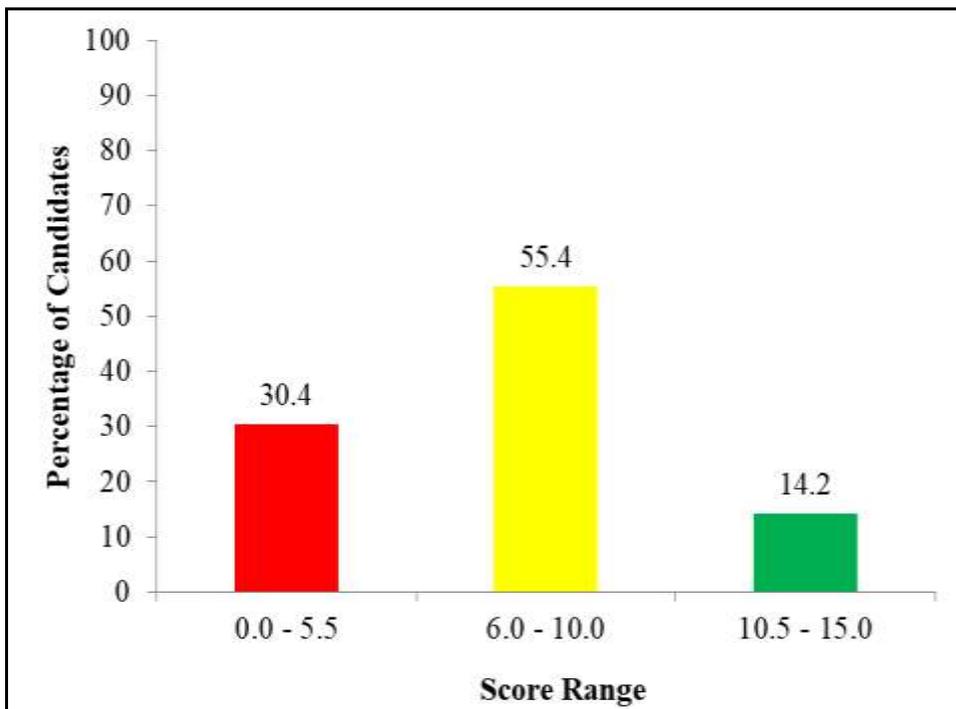

Figure 11: *Candidates' Performance on Question 14*

The analysis of data in Figure 11 indicates that 55.4 per cent of candidates scored 6.0 to 10.0 marks, 14.2 per cent scored 10.5 to 15.0 marks and 30.4 per cent scored 0.0 to 5.5 marks. The candidates (14.2%) who scored high marks were knowledgeable enough about the steps of conducting action research. They knew that appropriate steps of carrying action research enable researchers to find solutions to immediate problems facing the

45

society. The explanations were supported with relevant examples. The candidates described about hypothesis which enables the researchers to find out if the expectation of the investigation is true or not. Extract 14.1 is a sample of the correct responses from one of the candidates:

Action research: Is the type of research which deal with solving problem with close collaboration between academic and managerial staffs. The action research divided into two major types which are cyclic and spiral action research.

The following are the major steps to be followed in carrying out the task that is action research.

Identifying the problem: A researcher should identify the purpose of that research that why your conducting such kind of research. Forexample the problem can be high dropout rates of girls in sciences subject.

formulate hypothesis: Is the second stage in action research where by a research can develop his or her own answer or make assumption in the problem. forexample he or she assume that girls dropout because of the poor belief in the community that girls are weak object.

Establish study design: This is the Third stage where by a researcher design who how a research can be conducted or the plan of obtain answer to the research question which high dropout of girls in science subject. Forexample accademic committee can choose appropriate sample or population in order to obtain the data.

Data collection: the accademic committee should collect data about the research problem. for example they identify instruments used in collecting data like Questionnaire, interview and observation. Also this instruments can be choosen according to the research problem, that the committee choose the best method which do not cost them more,

Extract 14.1: A sample candidates' correct responses to question 14.

Despite the overall good performance of the candidates on this question, 55.4 per cent of the candidates scored average marks. The candidates partially described steps of conducting action research which did not deserve full marks. Other candidates' descriptions were not supported with relevant examples. For example, one of the candidates wrote: "*emphasize on science subjects, women are given more priority than men on science subjects,*" while the other candidate stated: *"construct well physical infrastructure and facilities that support learning and facilities that support learning.*"

The candidates who scored below six marks either gave incorrect descriptions to all or most of the steps of carrying action research. Specifically, most of the candidates scored zero basing on their description. For instance, one of the candidates explained the steps as: *(i) to decrease the fear of failure and encourage them through giving many tests and assignment (ii) to give conducive environment for learning science subjects and provide reward to learners*. The candidates described concepts related to reinforcement and not steps of conducting action research. Extract 14.2 is a sample of candidates' incorrect responses:

Research design, refers to the ways that used by a researcher to get data or information to the people who going to research them.

The following are the steps of research design

✗ To know the invironment which your going to do a field, so as to know if you are fit (capable) with that invironment by considering your health status. For example a person who have asthma not safe for him/her to live in areas with Low temperature.

You should know the culture of area of field, so as to cope with it during collection of data and that help to know which method of data collection is better for that society, so as to have good information. For example a society which have a culture of discriminate girls it is not easy to get information about why girls dropout science subject unless you provide enough education for them.

To know nature of the people of the society of do a field, because some societies do not want researcher to investigate in their society so they can give a researcher information about a problem want or they can give a wrong data.

To know the budget you have by consider transportation, accommodation, food and methodology in research design used, so as to afford a research you are going through.

Extract 14.2: A sample of candidates' incorrect responses to question 14.

In Extract 14.2, the candidate incorrectly regarded action research as research design and thus gave unclear descriptions about major steps to follow in conducting action research.

3.0 ANALYSIS OF CANDIDATES' PERFORMANCE IN EACH TOPIC

A total of seven (7) topics were examined in Educational Research, Measurement and Evaluation examination papers. These topics were: *Test Construction, Educational Research*, *Assessing Achievement*, *Educational Assessment and Evaluation, Educational Measurement*, *Analysis and Interpretation of Test Results* and *Qualities of Tests*.

The overall analysis of the candidates' responses in each topic in Educational Research, Measurement and Evaluation shows that good performance was reflected on the topics of *Educational Assessment and Evaluation* (95.7%) and *Educational Measurement* (71.9%). On the other hand, average performance was observed in the topic of *Educational Research* (68.62%). Moreover, the topics of *Analysis and Interpretation of Test Results* (34.8%), *Assessing Achievement* (29.6%), *Qualities of Tests* (15.9%) and *Test Construction* (13.2%) had weak performance (See Appendix I).

A comparative analysis suggests that, the performance in the topic of *Educational Measurement* has been maintained as good for two consecutive years from 2021 to 2022. On the other hand, the performance in the topics of *Educational Assessment and Evaluation* and *Educational Research* has significantly improved from weak performance in 2021 to good and average performance in 2022 respectively. Contrary however, the performance of the topics of *Qualities of Tests* and *Test Construction* has decreased from good to weak performance. Nonetheless, the performance in the topic of *Assessing Achievement* remained weak in 2021 and 2022 (See Appendix II).

4.0 CONCLUSION

In general, the candidates' performance in Educational Research, Measurement and Evaluation was good as 99.11 per cent of the candidates passed the examination and only 0.89 per cent failed. The good performance was due to the candidates' ability to understand the demands of the questions having knowledge and competence on the subject matter, as well as good mathematical skills and correct application of the principles

in interpreting the subject concepts. They were able to explain and elaborate the points using appropriate subject's principles and theories.

On the other hand, the reasons for weak performance included; inadequate knowledge of the subject matter, little understanding of the concepts in the questions, inability to apply formula and misinterpretation of some subject's principles and theories.

5.0   RECOMMENDATIONS

In order to improve the performance of the candidates (prospective teachers) in Educational Research, Measurement and Evaluation subject student-teachers are strongly advised to:

(a)   do more discussions and presentations about the concept of item analysis, measure of dispersion, processing of test scores and qualities of tests. This will help them to correctly differentiate the concepts of discrimination index and difficult index; application of ranking and histogram in processing test scores; scoring procedures of tests; behavioral assessment techniques and validity and reliability of tests.

(b)   solve various questions about the concept of correlation coefficient of reliability, standard deviation and measures of average. The practice will help student-teachers to develop competencies and skills of derivation and application of appropriate formulas.

SUMMARY OF THE CANDIDATES' PERFORMANCE IN 762 EDUCATIONAL RESEARCH, MEASUREMENT AND EVALUATION SUBJECT DSEE 2022

S/N	Topic	Question Number	Performance in each question (%)	Average performance per topic (%)	Remarks
1.	Educational Assessment and Evaluation	12	95.7	95.7	Good
2.	Educational Measurement	5	71.9	71.9	Good
3.	Educational Research	7	98.8	68.62	Average
		2	70.2		
		4	68.6		
		14	69.9		
		13	65.5		
		8	38.7		
4.	Analysis and Interpretation of Test Results	6	48.6	34.8	Weak
		9	28.8		
		11	27.7		
5.	Assessing Achievement	3	29.6	29.6	Weak
6.	Qualities of Tests	10	15.9	15.9	Weak
7	Test Construction	1	13.2	13.2	Weak

COMPARISON OF THE CANDIDATES' PERFORMANCE IN 762 EDUCATIONAL RESEARCH, MEASUREMENT AND EVALUATION SUBJECT DSEE BETWEEN 2021 AND 2022 TOPIC-WISE

S/N	Topic	2021			2022		
		Number of Question(s)	Average Performance of the Candidates per topic (%)	Remarks	Number of Question(s)	Average Performance of the Candidates per topic (%)	Remarks
1.	Educational Assessment and Evaluation	1	39.2	Weak	1	95.7	Good
2.	Educational Measurement	1	98.4	Good	1	71.9	Good
3.	Educational Research	4	38.2	Weak	6	68.62	Average
4.	Analysis and Interpretation of Test Results	2	61.3	Average	3	34.8	Weak
5.	Assessing Achievement	1	10.60	Weak	1	29.6	Weak
6.	Qualities of Tests	1	90.10	Good	1	15.9	Weak
7.	Test Construction	6	70.65	Good	1	13.2	Weak