

37

**THE IMPACT OF THE MULTIPLE-CHOICE AND THE
SHORT-ANSWER FORMATS ON STUDENTS'
PERFORMANCE: THREE GOVERNMENT
SENIOR HIGH SCHOOLS IN
ADDIS ABABA IN FOCUS**

**A THESIS
PRESENTED TO
THE SCHOOL OF GRADUATE STUDIES
ADDIS ABABA UNIVERSITY**

**IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE
MASTER OF ARTS IN TEFL**

**BY
WOGARI NEGARI**

JUNE, 1995

ADDIS ABABA UNIVERSITY
SCHOOL OF GRADUATE STUDIES

THE IMPACT OF THE MULTIPLE-CHOICE AND THE
SHORT-ANSWER FORMATS ON STUDENTS'
PERFORMANCES: THREE GOVERNMENT
SENIOR HIGH SCHOOLS IN
ADDIS ABABA IN FOCUS.

BY

WOGARI NEGARI

Approved by _____

Dejenie Keta



Advisor

Signature

Examiner

DON PORTER

Signature



Examiner

Hailom Bantezerga

Signature



Examiner

Signature

Examiner

Signature

ACKNOWLEDGMENTS

I would like to express my heartfelt gratitude to Dr. Dejenie Leta, my advisor, who devoted his time and knowledge in helping me in writing this thesis. I am most grateful for his constructive criticism, patience and encouragement without which this study would have never reached its present stage.

I would also like to express my appreciations to Ato Markos Tadesse, Ato Tola Gamada, Ato Yalew Endawoke, and Ato Mohammed Hassen who advised me on the statistical part of the paper.

I owe special thanks to Dr. Waqtasu Negari and 'Woizero' Majitu Tarfa for their invaluable advice and moral support.

Especially, I wish to express my indebtedness to the Swedish Agency for Research Cooperation with developing countries (SAREC) for providing me with the necessary fund during the various stages of the research work.

I also wish to thank the directors, the unit leaders, and the English language teachers of Abyot Kirs, Ayer Tena, and Higher Four Government senior high schools in Addis Ababa for their participation and contribution in this study.

TABLE OF CONTENTS

	Page
ACKNOWLEDGMENTS	i
LIST OF TABLES	iv
LIST OF GRAPHS	v
LIST OF APPENDICES	vi
ABSTRACT	viii
CHAPTER ONE: INTRODUCTION	1
1.1. The Statement of the Problem	1
1.2. The Purpose of the Study	3
1.3. The Importance of the Study	3
1.4. Limitations of the Study	4
1.5. Delimitation of the study	4
1.6. Definition of Terms Used	4
CHAPTER TWO: REVIEW OF RELATED LITERATURE	6
2.1. Achievement Tests	6
2.2. Test Specifications	6
2.3. The Multiple-Choice and the Short-Answer Formats	7
2.4. Discrete Point Item and Integrated Tests	11
2.5. Communicative and Traditional Tests	13
CHAPTER THREE: METHODOLOGY	14
3.1. Subjects	14
3.1.1. Selection of Schools	14

3.1.2. Selection of Students	14
3.1.3. Selection of Teachers	15
3.2. Instruments	15
3.2.1. Tests	15
3.2.2. Questionnaires	17
3.3. Procedures	17
3.4 Statistical Information	20
CHAPTER FOUR: RESULTS AND DISCUSSION	22
4.1. Results of the Tests	22
4.2. Results of the Questionnaires	32
4.2.1. Results of the students' Attitudes	32
4.2.2. Results of the Teachers' Attitudes	33
4.2.3 Results of the Students' Responses to Actual Classroom Practices Questionnaire	35
4.2.4. Results of the Teachers' Responses to Actual Classroom Practices' Questionnaire	40
CHAPTER FIVE: SUMMARY, CONCLUSION AND RECOMMENDATIONS	45
5.1. Summary	45
5.2. Conclusion	47
5.3. Recommendations	49
BIBLIOGRAPHY	50

LIST OF TABLES

Table 1: Table of Specifications for the Two Tests' (Multiple-Choice and Short-Answer)	16
Table 2: Analysis of Variance of the Students According to their performances in the Two Tests	22
Table 3: A t-test Computed for the Mean Difference in the Students' Performance on both Tests having the Same Content Presented in the Multiple-Choice and the Short Answer Formats	23
Table 4: A t-test for a Difference in performances in the Components (Comprehension, Vocabulary, and Structure) of the Multiple-Choice Test.....	24
Table 5: A t-test for a Difference in Performances in the Components (Comprehension, Vocabulary, and Structure) of the Short-Answer test.....	25
Table 6: Difference between the Students' Attitude Scores on the multiple-Choice and the Short-Answer Formats	28
Table 7: Difference between the Teachers' Attitude Scores on the Multiple-Choice and the Short-Answer Formats	29
Table 8: The Students' Responses to Actual Classroom Practices' Questionnaire.....	30
Table 9: The Teachers' Responses to Actual Classroom Practices' Questionnaire.....	33

LIST OF GRAPHS

Graph(a): A Summary of the Students' Scores on both the Multiple-Choice and the Short-Answer	27
Graph(b): A Summary of the Students' Scores on each Component (Comprehension, Vocabulary, and Structure) of the Multiple-Choice and the Short-Answer Tests	31

LIST OF APPENDICES

	PAGE
Appendix I: Short-Answer Test for Grade Eleven -----	58
Appendix II: Puzzle Test-----	64
Appendix III: Multiple-Choice Test for Grade Eleven -----	65
Appendix IV (a): Answer Key for the Multiple-Choice Test-----	75
Appendix IV (b): Answer Key for the Short-Answer Test-----	75
Appendix V: Marking Guidelines -----	77
Appendix - VI: Students' Questionnaire -----	79
Appendix - VII: Teachers' Questionnaire -----	84
Appendix VIII: The Short-Answer Test Scores Marked by Three Raters -----	91
Appendix IX: Inter-Raters Reliability -----	97
Appendix X: The Students' Scores on the Multiple-Choice and the Short-Answer Tests -----	98
Appendix XI(a): Correlation between the Students' Scores on the Multiple-Choice and the Short-Answer Tests. -----	101
Appendix XI (b): Mean Differences Computed with the help of t-dependent for Repeated Measures -----	103
Appendix XII (a): The Students' percentage and Average Scores on the Multiple-Choice and the Short-Answer Tests. -----	106
Appendix XII (b): Row and Column Totals of the Students' Scores on the Multiple-Choice and the Short-Answer Tests. -----	109
Appendix XII (c): The Summary of Row and Column Totals of the Students' Scores on the Multiple-Choice and Short-Answer Tests-----	111
Appendix XII (d): Analysis of Variance (ANOVA) of the Students According to Their Performances in the Two Tests-----	112
Appendix XIII :The Reliability of the Multiple-Choice Test. -----	114
Appendix XIV (a): The Students' Scores on their Attitudes Towards the Multiple-Choice and the Short - Answer Tests. -----	115
Appendix XIV (b): The Reliability of the Students' Attitude, Questionnaire-----	121

Appendix XIV (c): The Students' Attitudes' Mean Difference	122
Appendix - XV (a): The Students' Responses to their Actual Classroom Practices' Questionnaire.	123
Appendix - XV (b): The Reliability of the Students' Actual Classroom Practices' Questionnaire.	130
Appendix - XVI (a): The Reliability of the Teachers' Attitude; Questionnaire	131
Appendix - XVI (b): Mean Difference of Teachers' Attitudes towards the Multiple-Choice and the Short-Answer Formats.	133
Appendix XVII: The Reliability of the Teachers' Questionnaire on Their Actual Classroom Practices.	135
Appendix XVIII (a): The Students' Performances in the Components (Comprehension, Vocabulary, and Structure) of the Multiple-Choice Test.	136
Appendix XVIII(b): Correlation between the Students' Scores on Comprehension, Vocabulary, and Structure parts of the Multiple-Choice Test.	141
Appendix XVIII (c): Differences of the Students Performances in the Components (Comprehension, Vocabulary, and Structure) of the Multiple-Choice Test.	142
Appendix - XIX (a): The Students' Performance, in the Components (Comprehension, Vocabulary, and Structure) of the Short-Answer Test.	143
Appendix - XIX (b): Correlation between the Students' Scores on Comprehension, Vocabulary and Structure.	148
Appendix XIX (c): Differences among the Students' Scores on Comprehension, Vocabulary and Structure.	150

ABSTRACT

In this study, an attempt was made to look into the impact of the multiple-choice and the short answer formats on students' performances.

The study population comprised three senior secondary schools (Abyot Kirs, Ayer Tenna, and Higher Four) in Addis Ababa. One hundred and twenty Grade Eleven students (65 male and 55 female) enrolled during the 1994/95 academic year participated in the study. The students were divided according to achievement levels (high, average, and low) based on their first semester English results. In addition, nine Grade Eleven English teachers were sampled for the study.

The instruments employed were two tests on the same content in three parts each (comprehension, vocabulary, and structure), and two questionnaires comprising three parts each (bio-data, attitudes, classroom practices).

The results were computed on the basis of a two-way analysis of variance (ANOVA), correlated samples of two-tailed t-test (t-dependent for repeated measures), the Pearson product moment correlation, r , and percentages. The level of significance was set at 0.05.

The results of the computation exhibited that there was statistically significant difference between the students' performances in the two tests. The students invariably scored higher on the multiple-choice test than on the short-answer test.

Regarding the attitude description questionnaires, there was no significant difference between the students' attitudes towards the two formats, whereas the teachers had favourable attitudes towards the short-answer format.

Inconsistencies were seen in the students' and teachers' responses to the items dealing with the frequency of employing the multiple-choice and the short-answer formats.

Based on the results of the findings, conclusions were drawn and recommendations made.

CHAPTER ONE: INTRODUCTION

This chapter consists of the statement of the problem, the importance of the study, limitations of the study, delimitation of the study, hypothesis, and definition of terms used.

1.1. The Statement of the problem

In Ethiopian junior and senior high schools these days, the multiple-choice format is predominantly being employed in tests and examinations though the English teaching syllabus recommends that both the multiple-choice and the short-answer formats are to be employed in teaching/learning. This indicates that not all formats used in teaching given due regard in testing. When the multiple-choice is over-emphasized as a sole format of measuring students' achievements, it doesn't seem that attention is given to the likely damages that the mere use of the multiple-choice format could inflict upon the whole teaching/learning process. Nevertheless, the importance that this format has should have been weighed against its limitations (Bogale, 1994; Dejene, 1990; Tadesse, 1990; Tesfaye, 1982).

Bogale(1994) and Tadesse(1990) add that the teachers complain that many of the Ethiopian junior and senior high schools' students hardly write and comprehend very simple English sentences. Moreover, the English Curriculum Development, as quoted by Tadesse(1990), suggests that investigations are to be made so as to diagnose problems that give rise to the decline of the Ethiopian students' command of English.

As to whether to use the multiple-choice format or the short-answer in measuring students' performances requires an appropriate investigation to be made. Frary (1985) and Oller(1979) note that how to make use of these formats in tests and examinations is a persistently recurring question.

In sum, what has been said so far implies that there is a mismatch between teaching and testing English in Ethiopian Schools. That tests and examinations measure what they are intended to do thus becomes a controversial issue.

So, the present study attempts to find out the impact of the multiple-choice and the short-answer formats on students' achievements by administering tests on the same content, but using the two formats: multiple-choice and short-answer.

In general, this study aims at answering the following questions:

- Do the two formats (multiple choice and short-answer) have an impact on students' achievements?
- Is there an interaction between the formats and achievement levels?
- Do students' scores on the two tests have a statistically meaningful difference?
- Do students' attitudes towards the multiple-choice and the short-answer formats have a statistically meaningful difference?
- Do teachers' attitudes towards the multiple-choice and the short-answer formats have a statistically meaningful difference?

1.2. The Purpose of the Study

In Ethiopia, except complaints and informal talks, no comprehensive study has so far been carried out on students' achievements in the multiple-choice and the short-answer formats. Thus, this paper is an attempt to look into the differences and/or similarities that students' scores exhibit in the same content tests presented in the two formats stated above.

1.3. The Importance of the Study

Besides revealing the impact that the multiple-choice and the short-answer formats could have on students' performances, the study is hoped to be of paramount relevance in that its findings could help:

- teachers and test writers know the relationships and/or differences between the two formats (multiple-choice and short-answer) and their impact on the scores of students at the three achievement levels: high, average, and low.
- teachers make efforts to reconcile teaching and testing so that they reflect and/or support each other.
- test constructors introduce eclecticism in measuring students' achievements.
- teachers or any interested persons be initiated in conducting further investigations into other formats to be used in testing without giving preference to certain formats over the others so long as the instructional objectives require them,
- the English curriculum designing authorities realize that teaching and testing are inseparable.

1.4. Limitations of the Study

Due to time and financial constraints, the study focused only on Grade Eleven students enrolled in three government senior high schools in Addis Ababa during the academic year 1994/95.

1.5. Delimitation of the study

Moreover, Grade Eleven students were made the study population the fact that they have gone a long way towards tertiary education where they are to get involved in different writing activities necessitates this study to look into their performances in the tests presented in the multiple choice and short-answer formats.

On the other hand, the restricting of the study only to senior high schools is to avoid the possible bias that could emanate from differences in the attention given to English language teaching in comprehensive high schools or the vocational schools.

1.6. Definition of Terms Used

The following terms or concepts are clarified to avoid ambiguities.

- Senior high schools: According to the Ethiopian educational set-up, they show secondary schools that deal only with pure academic subjects (Dejene, 1990).
- Comprehensive high schools or vocational schools: They refer to educational institutions that give training in different fields such as Electricity, Home economics, Wood work, Business, Metal work, and Commerce and other academic subjects as well (Ibid, 1990).

- Tests: Short procedures which are used to evaluate students' mastery of a material in a limited scope (Ebel, 1979).
- Examinations: More comprehensive procedures which are used to evaluate students' mastery of a material in a wider scope than tests (Ibid, 1979).
- Subjects: Sample of students who took part in the study.
- Low-achievers: Students who scored 53% and below on their first semester English examination.
- Average-achievers: Students who scored between 53% and 65% on their first semester English examination.
- High-achievers: Students who scored 65% and above on their first semester English examination.
- Reliability: The degree to which a test produces similar results on different occasions under similar condition (Heaton, 1988).
- Repeated measures: The same sample of students used for more than one treatment (Otto, 1976).

CHAPTER TWO: REVIEW OF RELATED LITERATURE

This chapter reviews the literature of particular importance to the present study.

2.1. Achievement Tests

Achievement tests are based on what has been taught and the way it has been taught. They are based on what students are presumed to have learned. This means that the teaching and testing are closely interrelated as it is impossible to separate one from the other (Heaton, 1988).

2.2. Test Specifications

Ebel(1979) suggests that test constructors should use the following important points as guidelines to prepare good achievement tests:

- Objectives of the items to be used.
- Formats of test items to be employed.
- Kinds of tasks the items will provide.
- Number of tasks of each type.
- Areas of content to be sampled.
- Number of items in each area.

Hughes(1989) also points out that a test constructor should primarily know what s/he wants to assess and the purpose of the test, and then write specifications in order to clearly know the areas to be included in the test.

Moreover, Heaton(1988:13) explains that:

Before starting to write any test items the test constructor should draw up a detailed table of specifications showing aspects of the skill being tested and giving a comprehensive coverage of the specific language elements to be included. A classroom test should be closely related to the ground covered in the class teaching, an attempt being made to relate the different areas covered in the test to the length of time spent on teaching these areas in class.

2.3. The Multiple-Choice and the Short-Answer Formats

The multiple-choice format requires students to select an answer from a number of options, whereas the short-answer format requires them to produce answers (Carey, 1988; Weir, 1990).

As regards the utilization of formats, Frary (1985:21) says: "Whether to use free-response or multiple-choice tests in specific educational settings is a persistently recurring question." He feels that there are limited studies in the literature concerning a comparative study of students' scores on the multiple choice and the short-answer formats.

However, some of the studies which have been carried out on this issue are the following: Sax and Collet, as quoted by Frary(1985), have found out that both formats are equally effective in encouraging student learning.

Paterson, as quoted by Ebel(1979), in a similar investigation found out that the two formats (multiple-choice and short-answer) measure the same thing.

Regarding students' mean scores on the multiple-choice and the short-answer formats, Frary(1985) notes further that there were significant mean differences because tests in the ~~Form~~one as they are affected by guessing and testwiseness. This situation appears to support Heaton's(1988:26) view that: " Objective tests are frequently criticized on the grounds that they are simpler to answer than subjective tests."

As concerns students' ranks, McMorris et al (1986) point out that many studies on the effects of different formats in achievement tests have reported that the use of different formats had little effect on ranking of students even though there can be significant mean differences between mean scores on different formats.

Nitko(1983) and Cangelosi(1990) are of the opinion that the short-answer format could help students exhibit what they have learned, while on the other hand, multiple-choice format avoids self-expression and restricts students to select what has been produced by others.

Furthermore, Lukhele et al(1993) rationalize that the multiple-choice format is suitable for evaluating static knowledge, and that it has been challenged as being inadequate to fully assesses students performances, have prompted a search for alternatives to the multiple-choice format. They believe that the short answer format could offer such an alternative. They further note that the short-answer format can enable one to measure traits that can't be tapped by the multiple-choice format, and that it also measures a dynamic cognitive process by replicating tasks students face in academic and work settings.

On the contrary, Brown(1983) argues that it is hardly possible to declare whether the multiple-choice format or the short-answer format is advantageous in measuring students' achievements, and that there is no definite evidence to say that a certain format

is better than the other. In brief, he feels that the situation in which a test is to be administered and the purpose it serves should be specified to conclude whether the multiple-choice format of the short-answer format is preferable or otherwise.

To remedy the problem which may occur as a result of using a single format, Allen(1965), Heaton(1988), Dejenie(1990), and Taddese(1990) advocate the use of eclectic formats to assess students' real language mastery. They add that tests could have inherent weaknesses when restricted to a single approach of testing no matter how attractive that approach might be.

Weir(1990:5) supports that:

There is also evidence in literature that the format of a task can unduly affect the performance of some candidates.... This makes it necessary to include a variety of test formats for assessing each construct rather than on a single overall measure.

Heaton(1991:27) also says: " A classroom test will usually contain both subjective and objective test items."

In Ethiopia, as Tesfaye(1982) and Taddese(1990) illustrate, the change in the ESLC English examination format in which precis and the writing components are totally avoided, gave rise to excessive focus on the multiple-choice format in senior high schools. This has resulted in a situation where productive skills have been given no consideration. This caused tests and examinations to have a negative washback effect on teaching.

Bogale(1994) says that testing is a very unnoticed area in the Ethiopian senior high schools' English syllabus. He states that the syllabus requires teachers to offer both objective and subjective (Sentence construction and short paragraph writing) questions.

Nevertheless, in practice starting from elementary school all the way to tertiary level, the multiple-choice and a few true-false items are being employed predominantly.

It is worth noting Heaton's(1988:6) view here that:

Just as it is necessary for the doctor first, to diagnose the patients' illness, so it is equally necessary for the teacher to diagnose the students' weaknesses and difficulties.

Ingram(1975:315) argues:

If you want to know how good a person is at writing essays, you ask him to write an essay, if you want to know how fluent he is in a foreign language, you ask him to talk to you.

According to the two quotations above, if teachers are aware of why, what, and how to evaluate, they will have an insight into possible changes to be made either in teaching or testing.

Carey(1988) also stresses that there should be:

- a match of items with conditions, behaviour, and content specified in the behavioural objectives.
- a match of items with the characteristics of students to be evaluated,
- clarity of items, and
- accuracy of measures.

He stresses that the primary consideration to assess an item is its match with the objective it measures. He adds that without this match, the appropriateness of the item for the target students, the clarity of item wording and the accuracy of measures are meaningless.

Remmers et al(1954) also underline that a teacher should determine the fitness of each format for each defined situation rather than making generalizations that a certain format is the best for all situations. Formats shouldn't be decided according to teachers' personal preference, but rather according to what the teaching syllabus dictates. If, however, the selection is made on the basis of what the teachers think, there will be a mismatch between the format and the objective it serves.

In Ethiopia, both teaching and testing appear to be dictated by the ESLC English examination format. Teachers don't seem to follow what the teaching syllabus of English recommends.

Furthermore, Dejenie(1990:3) says that:

As regards the teaching and testing of the language in particular, it is a common understanding that the multiple choice method of the ESLC English examination is dictating both the teaching and testing methods not only in secondary schools but also in the junior secondary and elementary schools.

2.4. Discrete Point Item and Integrated Tests

A discrete point item test emphasizes on specific points to assess students' mastery of each item. Students are expected to know each point individually. The assumption is that students either know or don't know the point questioned. A lot is required from the students to do as the test doesn't give clues. It encourages the students to answer questions irrespective of contexts, and enables teachers to diagnose the language item the students have weaknesses or strengths in (Davies, 1988).

On the contrary, an integrated test considers each language item at one time. It serves the situation when students are required to know the relationship between each point covered in the teaching syllabus. But it doesn't specifically show the area in which students have or don't have problems (Ibid,1988; Weir, 1990).

Oller, in Weir(1990:3), says that:

The concept of integrative test was born in contrast with the definition of discrete point tests. If discrete items take language skill apart, integrative tests put it back together. Whereas discrete items attempt to test knowledge of language one bit at a time, integrative tests attempt to assess a learner's capacity to use many bits all the same time, and possibly while exercising several presumed components of grammatical system, and perhaps more than one of the traditionally recognized skills or aspects of skills.

In addition, Davies(1988), and Weir(1990) say that no testing approach should be considered irrelevant to measure students' achievements so far it serves the teaching/learning process at a certain period of time.

They emphasize that discrete point item tests play an important role to reveal students' performances in a certain language aspect, whereas integrated tests measure the students' ability to use two or more skills simultaneously. In discrete point item tests, flaws in one item will not affect the other one because they are independent bits. However, in integrated tests, if there is an error in some aspects of the questions, there could be a problem in the whole system of the questions.

Concerning the utilization of all necessary testing approaches, Heaton (1988) states that a useful test will generally consist of features of varying approaches.

2.5. Communicative and Traditional Tests

Dejenie(1990) quotes that communicative tests deal with tasks which approximate 'real-life', while traditional tests deal with separate elements having little or no context.

Heaton(1990) says that students performances in communicative tests enable teachers to predict the students' ability to perform the same task in 'real-life' rather than in traditional tests.

Traditional tests lack the potential of measuring how students use the target language in 'real-life' like situation, whereas communicative tests lack the potential of measuring how students answer decontextualized questions (Davies,1988; Heaton,1990; Weir,1990.)

Rea(1991) states that the existence of differences between the two approaches is not a matter of superiority; it should be, rather, that one measure may be testing a language aspect in a different manner than the other.

On the other hand, because a communicative approach lacks a clear-cut definition, different people have different views on communicative teaching and testing. Davies(1988:14) affirms this by saying:

Communicative language tests have no clear-cut definition. As with communicative language teaching they often mean different thing to different people.

To sum up, the reviewed related studies are of great significance to tackle the stated problem of the present study in that they provide evidences for the probable outcomes of the investigation to be carried out on the impact of the multiple-choice and the short-answer formats on students' achievements.

CHAPTER THREE: METHODOLOGY

3.1. Subjects

3.1.1. Selection of Schools

As stated in chapter one, government senior high schools were made the target population of the study. Among them, three were randomly selected. These were Abyot Kirs, Ayer Tenna, and Higher Four.

3.1.2. Selection of Students

To investigate the multiple-choice and the short-answer formats' impact on what students are expected to perform, Grade Eleven students enrolled during the 1994/95 academic year in the three schools were sampled. This was because they have gone a long way to tertiary education and are on the verge of getting involved in different writing activities.

From among the three schools (Ayer Tenna, Higher Four, and Abyot Kirs) consisting 700, 1100 and 1300 students, respectively, 155 students were sampled with the help of the systematic random sampling method. This was done by taking every 20th student from each school's eleventh graders' list. The number of the students taken from Abyot Kirs, Ayer Tenna, and Higher Four schools were 65, 35, and 55, respectively.

Finally, the students' first semester results were collected, and their score mean and standard deviation were computed to classify them into the three achievement levels. Accordingly, their mean and standard deviation were found to be 59 and 6,

respectively. For the purpose of identifying average-achievers, standard deviation was added to the students' mean score (59+6 and 59-6). On the basis of this computation those students whose marks fell between 53 and 65 were placed in the category of average-achievers, whereas those students whose marks were greater than or equal to 53 and 65 to the left and to the right were taken as low- and high-achievers, respectively. Then 45, 50, and 60 students were categorized under high-, average-, and low-achievers respectively. So that the number of the student at each achievement level correspond with each other and also could be manageable for comparison, their number in each achievement level was trimmed down to 45. Twenty Students were randomly left out. This reduced the students' number to 135. Of these, only 120 students participated in the study because 15 students didn't respond to the call.

3.1.3. Selection of Teachers

From among 12 English teachers of Grade Eleven in the three schools, 9 were randomly selected for the study. Three teachers were selected from each school.

3.2. Instruments

The instruments employed in this study were two tests (which were on the same content presented in the multiple-choice and the short-answer formats), and two questionnaires, one each for teachers and students.

3.2.1. Tests

In order to prepare the tests in a way they reflect the portions the English teachers covered, the researcher asked them the portions they have covered, and the approximate time they spent on teaching them. Accordingly, the tests' specifications were made:

Table 1: Table of Specifications for the Two Tests (Multiple-Choice and Short-Answer)

NO.	OBJECTIVES	CONTENT AREAS	KINDS OF TASK	THE PROPORTIONS OF QUESTIONS TAKEN FROM EACH PORTION	
				FREQUENCY	PERCENTAGE
1	Answering Comprehension questions	Reading passage	Answering according to the information given in the passage in their own words	6	24
2	completing blank spaces	Vocabulary	Using vocabulary items in other sentences according to the context given in the other preceding sentences	8	32
3	Applying grammar points which have been achieved so far	Grammar and usage	Connecting sentences (Questions 1,4,11)	3	12
			Sentence construction by using <u>wonder</u> (Question 2)	1	4
			Using appropriate prepositions (Questions 3, 8)	2	8
			Using <u>unless</u> in place of <u>if</u> . (Question 5)	1	4
			Changing words given in brackets into the impossible condition form. (Question 6)	1	4
			Using an appropriate tense (Question 7)	1	4
			Using <u>so...as</u> in place of <u>less...than</u> . (Question 9)	1	4
			Converting a sentence into an interrogation (Question 10).	1	4
Total				25	100

N.B. Questions are numbered according to their sequence in the short-answer test.

The test items were sampled from the specified sections, and then given to three Grade Eleven English teachers randomly selected from the three schools to check whether or not the tests are content and face valid. According to some suggestions raised by the three subject teachers, some modifications were made and the tests were re-written for final administration. [Please see Appendices I and III.]

3.2.2. Questionnaires

The questionnaires, both for the students and teachers were designed in two main parts to elicit information on their attitudes towards the multiple-choice and the short-answer formats, and their actual classroom practices. Each part had ten items.

In both the teachers' and the students' questionnaires, the items dealing with attitudes were divided into two sub-parts: items dealing with their attitudes towards the multiple-choice format, and items dealing with their attitudes towards the short-answer format. [Please see Appendices VI and VII.]

3.3. Procedures

Here are the procedures followed in the study:

- After completing the preparation of the instruments, the researcher oriented the subject teachers on the purpose of the study and told them to inform the selected students that they (the students) would have tests on both the multiple-choice and the short-answer formats.
- Sixty and hundred minutes were allotted for the multiple-choice and the short-answer tests, respectively.

- The two tests were administered successively, with a difference of fifteen minutes with the short-answer test administered first. To interrupt a positive knowledge transfer, in the interval, a puzzle test was administered, but with no credit. Finally, the puzzle test was collected and the multiple-choice test was distributed.
- The tests were administered by three subject teachers who were told to promise the students that the results in both tests would be given value and that the value would be added to their results at the end of the semester of the year.
- For both tests, marks were counted in different ways. Each item in the multiple-choice test was given one mark, whereas each item in the short-answer test was categorized according to the approximate time it requires to answer, as well as the accuracy and appropriacy of the answers required. Each comprehension question, each vocabulary and two structure question, and each of the remaining structure question was marked out of three, one, and two, respectively. In general, the multiple-choice test was marked out of a total of **25**, whereas the short-answer test was marked out of 46.
- After the administration of the tests, the students' questionnaire was administered by the three subject teachers. The teachers were told to read every item and give explanations where necessary to avoid misunderstandings. The students responded to the questionnaire at the same time in the respective halls of the three schools.

- For both tests, answer keys were prepared. And marking guidelines were prepared for the short-answer test. One rater was randomly selected from each of the three schools. The purpose the of the marking guidelines was to bring the raters into a line while marking. In order that they rate the students paper reliably, the raters were trained beforehand (please see Appendices IV and V.)

Both the Students' and the teachers' attitudes and actual class room practice items were coded in the following ways:

Each item on attitudes was coded ranging from strongly agree (5 points) to strongly disagree (1 point) to a positive statement, and from strongly agree (1 point) to strongly disagree (5 points) to a negative statement.

As regards the students' questionnaire, items 1a, 4a, 5a, 2b, 4b and 5b were negatively phrased, while 2a, 3a, 1b and 3b were positively phrased.

In the teachers' questionnaire, items 1a, 2a, 4a, 5a, 1b, 2b, 4b and 5b were negatively phrased, whereas 3a and 3b were positively phrased.

Each item on the frequencies of employing the two formats (Multiple-Choice and Short-Answer) in class room exercises and tests, was coded ranging from always true (5 points) to never true (1 point) to a positively phrased statement, and from always true (1 point) to never true (5 points) to a negatively phrased statement.

In the students' questionnaire, 1, 3, 5, 6, 8, 9 and 10 were positively phrased, whereas 2, 4 and 7 were negatively phrased.

Concerning the teachers' questionnaire, 1, 3, 4, 5, 8 and 9 were positively phrased, while 2, 6, 7 and 10 were negatively phrased.

3.4 Statistical Information

The scores on both tests were analysed on the basis of :

- ANOVA (to see interactive effects of the students' achievement levels and the test formats)
- t - dependent for repeated measures (to test mean differences of the students' and the teachers' attitudes towards the two formats, and the students' mean scores difference on the two tests).
- reliability (to check the dependability of the instruments employed), and
- correlation coefficient (to see whether or not the students' scores on both tests correspond with each other).

The dependability of each instrument was calculated as follows:

- The inter-rater reliability of the short-answer test was calculated with the help of

$$\text{Cronbach alpha } (\alpha = \frac{k}{k-1} \left(1 - \frac{\sum S_i^2}{S_t^2} \right))$$

Where k = the number of items

$\sum S_i^2$ = the sum of the variances of each score.

S_t^2 = the variance of the sample.

The reliability coefficient was found to be ($r = 0.998$) which is highly reliable (Please see appendices VIII and IX).

- The reliability of the multiple-choice test was calculated on the basis of Kuder

$$\text{Richardson 21 } \left(K - R_{21} = \frac{K}{K-1} \left(1 - \frac{\bar{x}(1-\bar{x})}{S_t^2} \right) \right)$$

Where k = the number of items

S_t^2 = the variance of the sample.

The reliability coefficient was found to be ($r = 0.81$). (Please see Appendix XIII.)

- The reliability coefficient of the students' questionnaire on their attitudes towards the multiple-choice and the short-answer formats, and their responses to actual English teaching practices were calculated by using the K-R21. The coefficients were ($r = 0.55$ and $r = 0.53$), respectively. (Please see Appendices XIV (a), (b), XV(a) and (b).)
- The reliability of the teachers' questionnaire dealing with their attitudes towards the two formats (multiple-choice and short-answer), and actual classroom practices were calculated with the help of Cranbach alpha

$\left(\alpha = \frac{k}{k-1} \left(1 - \frac{\sum S_i^2}{S_t^2} \right) \right)$ and were found to be ($r = 0.4$ and 0.9), respectively.

(Please see, Appendices XVI (a) and XVII.)

CHAPTER FOUR: RESULTS AND DISCUSSION

In this chapter are presented the results and the discussions of the two tests and the two questionnaires.

Among one hundred and twenty students which took part in the study, sixty-five were male and fifty-five were female. Only one student was above twenty years old, while the remaining were below twenty.

Nine teachers (one female and eight male) were considered in the study. One was below thirty years old, whereas the rest were between thirty-one and fifty.

Except two teachers who were university drop outs, the remaining ones were holders of Bachelor of Arts degree.

Three teachers were assigned to teach below twenty periods per week, whereas six were made to teach between twenty and thirty periods per week.

4.1. Results of the Tests

Analysis of the data collected by the aforesaid instruments has yielded the following results.

In this phase of the study, a two-way analysis of variance (ANOVA), and t-dependent for repeated measures were used to see the test formats' and the achievement levels' impact on the students' achievements, and mean comparisons of their scores, respectively.

Table 2: Analysis of variance of the students according to their performances in the two tests.

Source of Variation	Sum of squares	DF	Mean Square	F
Row (achievement level)	47540.063	2	23770.0315	75.114*
Column (test format)	32139.147	1	32139.147	101.56*
Interaction (achievement level X test format)	1892.2	2	946.1	2.989
Error variance (within subjects)	74049.54975	234	316.45	

* $P < 0.05$

DF = degree of freedom

F = F-ratio

In the above table, sum of squares and mean squares across row (achievement levels) and column (test formats) are indicated. Furthermore, an interaction between the achievement levels and test formats, and within-subjects' variations are presented.

As concerns the analysis of variance of the students according to their performances in the multiple-choice and short-answer tests, Table 2 indicated that there exist a statistically significant difference in the students' performance across achievement levels ($F=75.114$) irrespective of the test formats. In brief, a significant difference appeared in the students' achievements on the basis of their achievement levels.

Similarly, there was a statistically significant difference in the students' performance across the test formats (the multiple-choice and the short-answer) ($F=101.56$) regardless of their achievement levels. In gross, differences occurred in

the students' performance with changes in the test formats and the students' achievement levels.

Furthermore, as the same table displays, there was no statistically significant interaction between the test formats and the students' achievement levels. This was made clear by a nonsignificant value of F ($F=2.989$). That was, the two variables did not interact with each other as the students' achievement levels were not found to be affected by the test formats. Students at the three achievement levels invariably scored higher on the multiple-choice test than on the short-answer.

Table 3: A t-test computed for the mean difference in the students' performance on both tests having the same content presented in the multiple-choice and the short-answer formats. ($N=40$ for each achievement level).

Achievement level	Multiple-choice		Short-answer		Multiple-choice versus short-answer	
	\bar{x}	Sd	\bar{y}	Sd	t_0	t_{cr}
Low	42	17.01	18	14.14	11.27*	
Average	55	16.29	25	13.96	13.09*	
High	72	18.50	56	23.67	5.10*	
Total	56	21.01	33	24.18	17.7*	1.980

$df=38$, t_{cr} (for the three achievement levels)=2.021

* $P<0.05$

t_0 = the calculated value of t-test

t_{cr} = the table value of t-test

\bar{x} = mean score on the multiple-choice test

\bar{y} = mean score on the short-answer test

y

Sd = standard deviation

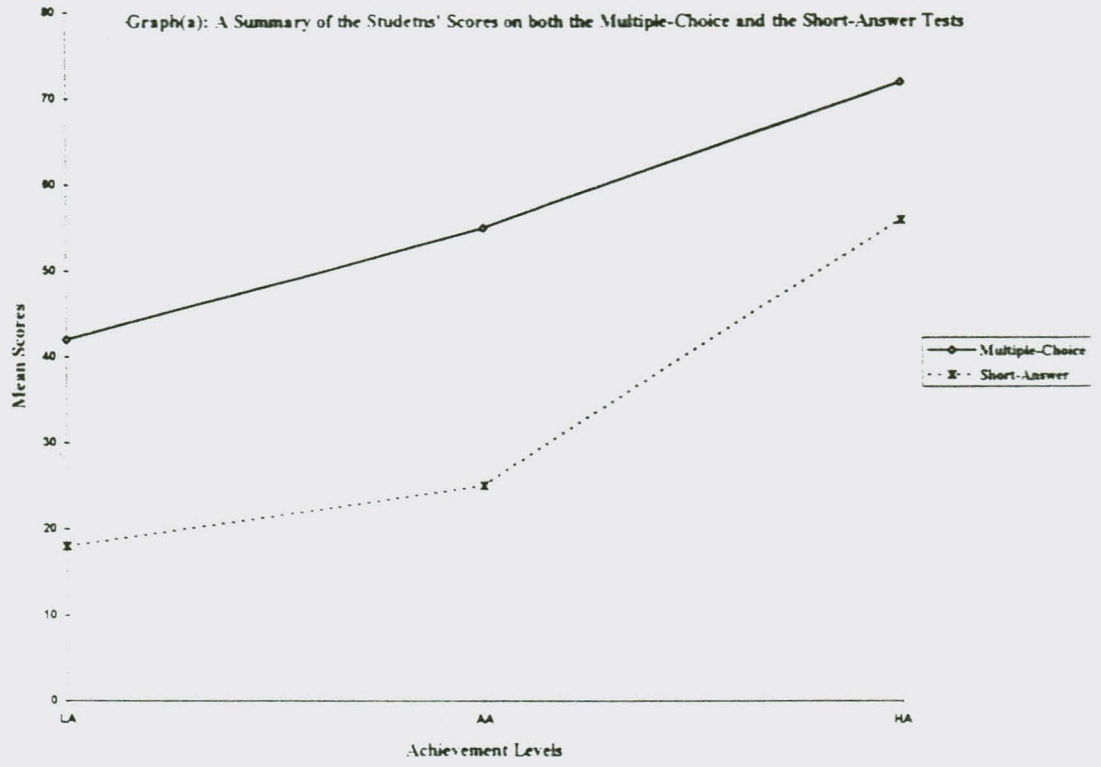
The above table indicates the mean scores of each achievement level and that of the overall performance.

Table 3 displays that there was a statistically significant mean difference in the students' performance in the multiple-choice and the short-answer tests. In all achievement levels (high, average, and low), the t-calculated exceeds the t-critical. The t-calculated for each achievement level were found to be 11.27, 13.09, and 5.10, respectively, whereas the critical value of t for all of them was 2.021. In general, the students had higher mean score on the multiple-choice test as compared to their mean score on the short answer test. The students, as a group, had less mean score in the short-answer test than in the multiple-choice test. The present finding falls in line with what Frary(1985) has come up with. He found out that statistically significant mean difference occurs in students' performance when the same content tests are presented in the multiple-choice and short-answer formats and that lower mean scores were obtained on the short-answer test. He concluded that most likely the differences occurred due to guessing and testwiseness effects in the test presented in the multiple-choice format. Similarly, McMorris et al (1986) reported that presenting the same content test in the multiple-choice and the short-answer formats causes differences in students' achievement.

A closer look at the raw mean scores of the students shows that the low-achievers failed in the two tests, whereas the average-achievers passed in the multiple-choice test and failed in the short-answer test. But as can be discerned from Table 2, the high achievers have passed in both tests even though their mean score in the short-answer test was lower than that in the multiple-choice test.

One of the reasons for the students' lower achievements in the short-answer test could be due to the dominantly employed multiple-choice tests and/or examinations, and that the students base their study procedures on the nature of test and/or examinations they expect to appear in the ESLC English examination. Dejenie(1990) says that the ESLC English examination is dictating both the teaching and testing methods in secondary schools and elementary schools as well. He adds that rejecting other formats which are treated in the English textbooks and relying completely on the multiple-choice format makes teachers and students be certain about what to expect in the ESLC English examination.

The other possible reason for earning lower mean score on the short-answer test might be because of the students' format preference. As Weir(1990) remarks, some students may have a preference for a certain test format. He feels that this tendency could have adverse effects on students' achievements.



LA = Low-Achiever
AA = Average-Achiever
HA = High-Achiever

Table 4: A t-test for a difference in performances in the components (comprehension, vocabulary, and structure) of the multiple-choice test.

df=118, N=120

Comprehension (x)		Vocabulary (y)		Structure (z)		Comprehension versus vocabulary		Comprehension versus structure		Vocabulary versus structure	
\bar{x}_i	SD	\bar{y}_i	sd	\bar{z}_i	sd	to	tcr	to	tcr	to	tcr
46	24.9	55.7	24.3	63.8	25.9	4.04*	2.0	7.7*	2.0	2.8*	2.0

* $P < 0.05$

\bar{x}_i = mean score on comprehension

\bar{y}_i = mean score on vocabulary

\bar{z}_i = mean score on structure

In the table presented the students' mean scores on each component of the multiple-choice test. Moreover, standard deviation, the t-obtained (the calculated value of t-test) and t-critical (the table value of t-test) are indicated.

Moreover, the students' achievement in each component (comprehension, vocabulary, and structure) of the two tests were calculated to see whether or not they have statistically significant mean differences. Accordingly, in the multiple-choice test, as indicated in Table 4, there were statistically significant mean differences between the students' performance in comprehension versus vocabulary, comprehension versus structure, and vocabulary versus structure. In the structure, the students scored the highest mean score, while they scored the least on the comprehension. The students' low mean score on the comprehension

could imply that they did not cultivate appropriate reading skills that they read and comprehend to answer questions drawn from passages.

On the other hand, it could also be because the degree of blind guessing is reduced by the reading passage. Smith and Smith in Frary(1985) say: "...the opportunity to look back at the reading passage nullified the advantage from guessing on multiple-choice items."

Generally, the students failed in the reading comprehension, whereas they passed in the vocabulary and the structure parts. Despite the fact that they passed in the vocabulary and the structure parts, there was a statistically significant difference between their mean scores.

Table 5: A t-test for a difference in performances in the components (comprehension, vocabulary, and structure) of the short-answer test.

df=118, N=120

Comprehension (xo)		Vocabulary (yo)		Structure (zo)		Comprehension versus vocabulary		vocabulary versus structure		Comprehension versus structure	
\bar{x}_o	sd	\bar{y}_o	sd	zo	sd	to	tcr	to	tcr	to	tcr
31	24.8	52	35.5	28	26	7.67*	2.0	9.85*	2.0	1.41	2.0

* P<0.05

\bar{x}_o = mean score on comprehension

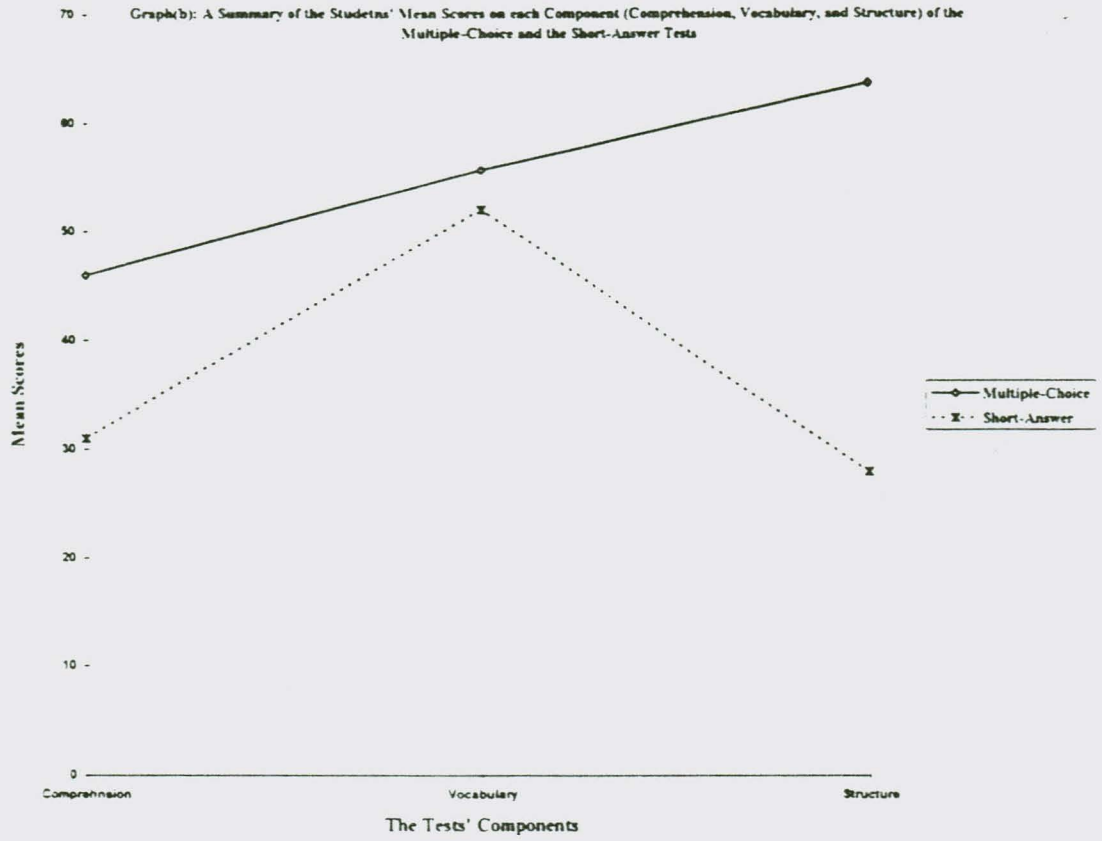
\bar{y}_o = mean score on vocabulary

\bar{z}_o = mean score on structure

The above table presents mean scores and standard deviations. In addition, the calculated value of t-test and the table value of t-test are indicated.

In the short-answer test, as indicated in Table 5, the students scored the highest mean score on the vocabulary part, whereas they scored the least mean score on the structure part. There were statistically significant mean differences between the students' mean scores on comprehension and vocabulary, and vocabulary and structure. But the mean difference between comprehension and structure was not statistically significant.

In sum, the students passed in the vocabulary, while they failed in the comprehension and the structure. This shows that they had serious problems in the components of the short-answer test as compared to that of the multiple-choice test. In other words, this implies that the students had lower performance in the short-answer than in the multiple-choice test. This result is consistent with those results which are indicated in Tables 1 and 2.



4.2. Results of the Questionnaires

Analysis of the questionnaires dealing with the attitudes of the students and those of the teachers, and the frequency of employing the multiple-choice and the short-answer formats in classroom exercises and examinations is presented hereunder.

4.2.1. Results of the students' Attitudes

Table 6: Difference between the Students' Attitude Scores on the Multiple-Choice and the Short-Answer Formats.

Formats	\bar{X}	to	tcr	df	Significance level
Multiple-choice	15.1	1.4	2.0	118	Not significant
short-answer	15.7				

$P > 0.05$

In the above table, mean scores of the students' attitudes towards the multiple-choice and the short-answer formats are presented. A t-value of 1.4 is obtained for the difference in mean scores of their attitudes towards the mentioned formats with 118 degree of freedom.

Table 6 shows that there was no statistically significant difference between the students' attitudes towards the multiple-choice and the short-answer formats. In spite of their favourable attitudes towards the two formats, their performance did not fall in line with their attitudes towards the formats.

4.2.2. Results of the Teachers' Attitudes

Table 7: Difference between the Teachers' Attitude Scores on the Multiple-Choice and the Short-Answer Formats.

Formats	\bar{X}	to	tcr	df	Significance level
multiple-choice	13.78	5.5*	3.335	118	Significant
short-answer	17.33				

* $P < 0.05$

As can be discerned from Table 7, there was a statistically significant difference between the teachers' attitudes towards the multiple-choice and the short answer format in favour of the latter one. However, this did not seem to contribute to the achievements of the students. As regards the impact of teachers' attitudes on students' attitudes and performance, Medhanie(1986) described that if teachers' attitudes towards English teaching is positive, the students' attitudes will most likely be positive and their performances satisfactory. If the contrary, however, the attitudes of the students will basically be negative and their performances unsatisfactory.

Therefore, according to what Medhanie(1986) has described the students' achievements in the short-answer test should have exceeded their achievements in the multiple-choice test.

Overall, as far as the attitude items in both questionnaires are concerned, the findings have shown that the teachers' attitudes towards the short-answer was by and large favourable. But the students' attitudinal tendency, as shown in Table 5,

was not found to be favourable to a particular format. They seem to be happy if eclectic formats are used in teaching and testing.

Hereunder, the students' and the teachers' responses to actual classroom practices were discussed.

4.2.3 Results of the Students' Responses to Actual Classroom Practices Questionnaire

Table 8: The Students' Responses to Actual Classroom Practices Questionnaire

Item no.	STATEMENTS	Responses											
		always true		usually true		sometimes true		rarely true		never true		Total	
		f	%	f	%	f	%	f	%	f	%	f	%
1	My teacher employs the multiple-choice format in classroom exercises	48	40	47	39	5	4	13	11	7	6	120	100
2	Even though classroom exercises are in the multiple-choice format tests and/or examinations are prepared in the short answer format.	17	14	31	26	8	7	34	28	30	25	120	100
3	My teacher uses the multiple-choice format in tests and/or examinations	24	20	45	38	10	8	27	22	14	12	120	100
4	Classroom exercises are in the short-answer format, but tests and/or examinations are in the multiple-choice format.	32	17	33	28	9	6	33	28	13	11	120	100

		Responses											
Item no.	STATEMENTS	always true		usually true		sometimes true		rarely true		never true		Total	
		f	%	f	%	f	%	f	%	f	%	f	%
		5	My teacher uses the short-answer format in classroom exercises.	35	29	59	49	6	5	12	10	8	7
6	Both classroom exercises, and tests and/or examinations are in the short-answer format.	15	12	19	16	13	11	41	34	32	27	120	100
7	My teacher avoids the short-answer format from tests and/or examinations.	34	28	53	44	14	12	11	9	8	7	120	100
8	My teacher advises me that I prepare myself to answer questions prepared in the short-answer format.	31	26	39	33	17	14	18	15	15	12	120	100
9	My teacher encourages me that I prepare myself to answer questions prepared in the multiple-choice format.	19	16	28	23	19	16	40	33	14	12	120	100

		Responses											
Item no.	STATEMENTS	always true		usually true		sometimes true		rarely true		never true		Total	
		f	%	f	%	f	%	f	%	f	%	f	%
		10	In tests and/or examinations, my teacher gives equal chance to both the multiple-choice and the short-answer formats	26	21	45	38	13	11	19	16	17	14

The Students' responses to actual classroom practices were discussed as follows:

In response to item one that says the multiple-choice format is used in classroom exercises, 40%, 39% and 4% said always true, usually true, and sometimes true, respectively, while, 11% and 6% reported rarely true and never true, respectively. In general, 83% reported positive responses whereas 17% reported negative responses.

To item two that says classroom exercises are in the multiple-choice format while tests and/or examinations are in the short-answer format, 14%, 26% and 7% said always true, usually true, and sometimes true, respectively, while 28% and 25% reported rarely true and never true, respectively. These differences in employing formats in teaching and testing could imply that some teachers follow what the English teaching syllabus while others might not do that.

In item three, the students were required to report whether or not their teachers undertake the multiple-choice format in test and/or examinations. On the basis of this item, 20%, 38% and 8% remarked always true, usually true and sometimes, respectively. On the other hand, 22% and 12% reported rarely true and never true, respectively.

In response to item four, 61% of the students indicated positive responses (always true=27%; usually true=28%, and sometimes=61%), whereas the remaining responses were negative (rarely true=28%, and never true=11%).

As regards item five that says the teachers employ the short-answer format in classroom exercises, and tests and/or examinations are in the short-answer format, 39% of the students revealed positive responses (always true=12%, usually true=16%, and sometimes=11%), while 61% reported negative responses (rarely true=34%, and never true=27%).

In item seven, the students were required to report whether or not their teachers avoid the short-answer format in tests and/or examinations. In response to this item, 84% positive responses (always true=28%, usually true=44%, and sometimes true=12%) were reported, while 16% negative responses (rarely true=9% and never true =7%) were demonstrated.

As concerns item eight that requires the students to report as to whether their teachers encourage them to write, 73% of the students reported positively (always true=26%, usually true=33%, and sometimes=14%), whereas the remaining (27%) reported negatively (rarely true=15% and never true=12%).

In response to item nine that requires the students to respond to how often their teachers encourage them so that they prepare themselves to answer questions presented in the multiple-choice, 55% of the students remarked positively (always true=16%, usually true=23%, and sometimes=16%). On the other hand, 45% of the students indicated negatively (rarely true=33% and never true=12%).

4.2.4. Results of the Teachers' Responses to Actual Classroom Practices' Questionnaire.

Table 9: The Teachers' Responses to Actual Classroom Practices' Questionnaire

ITEM NO.	STATEMENTS	RESPONSES											
		Always true		Usually true		Sometimes true		Rarely true		Never true		Total	
		f	%	f	%	f	%	f	%	f	%	f	%
1	I employ the multiple-choice format in classroom exercises.	2	22	3	34	1	11	2	22	1	11	9	100
2	I use the multiple choice format in classroom exercises because my students like exercises of that type.	-	-	4	45	3	33	1	11	1	11	9	100
3	Because the English teaching syllabus doesn't recommend the use of the multiple-choice format in classroom exercises, I don't employ it.	3	33	2	23	3	33	1	11	-	-	9	100
4	I employ the short answer format in classroom exercises, because this format is given a special attention in the English teaching syllabus.	4	45	1	11	-	-	1	11	3	33	9	100

ITEM NO.	STATEMENTS	RESPONSES											
		Always true		Usually true		Sometimes true		Rarely true		Never true		Total	
		f	%	f	%	f	%	f	%	f	%	f	%
5	I use the short-answer format in classroom exercises to improve my students' writing ability.	2	22	4	45	1	11	1	11	1	11	9	100
6	Because my students don't like the short-answer format, I don't employ it in classroom exercises.	1	11	5	56	1	11	2	22	-	-	9	100
7	I make use of the short-answer exercises in class, however, my tests and/or examinations are purely in the multiple choice format.	2	22	4	45	-	-	2	22	1	11	9	100
8	I don't employ the short-answer format in classroom exercises because it is not dictated by the English teaching syllabus.	2	22	2	22	2	22	1	12	2	22	9	100
9	As the English teaching syllabus recommends the use of both the multiple-choice and the short-answer formats in classroom exercises, I employ both formats.	1	11	6	67	-	-	2	22	-	-	9	100

ITEM NO.	STATEMENTS	RESPONSES											
		Always true		Usually true		Sometimes true		Rarely true		Never true		Total	
		f	%	f	%	f	%	f	%	f	%	f	%
10	Because the marking of short-answer tests and/or examinations is time consuming, I do not employ the short-answer format in tests and/or examinations.	1	11	4	45	1	11	2	22	1	11	9	100

The teachers' responses to actual classroom practice were discussed as follows:

To item one that inquires the teachers to report whether or not they employ the multiple-choice, 67% of them responded to it positively (always true=22%, and never true=11%). Nevertheless, 33% responded to the same item negatively (rarely true=22%, and never true=11%).

In item two that requires the teachers to report whether or not they use the multiple-choice format because their students like, 78% of them commented positively (usually true=45%, and sometime=33%). But 22% responded to the same item negatively (rarely true=11%, and never true=11%).

In item three, 89% of the teachers reported positively (always true=33%, usually true=23%, and sometimes=33%), that they don't employ the multiple-choice format in exercises because the English teaching syllabus doesn't recommend to use it. On the contrary, 11% remarked that it is recommended by the syllabus and they use it in classroom exercises.

In item four, the teachers were inquired to remark whether or not the short-answer format is given a special focus in the English teaching syllabus. In response to this item, 56% of them responded positively (always true=45% and usually true=11%) that it is dictated by the teaching syllabus and they make use of it. Nevertheless, 44% reported negatively (rarely true=11% and never true=33%) that it is not recommended by the teaching syllabus and they don't employ it in teaching.

As concerns item five that requires the teachers to report as to whether or not they use the short-answer format to improve their students' ability to write, the responses of 78% of the teachers were positive (always true=22%, usually

true=45%, and sometimes true=11%), while the remaining (22%) responded to the same item negatively (rarely true=22% and never true=22%).

In response to item six, 78% of the teachers reported positively (always true=11%, usually true=56% and sometimes = 11%) that they don't employ the short-answer format in class exercises the fact that their students don't have favourable attitudes towards it. On the other hand, 22% said that they employ the short-answer format regardless of their students' attitudes towards the format.

In item seven, the teachers were asked to report whether their teaching and testing formats reflect and/or support each other. In response to this item, 67% of them have given positive responses (always true=22% and usually true=45%) that they utilize the short-answer format in classroom exercises, but their tests and/or examinations are in the multiple-choice format. The remaining (33%), however, reported that there is little mismatch between their teaching and testing formats. In short, these responses reveal that there is a mismatch between the formats utilized in teaching and testing.

In general, on the basis of the students' and the teachers' responses to actual classroom practices, it could be concluded that there is a mismatch between teaching and testing formats.

CHAPTER FIVE: SUMMARY, CONCLUSION AND RECOMMENDATIONS

This chapter deals with a brief summary of the main objectives, methods and results of the study as well as recommendations made.

5.1. Summary

To look into the impact of the multiple-choice and the short-answer formats, two test on the same content were prepared in three parts each.

The two tests were administered to 120 Grade Eleven students enrolled during the 1994/95 academic year in three government senior secondary schools in Addis Ababa. The students were divided into three achievement levels (high, average, and low) on the basis of their first semester English results in order to find out the impact of the two formats on their achievements. Moreover, two questionnaires, each in three parts, were designed for the students and their teachers to collect information concerning their attitudes towards the two formats and the frequency of employing them in classroom exercises, and tests and/or examinations.

The reliability coefficients of the instruments were computed to check whether or not they were dependable to use in the study. Accordingly, they were found to be reliable to yield consistent results.

The results of the tests revealed that there was a statistically significant difference between the students' mean scores on the two tests. The students as a group and at each achievement level invariably scored lower means on the short-answer test than on the multiple-choice test. In all cases, they were found to be weak in the short-answer test.

On the other hand, to diagnose the area in which the students had weakness or strength, in both tests, the mean scores of the students on the three components of each test were compared. As the correlated samples of the two-tailed t-test exhibited, there were statistically significant mean differences between the students' mean scores on each component of the multiple-choice test (comprehension versus vocabulary, comprehension versus structure, and vocabulary versus structure). But as concerns each component of the short-answer test, there were no statistically significant differences among all components. There were statistically significant differences between comprehension and vocabulary, and vocabulary and structure, while there was no statistically significant difference between their mean scores on comprehension and structure. Regarding the multiple-choice test, of all components, the students scored the highest and the least mean scores on structure and comprehension, respectively. Nevertheless, they scored the highest and the least mean scores on vocabulary and structure, respectively on the short-answer test.

As regards the results of the students' attitude questionnaire, there was no statistically significant mean difference between their attitudes towards the multiple-choice and the short-answer formats. However, there was a statistically significant mean difference between the teachers' attitudes towards the multiple-choice and the short-answer format, in favour of the latter.

Concerning the students' and the teachers' responses to actual classroom practices, the summary tables on their responses show that there is a mismatch between teaching and testing formats. However, there were inconsistencies between what the students and the teachers reported. In response to item one, the majority of the students (83% = always true =40%, usually true =39%, and

sometimes true = 4%) reported that their teachers use the multiple-choice format in classroom exercises. On the other hand, 89% of the teachers (always true = 33%, usually true = 23% and sometimes true = 33%) remarked that the multiple-choice format is not recommended by the English teaching syllabus and they don't use it in teaching.

As their responses to item six indicate, the majority of the teachers reported positive responses (always true = 11%, usually true = 56% and sometimes true 11%) that they don't use the short-answer format in classroom exercises because their students are not in favour of it. Here, the teachers' responses pinpoint that they tune their teaching to students' preferences.

5.2. Conclusion

The results of the study indicated that although the students scored significantly higher on the multiple-choice test, they failed in the short-answer which had the same content with the former test. Thus, the lower mean scores on the short-answer could show that the writing ability of the students is poorly cultivated. Presumably, the students' higher performance in the multiple-choice test might be attributed to guessing. Consequently, their high scores on the multiple-choice couldn't be an evidence for actual language mastery.

The ESLC English examination format could also have an impact on both the teachers' method of teaching and testing, and the nature of the study the students carry out. The nature of the examination expected by both teachers and students could dictate the teaching and learning system. No matter the efforts made to do exercises in the short-answer format, students might adjust the way they study to

activities involving selection since the format of the English examination expected to appear in the ESLC is purely in the multiple-choice format.

To detect the area in which the students had difficulties in the components of each test, their mean scores on the components of each test were compared. On this basis, the students scored below 50% in the structure and comprehension parts of the short-answer test, while they scored below 50% in the comprehension part of the multiple-choice test. Overall, the students lower mean scores in the components of the short-answer (comprehension and structure) revealed that they had difficulties in these areas. This might show that they were deficient in answering these components. In both tests, the students scored below 50% in the comprehension questions. This is likely to say that the students had more serious problems in reading comprehension than in any other components.

In other respects, despite the fact that their achievements were unsatisfactory in the short-answer test, it was statistically justified that they had significantly favourable attitudes towards both formats (multiple-choice and short-answer). This implies that the students seemed to be happy if eclectic language teaching and testing approaches are introduced to Ethiopian senior high schools' English teaching and testing system. But that the teachers had statistically significant positive attitudes towards only the short-answer format could encourage the use of a single format. This indicates that they seemed to say the short-answer format is the best.

Finally, the students' and the teachers' responses to actual classroom practices demonstrated tests and/or examinations are predominantly in the multiple-choice format, whereas both formats are employed in classroom exercises.

5.3. Recommendations

Based on the findings, the following recommendations are made.

- The Ministry of Education should orient English teachers so that they give appropriate attention to each format dictated by the English teaching syllabus.
- To evaluate students' achievements in the way they have been taught and diagnose problems they may have in certain areas, all possible formats treated in teaching should be employed in tests and/or examinations.
- Eclectic teaching and testing approaches should be introduced to the teaching-learning process.
- Appropriate teaching materials geared to developing language skills should be prepared.

BIBLIOGRAPHY

- Ademe and Gebre (1990). "The Impact of objective Type Tests on the Learning of High School Mathematics." The Ethiopian Journal of Education, Vol.11, No. 2: 28-49.
- Allen, Harold B. (1965). Teaching English as a Second Language. New York: McGraw Hill Book Company.
- Alderson, J.C. (1993). "Standards in Testing", Language Testing. Vol. 10, No.1: 1-92.
- Anastasi, Anne (1976). Psychological Testing (4th ed.). New York: Collier Macmillan, Ltd.
- Bachman, L.F. (1990). Fundamental Considerations in Language Testing. Oxford: Oxford University Press.
- Berk, R.A. (1982). Hand book of Methods for Detecting Test Bias. Baltimore: The John Hopkins University Press.
- Bogale Tesema (1994). "Continuous Assessment: Its validity, Reliability and Feasibility" (M.A. Thesis), Addis Ababa University.
- Brown, Annie (1993). "The Role of Test-Taker Feedback in the Test Development Process: Test-Takers' Relations to a Tape-Mediated Test of Proficiency in Spoken Japanese," Language Testing, Vol. 10, No. 3: 277-304.
- Brown, Frederick G. (1983). Principles of Educational and Psychological Testing. New York: Holt, Richard and Winston, Inc.

- Cangelosi, J.S. (1990). Designing Tests for Evaluating Student Achievement. London: Longman Group Ltd.
- Henning, Grant (1993). "Automated Assembly of Pre-equated Language Proficiency Tests," Language Testing, Vol.11, No. 1: 14-28.
- Carey, L.M. (1988). Measuring and Evaluating School Learning. Boston: Allyn and Bacon, Inc.
- Carrol, B.J. (1980). Testing Communicative Performance. England: Pergamon Press Ltd.
- _____ (1983). "The English Language Testing Service," Case Studies in ELT, London: William Collins sons and Co. Ltd., pp:99-101.
- Cooper, Robert L. (1968). "An Elaborated Language Testing Model," Language Learning Special Issue. Vol. 18, No. 3: 55-77.
- Davies, Alan (1988). "Communicative Language Testing," Testing English for University Study, ELT Documents: 127. pp:1-4.
- _____ (1991). "Principles of Language Testing," Testing in 1990s, Vol. 1, No. 1: 136-149.
- Davies, Charles W.(1984). The Uses and Misuses of Tests. London: Jossey-Bass Ltd.
- Dejenie Leta (1990). "Measurement, Washback and Proficiency in School Leaving Examination: A Case of Innovation in an Ethiopian Setting." (Ph.D. Thesis). Addis Ababa University.

- _____ (1994). "Language Testing and Its Practical Application: The Testing of Reading in Focus" (A paper presented at a Workshop on the English Language Teaching at Nazareth). Addis Ababa University.
- Desforges, Charles (1989). Testing and Assessment. London: Eassell Education Ltd.
- Ebel, R.L. (1979). Essentials of Educational Measurement (3rd ed.). New Jersey: Prentice Hall, Inc.
- Ebel, R.L. and David A.F. (1991). Essentials of Educational Measurement (5th ed.). New Jersey: Prentice Hall, Inc.
- Fallik, Fred and Bruce Brown (1983). Statistics for Behavioural Sciences. Illinois: The Dorsey Press.
- Firth, D.S. and H.G. Mac Intosh (1987). A Teacher's Guide to Assessment. Glasgow: Bell and Bain Ltd.
- Frary, R.B. (1985). "Multiple-choice Versus Free-Response: A simulated Study," Journal of Educational Measurement, Vol. 22, No.1: 21-31.
- Gerberich, J.R. etal. (1962). Measurement and Evaluation in the Modern School. New York: David McKay Company, Inc.
- Goldberg, M.L. etal.(1966). The Effects of Ability Grouping. New York: Teachers' College Press.
- Gronlund, N.E.(1976). Measurement and Evaluation in Teaching. New York: Macmillan Publishing Co., Inc.
- _____ (1988). How to Construct Achievement Tests (4th ed.). New Jersey: Prentice Hall, Inc.

- _____ (1990). Measurement and Evaluation in Teaching(6th ed.). New York: Macmillan, Inc.
- Harrison, Andrew (1989). A Language Testing Handbook. London: Macmillan Publishers Ltd.
- Heaton, J.B. (1988). Writing English Language Tests. London: Longman Group UK Ltd.
- _____ (1990). Classroom Testing. London: Longman Group UK Ltd.
- _____ (1991). Writing English Language Tests. (New ed.). London: Longman UK Ltd.
- Hedge, Tricia (1988). Writing. Oxford: Oxford University Press.
- Hughes, Arthur (1989). Testing for Language Teachers. Cambridge: Cambridge University Press.
- Ingram, Elizabeth (1968). "Testing in the Context of Language Learning," Language Learning. Special Issue. Vol. 18. No. 3: 147-161.
- _____ (1974). "Language Testing," Testing Techniques in Applied Linguistics. Vol. 3: 313-343.
- _____ (1977). "Basic Concepts in Teaching," Testing and Experimental Methods. Vol. 4: 11-35.
- Jaradat, Derar and Sari Sawaged (1986). "The Sub-Selection Technique for Multiple-Choice Tests: An Empirical Inquiry." Journal of Educational Measurement. Vol. 23. No. 4: 369-376.
- Lado, Robert (1961). Language Testing: The Construction and Use of Foreign Language Tests. London: Longmans.

- Little John, Andrew and Diana Hicks (1987). "Task centered Writing Activities," Language Learning Tasks, Englewood Cliffs: Prentice-Hall, Inc. Vol. 7: 69-92.
- Lord, Frederick (1980). Applications of Item-Response Theory to Practical Testing Problems. New Jersey: Lawrence Erlbaum Associates Publishers.
- Lukhele, Robert et al. (1993). On the Relative Value of Multiple-choice, Constructed-Response, and Examine-Selected Items on Two Achievement Tests. New Jersey: Educational Testing Service.
- Lyman, H.B. (1986). Test Scores and What They Mean (4th ed.). New Jersey: Prentice-Hall Englewood Cliffs.
- Madson, H.S. (1968). "A Statistical Analysis of the 1968 ESLC English Language Examination," Ethiopian Journal of Education, Vol. 3, No. 1: 2-9.
- _____ (1983). Techniques in Testing. New York: Oxford University Press.
- Mcmorris, R.F. et al. (1986). "Differences in Instructional Sensitivity between Item formats and between Achievement Test Items," Journal of Educational Measurement. Vol. 23. No. 1: 1-12.
- Melvin, R.N. (1985). Standards for Educational and Psychological Testing. New York: American Psychological Associations, Inc.
- Mulugeta Wadajo (1959). Ethiopian School Leaving Certificate Examination Report. Addis Ababa: Curriculum Department.