

ADDIS ABABA UNIVERSITY
SCHOOL OF GRADUATE STUDIES
FACULTY OF LANGUAGE STUDIES
DEPARTMENT OF ENGLISH

AN ASSESSMENT OF THE CONTENT VALIDITY OF GRADE 9
ENGLISH LANGUAGE TEACHER-MADE CLASSROOM
TESTS: (Four Secondary Schools of Bahir Dar in Focus)



BY: Yeshambel Almaw Bitew

June, 2011
Addis Ababa

**AN ASSESSMENT OF THE CONTENT VALIDITY OF GRADE 9
ENGLISH LANGUAGE TEACHER-MADE CLASSROOM**

TESTS: (Four Secondary Schools of Bahir Dar in Focus)

By

Yeshambel Almaw Bitew

A Thesis Submitted to:

The School Of Graduate studies of

Addis Ababa University



In Partial Fulfillment of the Requirements for the Degree of Master
of Arts in Teaching English as a Foreign Language (TEFL)

**June, 2011
Addis Ababa**

ADDIS ABABA UNIVERSITY
SCHOOL OF GRADUATE STUDIES
FACULTY OF LANGUAGE STUDIES
DEPARTMENT OF ENGLISH

**AN ASSESSMENT OF THE CONTENT VALIDITY OF GRADE 9
ENGLISH LANGUAGE TEACHER-MADE CLASSROOM
TESTS: (Four Secondary Schools of Bahir Dar in Focus)**


BY: Yeshambel Almaw Bitew



Approved by the Board of Examiners

Dr. Mekasha Kassaye
Advisor

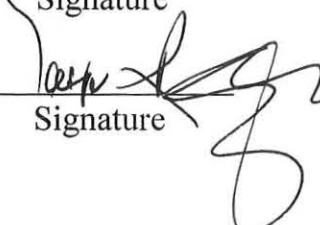
Dr. Taye Regassa
Examiner



Signature

15 June 2011

Date



Signature

June 15, 2011

Date

Acknowledgements

I would like to express my heart felt gratitude to my adviser, Dr. Mekasha Kassaye, who devotedly corrected my work and gave me constant guidance through out this study.

I also would like to extend my thanks to the English language teachers of Tana Haik, Ghion, Fasilo and Bahir Dar Zuria secondary schools, who have devoted their time in allotting the time for the tasks of the textbook, and to Addis Ababa university for financial assistance it provided.

Table of Contents

| | Pages |
|--|-------|
| Contents | i |
| Acknowledgements ----- | i |
| Table of Contents ----- | ii |
| List of Tables ----- | v |
| List of Appendices ----- | vi |
| Abstract ----- | vii |
| Chapter One: Introduction | |
| 1.1. Background of the Study----- | 1 |
| 1.2. Statement of the Problem----- | 2 |
| 1.3. Purpose of the Study----- | 3 |
| 1.4. Research Questions----- | 4 |
| 1.5. Significance of the Study----- | 4 |
| 1.6 .Scope of the Study----- | 4 |
| 1.7. Limitation of the Study----- | 5 |
| Chapter Two: Review of Related Literature | |
| 2.1. Teacher- made Classroom Tests ----- | 6 |
| 2.2. The Link between Language Teaching and Testing----- | 7 |
| 2.3. Purpose of Tests----- | 8 |
| 2.4. Quality of Tests----- | 10 |
| 2.4.1 Validity Types----- | 12 |
| 2.4.1.1 Face Validity----- | 12 |
| 2.4.1.2 Criterion-related Validity----- | 13 |
| 2.4.1.3 Construct Validity ----- | 13 |
| 2.4.1.4 Content Validity----- | 14 |
| 2.4.1.4.1 What Happens if Tests Lack Content Validity?----- | 14 |
| 2.4.1.4.2 Threats to Content Validity ----- | 16 |
| 2.4.1.4.3 Guidelines to Establish Content Validity of Tests----- | 17 |
| 2.4.1.4.4 Table of Test Specification and Content Validity ----- | 17 |
| 2.4.1.4.5 How to Assess Content Validity? ----- | 18 |
| 2.5. Test Formats ----- | 19 |
| Chapter Three: Research Methodology | |
| 3.1. Sources of Data----- | 21 |
| 3.2. Sampling Techniques and Sample Size----- | 21 |

| | |
|--|----|
| 3.2.1 Selection of Schools ----- | 21 |
| 3.2.2 Selection of Grade Level----- | 21 |
| 3.2.3. Selection of Teachers ----- | 22 |
| 3.2.4. Selection of Students----- | 22 |
| 3.3. Instruments for Data Collection ----- | 22 |
| 3.3.1. Document Analysis ----- | 22 |
| 3.3.2. Questionnaire ----- | 23 |
| 3.3.2.1 Questionnaire for Teachers----- | 23 |
| 3.3.2.2. Questionnaire for Students----- | 23 |
| 3.4. Procedure of Data Collection ----- | 24 |
| 3.4.1. Textbook Analysis----- | 24 |
| 3.4.2. Questionnaire Administration ----- | 25 |
| 3.4.3. Sample Test Papers Collection ----- | 26 |
| 3.5. Data Presentation and Analysis----- | 26 |
| 3.5.1. Contents of the Textbook and Sample Test Papers ----- | 26 |
| 3.5.2. Formats of the Textbook and Sample Test Papers ----- | 27 |
| 3.5.3. Questionnaire of Teachers and Students ----- | 27 |

Chapter Four: Data Analysis Results and Discussion

| | |
|--|----|
| 4.1. General description of Grade 9 English language Text book----- | 28 |
| 4.2 Content Analysis of Grade 9 English Language Textbook and Sample Test Papers ----- | 29 |
| 4.2.1. The List of Reading Tasks in the Textbook with the Allotted Time and the Frequencies of Test Items in Each Task----- | 29 |
| 4.2.2. The List of Vocabulary Tasks in the Textbook with the Allotted Time and the Frequencies of Test Items in Each Task ----- | 30 |
| 4.2.3. The List of Grammar Tasks in the Textbook with the Allotted Time and the Frequencies of Test Items in each Task----- | 31 |
| 4.2.4. The List of Speaking Tasks in the Textbook with the Allotted Time and the Frequencies of Test Items in Each task----- | 32 |
| 4.2.5. The List of Listening Tasks in the Textbook with the Allotted Time and the Frequencies of Test Items in each Task ----- | 33 |
| 4.2.6. The List of Writing Tasks in the Textbook with the Allotted Time and the Frequencies of Test Items in Each Task----- | 34 |

| | |
|--|----|
| 4.2.7. A Comparison between the Summary of the Time Coverage Given for the Main Content Areas and the Summary of Frequencies of Tasks in the Sample Test Paper ----- | 35 |
| 4.2.8. Ranking the Six Main Content Areas in their Time Allotment and Frequencies of Test Items ----- | 36 |
| 4.3. Analyses of the Formats of Grade 9 English Language Textbook and the Sample Test Papers ----- | 40 |
| 4.3.1 Frequencies of Test Formats in the Textbook and in the Sample Tests ----- | 40 |
| 4.3.2. Comparison of the Test Formats in the Textbook and Sample Test Papers ----- | 41 |
| 4.4. Responses of Subjects----- | 43 |
| 4.4.1. Responses of Teachers----- | 43 |
| 4.4.1.1. Teachers' Responses to Limitation, Coverage and Proportion of Content----- | 43 |
| 4.4.1.2. Teachers' Responses to Variety and Proportion of Test Formats----- | 46 |
| 4.4.1.3. Teachers' Responses to the Emphasis they Give while Teaching and Testing----- | 47 |
| 4.4.2. Students' Responses ----- | 49 |
| 4.4.2.1. Students' Responses to the Match, Coverage and Proportion of Item content ----- | 49 |
| 4.4.2.2. Students' Responses to Variety and Proportion of Test Formats ----- | 50 |
| 4.4.2.3. Students Response to the Emphasis Teachers Give in Teaching and Testing----- | 51 |
| Chapter Five: Summary, Conclusion, and Recommendation | |
| 5.1. Summary of the Findings----- | 53 |
| 5.2. Conclusion ----- | 55 |
| 5.3. Recommendations----- | 56 |
| Bibliography ----- | 57 |

List of Tables

| | |
|---|----|
| Table 1. The List of Reading Tasks in the Textbook with the Allotted Time and the Frequencies of Test Items in Each Task ----- | 30 |
| Table 2. The List of Vocabulary Tasks in the Textbook with the Allotted Time and the Frequencies of Test Items in Each Task ----- | 31 |
| Table 3. The List of Grammar Tasks in the Textbook with the Allotted Time and the Frequencies of Test Items in each Task----- | 32 |
| Table 4. The List of Speaking Tasks in the Textbook with the Allotted Time and the Frequencies of Test Items in Each task ----- | 33 |
| Table 5. The List of Listening Tasks in the Textbook with the Allotted Time and the Frequencies of Test Items in each Task ----- | 34 |
| Table 6. The List of Writing Tasks in the Textbook with the Allotted Time and the Frequencies of Test Items in Each Task ----- | 35 |
| Table 7. A Comparison between the Summary of the Time Coverage Given for the Main Content Areas and the Summary of Frequencies of Tasks in the Sample Test Papers -- | 36 |
| Table 8. Ranking the Six Main Content Areas in their Time Allotment and Frequencies of Test Items ----- | 36 |
| Table 9. Frequencies of Test Formats in the Textbook and Sample Tests ----- | 41 |
| Table 10. Comparison of the Test Formats in the Textbook and Sample Test Papers----- | 41 |
| Table 11. Teachers' Responses to Limitation, coverage and Proportion of Content ----- | 45 |
| Table 12. Teachers' Responses to Variety and Proportion of Test Format----- | 47 |
| Table 13. Teachers' Responses to the Emphasis they Give while Teaching and Testing----- | 48 |
| Table 14. Students' Responses to the Match, Coverage and Proportion of Item Content----- | 50 |
| Table 15. Students' Responses to Variety and Proportion of Test Formats ----- | 51 |
| Table 16. Students' Responses to the Emphasis Teachers Gave in Teaching and Testing----- | 52 |

List of Appendices

| | Page |
|--|------|
| 1. Time Allotted by Teachers for the Different Tasks of the Textbook ----- | 60 |
| 2. Determination of the Inter Rater Reliability among Raters ----- | 62 |
| 3. Sample Test Papers | |
| 3A. Tana Haik Secondary School Mid- Test----- | 63 |
| 3B. Tana Haik Secondary School Final- Test----- | 65 |
| 3C. Ghion Secondary School Mid- Test----- | 68 |
| 3D. Ghion Secondary School Final- Test----- | 69 |
| 3E. Fasilo Secondary School Mid- Test ----- | 71 |
| 3F. Fasilo Secondary School Final- Test ----- | 72 |
| 3G. Bahir Dar Zuria Secondary School Mid- Test----- | 75 |
| 3H. Bahir Dar Zuria Secondary School Final- Test----- | 76 |
| 4. Frequency of Sample Test Items by Question No. and Content Areas | |
| 4A Tana Haik Secondary School Mid and Final Tests. ----- | 79 |
| 4B. Ghion Secondary School Mid and Final Tests----- | 80 |
| 4C. Fasilo Secondary School Mid and Final Tests----- | 81 |
| 4D. Bahir Dar Zuria Secondary School Mid and Final Tests----- | 82 |
| 5. Teachers' Questionnaire----- | 83 |
| 6. Students' Questionnaire English Version----- | 86 |
| 7. Students' Questionnaire Amharic Version----- | 88 |

Abstract

The purpose of this study was to assess the content validity of Grade 9 English language teacher-made classroom tests of four government secondary schools at BahirDar.

For this purpose, the time allotted for the different content areas of the textbook, one progress and one achievement tests of the four sample schools and test formats of the textbook and the sample tests were used as main data sources. To supplement the data that were obtained from the above sources, questionnaires were also administered to Grade 9-English language teachers and their students. The time coverage for the different content areas of the textbook was allotted by four experienced English language teachers of Grade 9. The frequencies of the contents of the sample test items were analyzed under the main content areas. The test formats of the textbook and sample tests were analyzed by counting frequencies. Questionnaires were also used to assess the teachers' opinions about the representativeness of the classroom tests. To crosscheck teachers' responses, questionnaires were also administered to students.

The collected data were analyzed quantitatively and qualitatively. The time allotted to the content areas of the textbook and the frequencies of the contents of the sample tests, and the frequencies of the test formats of the textbook and the sample tests were analyzed in figures and percentages. Then, the magnitude of the relationship between the time allotment for the content areas in the text book and the contents of the sample test items, and the formats of the textbook and the formats of the sample test papers were determined by using the Spearman's Rank Order Correlation Coefficient. Figures, Percentages and qualitative analyses were also used to analyze the data from the questionnaire.

The findings show that the contents of the sample test items generally maintained moderate (0.48) correlation with the sum of the time allotted to the content areas of the textbook. However, in specific terms, the different content areas in the textbook like listening, writing and vocabulary were not adequately represented in the sample tests. The formats of the sample test papers and the textbook were also poorly (0.1) correlated. The dominant test formats in the textbook were not considered in the sample tests.

Based on the findings, it is concluded that the Grade 9 English language classroom teacher-made tests relatively lack content validity. Then, recommendations were made according to the findings.

Chapter One: Introduction

In this chapter, the background of the study, statement of the problem, purpose of the study, research questions, significance of the study, scope of the study, and limitation of the study were treated.

1.2. Background of the Study

As English language is used as medium of instruction for most subjects in secondary schools of Ethiopia, students at this grade levels are expected to be proficient in this language in order to achieve general academic success. In relation to this, English language teaching materials, as stated in the syllabus, are prepared in the way that can fulfill the above needs of students. All the major and sub- language skills are treated in the textbooks.

As testing and teaching should have close relationship, the classroom testing practice at these grade levels is, then, expected to reflect the teaching which has been practiced before. In this regard, Heaton (1991:14) has said the following:

if students have followed the structural approach to language learning, it is clearly unfair to administer communicative achievement tests at the end of their course. It is equally unfair to administer a structured based test to those students who have followed a communicative approach to learning a target language.

The above idea informs one that as the Grade 9 and 10 text books reflect communicative approach, the testing trend at these grade levels is expected to be communicative. The classroom tests need to be the direct reflection of the contents of the textbook. This can help learners to be proficient and successful in their learning as it has been intended. This requires a test which provides learners with all the major and sub-language skills and appropriate testing techniques. Such tests can provide information about students' performance in every content area.

It is obvious that the fundamental task of testing in education is to provide information for making decisions. This reminds one that whenever there is a test there is a decision to be made.

In relation to this, Harrison (1983) has informed test makers that the usefulness of the information that could be derived from a test will depend up on the amount of care that is taken in its preparation. Then, to be relied up on with the information they provide, tests should

maintain the desired qualities. These are validity, reliability, and practicality. Validity, specifically content validity is a major concern (quality) in achievement testing (Grondlund 1982).

Different scholars, Hughes (2008), Heaton (1991) and Grondlund(1982) have defined content validity as a term that concerns with the adequacy of sampling. Tests are expected to sample adequately the contents and the methods of the concerned textbook. Content validity helps to check students' performance/ levels of progress in each content area of the syllabus/ textbook. And, as test method might affect students' performance, the different test formats in the teaching material should be used in the tests. A test which lacks content validity is unlikely to be accurate. Such a test is likely to have harmful backwash effect. Areas which are not tested are likely to become areas ignored in teaching and learning (Hughes, 1989).

In relation to this, many testing scholars have emphasized the need for content validity for classroom progress and achievement tests. For example, Brown (2004:32) says, "the major source of validity in a classroom test is content validity". This informs one that English language teachers, as much as possible, should strive to achieve the content validity of their classroom tests.

Then, though there are different types of tests aimed at different purposes, this paper aimed at assessing the content validity of Bahir Dar town Secondary schools classroom progress and achievement tests because nearly all English language teachers are involved in preparation and use of such tests. So, the researcher feels that a study on content validity, especially, in relation to classroom test is vital.

1.2. Statement of the Problem

Baker (1991:3) has said, "Testing is invariably associated with the making of decisions." Test results can be used as a feedback that serves as a basis for instructional analysis, the determination of students' performance and an assessment as to whether a lesson or a course needs to be continued or readjusted. Then, it can be said that decisions which can be made from test results, especially in the case of progress and achievement tests, can be misleading if the decision is based on tests which lack content validity.

In language testing, as Heaton (1991:13) says, “there is a constant danger of concentrating too much on testing those areas and skills which most easily lend themselves to being tested. This indicates that language tests commonly lack content validity. As a result, statements that could be drawn about students’ language ability from such tests could be misleading. In relation with this, Grondlund(1982) notes test makers that their primary concern in constructing and using achievement test should be with content validity”.

Thus, assessing a test helps to determine the power of the test in functioning as an appropriate and effective measuring instrument of language instruction.

There are different local studies on content validity. Argaw (2004) and Kassaw (2006) have assessed the content validity of Grade 10 and 12 English Language National Examinations respectively. They have concluded that both examinations lacked content validity. Since only multiple choice format is used in national examinations, it is difficult to include some language contents. This can make the situation different from classroom tests. On the other hand, Asmare (2008) and Kifle (1995) have studied college and Grade 10 classroom tests. They both have also concluded that tests lacked content validity.

Asmare did not consider test formats in his study. On the other hand, in Kifle’s study, logistic problems, class size and teachers beliefs have been attributed as the cause for the low content validity of English language teacher-made classroom tests. This study has been conducted almost twenty years ago. As there is a change in textbook and teaching methodology at the present, there might be a change in the practice of testing. And as the researchers experience, class size and logistic problems were not serious issues in Bahir Dar. Furthermore, Sireci (2007) has said that evaluating test validity is not a static, one time event; it is a continuous process. Then, the researcher was interested to assess the content validity of Grade 9 English language teacher-made classroom tests of Bahir Dar secondary schools.

1.3. Purpose of the Study

The focus of this study was to assess the content validity of 2003 E.C Grade 9 English language teacher-made classroom tests in relation to Grade 9 English language textbook.

1.4. Research Questions

In order to attain the above purpose, this study attempted to answer the following questions:

- Are all the major and sub-language skills of Grade 9 English language textbook proportionally included in Grade 9 English language classroom tests?
- Do the formats in the classroom tests adequately reflect the formats in the Grade 9 English language textbook?

1.5. Significance of the Study

The result of this study is hoped to :

- provide information on which content areas have been given and have not been given attention in classroom language tests.
- help administrators of Bahir Dar zone education office to take measure as I have promised to give them a copy of this study.
- increase teachers awareness of language testing..
- help teachers organize and develop content valid English language test

1.6 .Scope of the Study

The main focus of this study was to assess the content validity of classroom tests of four government schools of Bahir Dar town in terms of content and format. It did not include other areas of test qualities. The study was also limited to 2003 E.C First Semester Grade 9 English language classroom progress (mid-term) and achievement (end-term) tests of the sample schools.

1.7. Limitation of the Study

As classroom tests are based on what has been taught, teachers were made to determine the time for comparing the test items with the emphasis teachers gave to different content areas. This means teachers (raters) determine content validity of the test. So, one limitation of the study was the result of the research work highly depended on the view of teachers. Content valid test of such practice may fail to serve the goal of the curriculum. It could have been more valid if the syllabus had specified the time limit for the different content areas of the textbook.

Chapter Two: Review of Related Literature

The aim of this study was to assess the content validity of Grade 9 English language classroom tests. In this chapter, then, theoretical bases for the study were presented. Concepts, which were related directly or indirectly to content validity of classroom tests, were presented from the most important works.

2.1. Teacher- made Classroom Tests

Before talking about teacher-made classroom tests, it is important to define what a test is. Brown (2004:3) has defined a test in the following way: "A test, in simple terms, is a method of measuring a person's ability, knowledge, or performance in a given domain". Brown further elaborated, "a well constructed test is an instrument that provides an accurate measure of the test taker's ability with in a particular domain".

The above definition of test also works for teacher made tests. But, Airasian (1991) further specifies that teacher-made classroom tests are the types of assessment techniques that are controlled by the classroom teacher. The teacher decides who to assess, what to assess, when to assess, how to assess, how to score, and how to use the results of the assessment.

These tests are fundamental ingredients in the teaching learning process in the class room. Heaton (1991) has said that good classroom tests will help to locate the precise areas of difficulty encountered by the class/ by an individual student. In line with this Heaton (1990:24) has also pointed out, "good class room tests should always reflect the teaching that has taken place beforehand" In this point, Heaton not only indicates the importance of classroom tests but also that these tests should be based on what has been taught. This means what has been practiced should appear in the test. Weir (1993:5) also adds, "Classroom testing should not be divorced from the teaching that precedes it". Then, the Grade 9 classroom tests are expected to be based on what has been practiced to have beneficial effect for teaching and learning. This, then, requires to be content valid.

2.2. The Link between Language Teaching and Testing

In no way testing is separated from teaching. They are so interwoven and interdependent that it is difficult to treat them differently. They are the two components of the whole teaching learning process. Scholars like Hughes (2008), Heaton (1991), Bachman (1990), Weir (1990) and Madsen (1983) have articulated the interrelatedness of language teaching and testing.

Every instructional sequence has a testing component. Then, testing should be a reflection of the way we teach. Testing should be a mechanism to recognize to what extent our method of teaching is appropriate to the level of the class, to know students weakness and strength in the teaching process and to indicate the general direction of the program. Pointing the relation ship that teaching and testing should maintain, Davies and Upshur (1993) have stated that the strength of most examinations lie in their relation ship to language teaching.

Whenever there is a change in teaching techniques and teaching materials, the test format and contents are expected to be changed in the same way to maintain their relationship. In this regard, Weir (1993:5) points out the relation ship of teaching and testing as, “changing a test is possibly the most powerful means of bringing about improvements in the learning experiences that go before it”. Similarly, Baker (1991) and Bachman (1990) have indicated that the shift in emphasis in language teaching has inevitably had consequences for language testing. Heaton (1991:14) strengthens the idea by saying,

if students have followed the structural approach to language learning, it is clearly unfair to administer communicative achievement tests at the end of their course. It is equally unfair to administer a structured based test to those students who have followed a communicate approach to learning a target language.

Harrison (1983:1) has also showed the relation ship between testing and teaching in the following way, “A test is seen as a natural extension of classroom work, providing teacher and students with useful information that can serve each as a basis for improvement”. In addition, Madsen (1983) further showed the relation by suggesting ways in which a well made tests can help students and teachers. He said tests help students: creating positive attitude towards the class, helping master the language and giving sense of accomplishment. In a similar way, tests help the teacher know if his/her teaching was effective and can provide insight in to ways to improve the testing process. Hughes (2008) further articulates the relation ship between teaching

and testing as,” testing should not be expected to follow teaching. It should be supportive of good teaching and, where necessary, exert a corrective influence on bad teaching”.

Thus from the interdependence nature of testing and teaching one can understand how powerful testing is in redirecting the focus of education to its essence of good teaching and learning.

2.3. Purpose of Tests

The type of test we give will depend very much on our purpose of testing. In line with this, Brown (1996) explained that each teacher must be clear about his/ her purpose for making a given decision and match the correct type of test to that purpose. According to their purpose, Alderson et al.(1995) classified tests as placement, proficiency, diagnostic, achievement and progress tests.

Placement tests are designed to assess the students’ level of language ability for placement in an appropriate course or class. This kind of tests indicate the level at which the student will learn most effectively. In line with this, Gronlund (1982) said that placement tests are designed to measure prerequisite skills which have low level of difficulty. Putting in a different way, Harmer (2009:166) said, “the purpose of placement tests is to find out not only what students know but also what they do not know”.

“Proficiency tests are designed to measure people’s ability in a language regardless of any training they may have had in that language”(Hughes,1989:10). These tests are designed to assess the over all language ability of students at varying levels. Their purpose is to describe what students are capable of doing in a language. The contents of proficiency test are not based on the content or objective of language courses which people taking the test may have followed. This means these tests are not limited to any course or curriculum. Instead, the contents of the proficiency tests are defined by the test constructor.

Commonly, proficiency tests are not developed by classroom teachers. As these tests are not based on a certain previously taken course, they exercise considerable influence over the method and content of language courses. As a result, these tests have harmful effect Hughes (1989).

Diagnostic tests seek to identify those language areas which students need further help. The information gained from diagnostic tests is crucial for further course activities and providing

students with remediation. Pointing the importance of diagnostic tests, Hughes (1989:14) has said the following:

the lack of good diagnostic test is unfortunate. They could be extremely useful for individualized instruction or self instruction. Learners would be shown where gaps exist in their command of the language, and could be directed to source of information, exemplification and practice.

In a similar vein, Gronlund(1982) has stated that diagnostic tests usually focus on the common errors students make , rather than attempting to sample broadly the intended outcomes of the instruction.

According to Heaton (1991) progress tests are designed to measure the extent to which the students have mastered the material taught in the classroom. This means Progress tests measure the progress that students are making towards defined course or program goals. These tests have the function of seeing how students are getting on with the lessons, and how well they have been taught over the last week, two weeks or a month (Harmer, 2009).

Progress tests are administered at various stages through out the language course to see what students have learned, after certain segments of instruction have been completed. These tests are generally teacher produced and are narrower in focus than achievement tests because they cover a smaller amount of material and assess fewer objectives.

Heaton (1990) says that results from progress tests aid teachers to become more familiar with the work of each student in particular, and with the progress of the class in general. Then, to get reliable information from progress tests, sampling the items from which has been treated in the instruction is vital. In line with this, Baker (1991:4) has said,

A properly written progress test sampling correctly from the course content can be an indication to learners which part of the course need more attention, and to course designers which part of the course have not been effective.

Achievement test is related directly to classroom lessons units or even a total curriculum. In line with this, Weir (1993:5) states, “achievement testing should be firmly rooted in previous classroom experiences in terms of activities practiced, language used, and criteria of assessment employed.” In certain ways, an achievement test is like a progress test as their purpose is to see

what student has learned with regard to stated course outcomes but it is usually designed to cover a longer period of learning than progress tests.

Hughes (1989:11) has said, “ the purpose of achievement tests is to establish how successful individual students, group of students or the course themselves have been in achieving objectives”. At the end of a term ,semester or year we may want to do a final achievement test to see how well students have learnt everything(Harmer,2009).Harmer further added achievement tests include a variety of test types(discrete, integrative) and measure the students abilities in all four skills as well as their knowledge of grammar and vocabulary.

A good achievement test should reflect the particular approach to learning and teaching that has previously been adopted (Heaton, 1991). This means testing should be the reflection of the way we teach. If students were asked to write in the classroom, then, their test should require them to do that.

From the definitions given above, Grade 9 English language end term classroom tests are achievement tests as it is based on the materials covered in the classroom. Specifically, as Heaton (1990), they are informal attainment tests as they measure a mastery of what has been learned over the past term or year.

2.4. Quality of Tests

Certain conditions must be fulfilled before the data which one obtains from tests can be used to make different sound academic decisions. Thus, tests should display certain qualities. If tests do not fulfill the desired qualities, it is difficult to obtain appropriate information from test results. The information which can be gathered from such tests could be misleading. For this reason testing experts, like Hughes (2008) and Weir (2005) said that every language test, be it standard or teacher made, should attain qualities of tests-practicality, comprehensiveness, reliability and validity.

Practicality is one of the qualities of a test. The term practicality involves the question of economy, ease of administration, scoring and interpretation of results. Even valid and reliable tests can be of little use when it is not practical one. In line with this Harmer (2009:167) said, “when designing tests, one of the things we have to take in to account is the practicality of tests”.

Harmer further elaborated “the test will be worthless if it is so long that no one has the time to do it”.

A good test should be teacher friendly that a teacher should be able to develop, administer mark it with in available resource and available time. This means practical issues should consider cost of the test development and maintenance, ease of marking, time, availability of trained marker and administrative logistics. In this regard, Weir (1990:34) has said, “there is clearly an imperative need to try and develop test formats and their evaluation criteria that provides the best over all balance among reliability, validity and efficiency in the assessment of communicative skills”.

A test should be comprehensive. A test may be said comprehensive when it contains an appropriate portion of items from all aspects of the language or material to be tested. In this regard, Atkins, Hailom and Nuru (1996:168) has said the following about comprehensiveness of tests: “the number of items need to be in proportion to the importance of the language area being tested, or in proportion to the problems of students.

However, when a test becomes to long, problems of administration, economy, and examinee fatigue easily arise and can destroy the tests utility. On the other hand, when a test is short, it gives rise to problem of lack of comprehensiveness and doubtful reliability

Reliability is another quality of a test. It refers to the consistency of test scores. It simply means that a test would give similar results if it were given another time. Madsen (1983:179) has elaborated the above point by saying: “a reliable test is one that produces essentially the same results consistently on different occasions when the conditions of the test remain the same”.

A test is reliable if it is consistent and dependable. In line with this, Heaton (1991:162) states, “reliability is the necessary condition of any good test: for it to be valid at all, a test must first be reliable as a measuring instrument” This reliability refers to the test itself and also to the external factors implied. In this sense it refers to condition and personal elements as reliable as possible, considering aspects such as the place of the examination (condition of noise, heat, etc), the person giving the test, the instructions on the test paper and the amount of time allowed. Apart from that, a test should be reliable in the way of correcting and marking. Objectivity and length are also among other factors that contribute to reliability. The longer the test is the more reliable

it becomes. According to Bachman (1990:119-112), the relative importance of the parts of the test, the testing environment, the test rubric, the kind of input provided, the format are among the factors that affect performance and consequently the reliability of the test.

Validity is another important quality of the test. It is one of the most important aspects of measurement and evaluation. Be it teacher made or standardized test, it should be valid. Scholars like, Hughes (2008), Brown (1996), Heaton (1991) have defined validity as the degree to which a test measures what it claims or purports to measure. Henning (1987) in Alderson, Clapham and Wall (1995:170) also defined validity as follows:

validity in general refers to the appropriateness of a given test or any of its component parts as a measure of what it is purported to measure. A test is said to be valid to the extent that it measures what is supposed to measure. It follows that the term valid when used to describe the test should usually be accompanied by the preposition for. Any test then may be valid for some purpose.

In light of the definition above, it can be said that validity has significant role in the language testing. Brown (2004) has confirmed that the most important principle in language testing is validity.

2.4.1 Validity Types

2.4.1.1 Face Validity

Except the difference in wording, scholars like Hughes (2008), Heaton (1991) and Weir (1990) have given the same definition of face validity. It is the tests surface credibility.

Although the concept of face validity is an important one, it is not generally included as one of the three major types of validity due to the questionable role that it plays in the validation process. In line with this Anastasia (1982) in Weir (1990:26) assures that face validity:

is not validity in technical sense; it refers to, not what the test actually measures but to what it appears superficially to measure. Face validity pertains to whether the test 'looks valid' to the examinees who take it, the administrative personnel who decide on its use and other technically untrained observers. Fundamentally, the question of face validity concerns rapport and public relation.

However, Alderson et al. (1995) have given due emphasis for the face validity as tests which do not appear valid for users may not be taken seriously and learners may not perform best of their ability.

2.4.1.2 Criterion-related Validity

This kind of validity is known as empirical validity by scholars like Heaton (1991), and criterion validity by scholars like, Weir (1990) and Hughes (1989). The definition of the concept lies the same: It refers to the extent to which a test is predictive and /or concurrent. According to Heaton (1991:161), this kind of validity can be obtained as a result of comparing the results of the test with the results of some criterion measures such as:

- *an existing test, known or believed to be valid and given at the same time; or*
- *the teachers ratings and any other such form of independent assessment given at the same time;*
or
- *the subsequent performance of the testees on a certain task measured by some valid test; or*
- *the teachers ratings or any other such form of independent assessment given later*

This means if the test and the criterion administered at the same time we speak of concurrent validity, and if it concerns the degree to which a test can predict the candidates future performance it is called predictive validity.

2.4.1.3 Construct Validity

Another kind of validity which is considered as the most important quality of tests by different scholars is construct validity. Heaton (1989:26) has defined the term as, ‘a test, part of a test, or a testing technique is said to have construct validity if it can be demonstrated that it measures just the ability which it is supposed to measure’. Similarly, “Construct validity concerns the extent to which performance on tests is consistent with predictions that we make on the basis of theory of abilities or constructs” (Bachman 1990:254-255).

Heaton (1991:161) also said “construct validity assumes the existence of certain learning theories or constructs underlying the acquisition of abilities and skills”. This means if a test has construct validity, it is capable of measuring certain specific characteristics in accordance with a theory of language behavior and learning. Or a test is said to have construct validity if it accurately reflects

the construct or theory underlying it, that is to say, if it is able to satisfy some previously stated theory against which we validate it.

2.4.1.4 Content Validity

Different scholars have given their own definitions of content validity. The following are some of the definitions: Hughes (1989:22) defined content validity as, “A test is said to have content validity if its content constitutes a representative sample of the language skills, structures, etc with which it is meant to be concerned”. A similar view is given by Kerlinger (1973) in Alderson et al.(1995).According to him, “content validity is the representativeness or sampling adequacy of the content-the substance, the matter, the topics of a measuring instrument”. Anastasia (1988) has also defined content validity as, “the systematic examination of the test content to determine whether it covers a representative sample of the behavior domain to be measured.

From all the above definitions, one can understand that content validity is concerned with what goes in to the test in relation to what has been taught or covered in the syllabus or textbook.

It should be clear from this discussion that the key element in content validity is the adequacy of the sampling. A test is always a sample of the many questions that could be asked. Content validity is a matter of determining whether a sample is representative of the larger universe it is supposed to measure. As Deale (1975) cited in Kifle (1995), this aspects of validity is very important in the context of achievement and progress tests at the classroom and school level.

2.4.1.4.1 What Happens if Tests Lack Content Validity? *

Sampling of the language skills, structures, etc. is very important factor in the success of language tests. Brown (2004:37) explained the importance of content validity in the following way: “A test that achieves content validity demonstrates relevance to the curriculum in question and thereby sets the stage for wash back”.

When all language elements and skills are included in the test, it can put pressure or motivation on the learner to study and cover the whole syllabus. Adequately sampled tests also make teachers focus to teach the specific material within the specific course in the specific time. This means content of the test puts the learner and the teacher on the right track by committing

themselves to the syllabus or textbook. In relation with this, Weir (1990:24-25) supported the above idea as follows:

A primary purpose of many communicative tests is to provide a profile of the student's proficiency, indicating in broad terms the particular modes where deficiencies lie. Content validity is considered especially important for achieving this purpose as it is principally concerned with the extent to which the selection of test tasks is representative of the larger universe of tasks of which the test is assumed to be sample.

It is obvious that tests-among other things-are tools of quality control through which we can judge the success of our pedagogical and educational objectives. Tests can help to keep both the teacher and the learner on the right track of the whole educational process. They can tell the teacher the degree of success or failure of his teaching methods, so he/she can change to new ones if the old ones do not work well. Tests can also help the learner to adapt and modify his/her learning strategies use in his/her career

However, if the test lacks content validity, it spoils the whole educational process. It would spoil the teaching as teachers would focus all their efforts to teach only content areas that may appear in the test. They cannot be engaged in doing real teaching to provide the learners with true language situations. Such tests also spoil the learners as all their learning strategies would be directed to focus on areas that may appear in the tests. In line with this, Hughes (2008) has said that content areas which do not appear in the tests are likely to be ignored in teaching and learning process and this is called backwash effect of tests.

This means tests which involve a proper content validity trigger students to study hard each content of the course. Otherwise, students will avoid studying and practicing those language areas which do not appear or appear less in tests. Hughes has elaborated tests can have negative or harmful wash back when the test contents and techniques deviate from teaching and objective of the course.

In their attempt to enable students pass, teachers may gear their teaching very closely to the tests. If the tests are good, it will have positive effect on teaching. If the tests are bad it may have negative effect on teaching.

Thus, being content valid is one of the criteria that can make tests good, if the test samples proportionally and adequately from all content areas of the syllabus, it can be good test. If it does

not sample proportionally and adequately, it can be bad test and it can have harmful wash back effect.

These all inform one that tests can be used as benchmarks to make different decisions. This can be realized appropriately if tests can achieve the test qualities. Content related validity is one of the qualities and it is the most important kind of *validity for classroom tests* (Brown 2004). Baker (1991:3) strengthened the above ideas by saying:

a properly written progress test sampling correctly from the course content can be an indication to learners which part of the course need more attention, and to course designers which part of the course have not been effective.

2.4.1.4.2. Threats to Content Validity

From the above it is made obvious that the higher content validity a test receives, the more likely it is to be an accurate measure of what it is supposed to measure. As a consequence, test developers, according to Henning (1987), have to avoid the threats to test validity such as misapplication of tests and in appropriate selection of content.

The most noticeable threat of validity arises from misapplication of test. Test developers must consider carefully the designed purpose of a test, including the exact content and objectives to be measured, and the type of examine to whom the measurement is given (Henning, 1987).

Secondly, inappropriate selection of content commonly occurs when items do not match the objective or content of instruction. This may happen when items are not comprehensive in the sense of reflecting all the major points of the instructional program (Henning, 1987).

Putting in a different way, the threats to content validity, Hynes, Richards and kubany (1995) have mentioned three hindrances to successful content qualification of educational assessment. These are: the zero presence of items representing a defined facet of the construct; the presence of items measuring irrelevant facets; and disproportionate weighting of items in deriving an aggregate score.

2.4.1.4.3. Guidelines to Establish Content Validity of Tests

Though content validity is essential for beneficial wash back, establishing it seems problematic because of the nature of language. However, Anastasia (1982) in Weir (1990:25) have recommended the following guidelines to establish content validity:

- *the behaviors domain to be tested must be analyzed to make certain that all aspects are covered by the test items, and in the correct proportion.*
- *the domain under consideration should be fully described in advance, rather than being defined after the test has been prepared.*
- *content validity depends on the relevance of the individuals test response to the behavior area under consideration, rather than on the apparent relevance of item content.*

This means when one embarks on constructing a test, it is important to follow these guide lines to achieve the content validity of tests. This in turn informs one the importance of table of test specification to achieve content validity.

2.4.1.4.4. Table of Test Specification and Content Validity

Testing specialists suggest the importance of test plan .The first important step in designing tests is to prepare test blueprint. Test specification is necessary because it helps to construct a balanced number of questions in each language skills and language areas according to the emphasis given for teaching them. In this regard Heaton (1991:13) has pointed out the use of test specification and what test writers should do when they embark on writing tests as follows:

The important point is that the test writer should attempt to quantify and balance the test components assigning certain value to indicate importance of each component in relation to the other components in the test. In this way, the test can achieve content validity and reflect the component areas and skills which the test writer wishes to include in the assessment.

Alderson, et al. (1995), Bachman (1990) and Hughes (1989) have also given more emphasis to the great role that test content specification plays in maintaining content validity. This means that if teachers use or design test frame work, it can help them to select representative samples and appropriate test format. This, then, contributes to achieve validity, more specifically content validity of the tests.

Furthermore, table of test specification can also help to judge whether or not a test has content validity. A comparison of test content and test specification can be the basis for the judgment. In this regard Brown (1996:233) has summarized as follows:

In order to investigate content validity, testers must decide whether the test is a representative sample of the content of what ever the test was designed to measure .To address this issue, testers or some of their colleagues usually end up making some sort of judgments. To maximize the efficiency of these judgments, the testers may need to focus particularly on the organization of the different type of items that they include on the test and on the specifications for each of item types.

From the above discussion, table of test specification is not only important to achieve content validity of tests but it is also important to judge the content validity of tests. This is because content validity is important both in test development and test use.

However, many scholars argued that it is difficult to decide what to include in language tests. In line with this, Heaton (1990) recommends, a teaching syllabus is of great help in designing a classroom progress or test of achievement”. This informs one that textbooks or syllabuses can be used to draw a test framework. Similarly, Spolsky (1975) in Chappelle (1999) has noted that for classroom tests the problem of validation is not serious, for the textbook or syllabus writer has already specified what should be tested. Then, for this study grade 9 English language text book can be used to draw up a test frame work. Furthermore, Heaton (1990) mentioned that importance of part of the course and length of time spent to teach that particular area can be taken as a criterion to draw appropriate sample of class room tests. Heaton (1991) also forwarded the solution to solve the problem of sampling in language tests as, “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 those areas in class”.

2.4.1.4.5. How to Assess Content Validity?

Though there are no objective methods of assuring the adequate content coverage of an instrument, scholars in testing have recommended different techniques that can help to assess the content validity of tests. To mention some: Alderson, et al. (1995:194-195) provided a useful checklist to evaluate content validity. These are:

- *to compare test content with specification*
- *questionnaire to, interview with experts such as teachers*

➤ *expert judge rate test items and texts according to precise list criteria*

Hughes (2008) also recommended that comparison of test specification and test content is the basis for judgment as to content validity. In a similar way Weir (1990:25) recommended, “a close scrutiny of the specification for proficiency test by experts in the field (or colleagues in the case of classroom achievement tests) and the relating of the specification to the test as it appears in its final form is essential”. This means, in the case of classroom progress and achievement tests, teachers can assess the content validity of tests by comparing contents of the test with the specification. Then in this research context, the content validity of grade 9 English language classroom tests can be assessed by comparing the content of the test items with the specification (the time allotted to the contents of the textbook).

In line with this the above ideas, (Brown, 2004) notes, “who is an expert judge depends on the type of test and its intended use. For example, expert judges of the content validity of an achievement test would be people who teach the courses in the subject area the test covers”. Alderson, et al. (1995) have strengthened the issue that validity is not a matter of all-or-nothing. Then test users can use their own or somebody else’s judgment when deciding on the validity of the test

2.5. Test Formats

Brown (2004:34) states, “the most feasible rule of thumb for achieving content validity in class room assessment is to test performance directly”. Brown further elaborates the idea that test content is not the only type of evidence to support the validity of test. This means test format has also influence to achieve validity, specifically content validity of language tests.

According to Hughes (1989:59), test techniques refer to “means of eliciting behavior from candidates which will tell us about their language abilities.” McNamara (2000:24) also defines test format as, “the way in which the candidates will be required to respond to the material’.

Testing scholars like, Alderson, et al.(1995), Baker (1991), Heaton (1991) and Weir (1990) have pointed out that test format could affect students performance and the scores of individual candidates. This is because reactions of each individual to different formats may vary from person to person. This indicates that when ever teachers embark on preparing test, they should

not forget to use a reasonable combination of different methods as it is much better than using a single method to measure students overall ability. In line with, Weir (1990:42) has articulated the following:

Given the limited state of knowledge concerning the effect of test formats, the only practical approach at present is to safe guard against possible format effect by spreading the base of a test more widely through employing a variety of valid, practical and reliable formats for testing each skill

In this regard Alderson, et al. (1995) also advised item writers to use more than one test method for testing any ability. In connection with this, Heaton (1990) has added that a good test should contain both subjective and objective test items. Furthermore, Davies and Upshur (1993) have suggested that formats should be as varied as possible to meet the needs and interests of most students.

Scholars in the field of testing note that any format has merit and demerit. Then, using variety of formats in line with the current view of language teaching is important. This means for the course which has been based on communicative language teaching, the test formats should be those which can be believed to implement the system. In this regard, Heaton (1990) reminded that subjective formats offer better way of testing language skills and certain area of language than objective format. This means if the test situation wants to test language skills, subjective tests are preferable.

In general, it is possible to conclude from the above discussion that the more different methods a test employs, the more confidence the tester can have that the test is not biased towards one particular method or to one particular sort of learner(Alderson, et al., 1995). Then, the different test formats which have been practiced in the concerned textbook should be sampled reasonably in the classroom tests.

Chapter Three: Research Methodology

This study was intended to assess the content validity of Grade 9 English language teacher-made classroom tests of Bahir Dar town secondary schools. Both quantitative and qualitative methods were employed in collecting and analyzing the data.

3.1. Sources of Data

In assessing the content validity of language tests, Weir (2005) and Alderson et al (1995) have recommended that document analysis-syllabuses, text books, and test papers- and stake holders like expert judges (teachers) and students could be used as source of information. The data sources of this study were the currently used Grade 9 English language Syllabus and Grade 9 English language textbook, which were published in 1996 and revised in 2005, 2003 E.C first semester Grade 9 English language teacher- made classroom progress (mid-term) and achievement (end-term) tests of sample schools, and Grade 9 English language teachers and students of the sample schools.

3.2. Sampling Techniques and Sample Size

3.2.1 Selection of Schools

There were four government secondary schools in Bahir Dar town. The schools were Tana Haik, Ghion, Fasilo and Bahir Dar Zuria secondary schools. Since the number of schools was manageable for the purpose of this research, all the four schools were taken as the sample of the study. The schools in Bahir Dar town were chosen deliberately for two reasons: the first was that the researcher worked as High school teacher in the area. He could not face much difficulty in producing information from respondents and getting test papers. The second reason was that, in the previous research results like Kifle's(1995), class size and logistic problem were attributed to lowering content validity of language tests. The researcher was interested to see the result where the above problems were not serious in urban towns like Bahir Dar.

3.2.2 Selection of Grade Level

Though it was important to assess validity of every grade level of tests regularly (Sireci 2007), Grade 9 was chosen for this study because it was at this grade level English becomes medium of

instruction for most subjects. Students at this level, then, need to attain sufficient mastery of the language to enable them to achieve general academic success. Then, it was important to lay a solid foundation for the basic language skills at this grade level. This could be made successful by having content valid classroom tests. It was important to check whether English language classroom tests achieve content validity or not.

3.2.3. Selection of Teachers

As the number of teachers at the sample schools was manageable, all of the 24 teachers who taught English at Grade 9 were taken. As this research was about classroom tests, four of the experienced teachers were taken purposively to limit the time given for the different tasks under the different content areas. The rest 20 teachers were made to provide supplementary information to the data that were collected from the document analysis. Six teachers from Tana Haik, five teachers from Ghion, seven teachers from Fasilo and six teachers from Bahirdar Zuria, a total of 24 teachers were taken as the sample.

3.2.4. Selection of Students

There were 15 sections from Tana Haik, 12 sections from Ghion, 16 sections from Fasilo and 17 sections from Bahir Dar Zuria. The minimum number of students in a class was 45. The maximum number of students in a class was 49 in Bahir Dar Zuria. A section in Bahir Dar Zuria was taken and two numbers out of 49 were drawn randomly by lottery method. The numbers were 17 and 23. Then, the students with these roll numbers from each section were taken as a sample of the study. A total sample of 120 students, 30 from Tana Haik, 24 from Ghion, 32 from Fasilo and 34 from BahirDar Zuria were taken.

3.3. Instruments for Data Collection

3.3.1. Document Analysis

In assessing content validity of tests, document analysis is a vital instrument. Then, this research was basically on document analysis. As a result, the contents and formats of Grade 9 English language textbook, the contents and the formats of the 2003 E. C. Grade 9 first semester English language sample tests were analyzed.

The time coverage for the units was consulted from the Grade 9 English language syllabus.

3.3.2. Questionnaire

To increase the reliability of the data collected from the document analysis, both close-ended and open-ended questionnaire for Grade 9 English language teachers and close ended questionnaire for Grade 9 students were used.

3.3.2.1 Questionnaire for Teachers

The purpose of this questionnaire was to elicit Grade 9 English language teachers' opinions about representativeness of their classroom tests.

The questionnaire had three parts. The first part dealt with general information, enquiring the number of students in a class, the number of sections that teachers had, the units teachers covered for the first semester and the teachers' awareness about language testing. The second part of the questionnaire had thirteen items which enquired the content and format of classroom tests. The third part had two items enquiring the emphasis teachers gave in teaching and testing.

3.3.2.2. Questionnaire for students

In order to cross check the Grade 9 English teachers' opinions on the representativeness of classroom tests, a questionnaire was administered to Grade 9 students.

The questionnaire had three parts. The first part dealt with background information, enquiring the name of school and the number of students in a section. The second part of the questionnaire had ten items which dealt with the contents and formats of the classroom tests. The third part had two items which dealt with the emphasis teachers gave in teaching and testing.

Question number 1,7,8, from part two and question number 1 and 2 from part three of teachers, and question number 1,7, 8, from part two and question number 1 and 2 from part three of students were taken from Kifle's (1995). They were slightly adapted and used for this research.

3.4. Procedure of Data Collection

3.4.1. Textbook Analysis

The first step was to analyze the contents of the Grade 9 English language textbook. This was done by classifying the contents of the textbook by headings of the major language aspects, such as Reading, Vocabulary Grammar, Listening, Speaking and Writing. The lists of the tasks in each units (from unit one to unit six, including the special unit) under the main language aspects were analyzed (see appendix-1).

The Grade 9 English language teacher's guide recommended that the first seven and half units should be completed by the end of the first semester. As a result the first seven and half units were analyzed. But, from the questionnaire teachers reported that they have completed only the first six units. The first six units including the special unit were analyzed for this study.

The researcher did not want to use directly the frequencies of the analyzed tasks under each language skill for the comparison of the contents of the textbook and the tests. This was because the researcher believed that taking the frequencies of tasks without considering the time spent could be misleading as a less frequent new task might take the same consideration as a common more frequent task. Taking the frequency of practice exercise was also difficult as the exercises in speaking and writing were general.

Heaton (1991 and 1990) recommended the time spent to teach a particular area can be taken as a criterion to sample classroom tests. He further elaborated the idea that an attempt should be made to relate the different areas covered in the test to the length of time spent on teaching those areas in class. However, the syllabus of Grade 9 English language provided only unit coverage. It did not provide the time coverage for different language skills. As the study was about teacher-made classroom tests, teachers were made to participate in deciding the time that each task could take while teaching them.

In favor of the decision which was made, Heaton (1990) also said that informal attainment tests should be based on what has been taught, instead of what should have been taught. The researcher believed it was reasonable to make teachers decide the time for the task that they have taught before. So, the copies of the analyzed tasks (see appendix-1) were given to four

experienced English language teachers who have taught at Grade 9 for a long time. These teachers were one from each sample school and volunteer to help the researcher. These teachers were made to allocate the time that could take to teach the analyzed tasks under the main language aspects of each unit. The time coverage, which was given for each unit in the syllabus of Grade 9 English, was used as a reference to allot time for each task under each unit. This was completed and returned in three weeks time. Inter rater reliability among the raters for the main six content areas was computed using Cronbach alpha (see appendix-2). The inter raters' reliability was found very high (0.9). To get more reliable data, instead of taking the average time, the four teachers were made to meet together including the researcher, for two morning sessions (Saturday and Sunday), and discussed and commented each other's time allocation reasonably. An agreement was reached for all tasks and optimum time was decided.

To get reliable data, teachers were only informed that the time allocation was needed for research purpose. They were not told the specific purpose it was used. Making the analysis before questionnaire administration and test paper collection was also thought to increase the reliability of the data.

The test formats used in the Grade 9 English language textbook were also analyzed. Frequency count for each format was made by the researcher.

3.4.2. Questionnaire Administration

The second step after content analysis was to administer the questionnaire for English language teachers and Grade 9 students. The teachers' questionnaire was distributed by the researcher in person for the respondents and collected back on the next day. All of the questionnaires were returned. The students filled in the questionnaire assisted by the researcher and collected on the same day. From 120 respondents, 118 questionnaires were returned. The reason to administer the questionnaire before test paper collection was to avoid artificial response from the teachers' side.

Before the administration of the questionnaire, pilot survey was made on four Grade 9 English language teachers who rated the time allotment for the different tasks, and thirty randomly selected Grade 9 students from Bahir Dar Zuria Secondary School. The results obtained from the survey showed that some subjects were unfamiliar with some of the terms used. Then, modifications were made for the words and an Amharic version was prepared for students.

3.4.3. Sample Test Papers Collection

After textbook analysis and administration of questionnaire to teachers and students, 2003 E.C first semester English language classroom progress and achievement tests of the four sample schools were collected. Their contents were analyzed. Headings such as, Reading, Vocabulary, Grammar, Listening, Speaking and Writing were used as major content areas for the analysis. Two English language teachers were made to participate in the sample test papers analyses to increase the reliability of the data. Test formats of the sample tests were also analyzed by counting the frequency.

The reason for taking only first semester English language tests was to match the response of the existing teachers with the tests they constructed. The inclusion of previous year test papers was not believed to seem reasonable as the present teachers might not be accountable for those papers. Furthermore, both teachers and students could have fresh memory about the recent tests while giving information.

3.5. Data Presentation and Analysis

The obtained data from different sources were analyzed as follows:

3.5.1. Contents of the Textbook and Sample Test Papers

1. The contents of the textbook and the sample tests were presented together in six different tables (see Table 1-6) under the main content areas (Reading, Vocabulary, Grammar, Speaking, Listening and Writing). The time allotted for the different content areas and the frequencies of the test items were determined in figures and percentages.
2. The summary of the time allotted for the contents of the textbook and summary of the frequency of the test items under the main content areas (Reading, Vocabulary, Grammar, Speaking, Listening and Writing) were presented in a table(see Table 7).

After presenting and analyzing the above data by using figures and percentages, the degree of relationship between the time allotted to the contents of the textbook and the frequencies of the test items that appeared in the sample tests was determined by using Spearman's Rank Order Correlation coefficient Statistical Method.

To compute the Spearman's Rank Order Coefficient of correlation, the following formula was used.

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

where r' = rank of correlation coefficient

D = difference between paired ranks

$\sum D^2$ = the sum of squared difference between ranks

N = number of paired ranks

According to Best and Kahn(1993), the following are interpretation of the correlation coefficient between two observations or ranks

| Coefficient(r') | Relationship |
|-------------------------------------|---------------------|
| 0 to 0.2 | Negligible |
| 0.2 to 0.4 | Low |
| 0.4 to 0.6 | Moderate |
| 0.6 to 0.8 | Substantial |
| 0.8 to 1 | high to very high |

3.5.2. Formats of the Textbook and Sample Test Papers

The different formats of the textbook and sample test papers were presented in a table showing the frequency of the different formats that appeared in the textbook and sample test papers. Then, the relation ship between the formats employed in the textbook and sample test papers were also determined by using the Spearman's rank order correlation coefficient.

3.5.3. Questionnaire of Teachers and Students

The responses from teachers and students were analyzed quantitatively and qualitatively to supplement the data collected from the document analyses. The quantitative data were analyzed in figures and percentages. The qualitative data were expressed in words.

Chapter Four: Data Analysis Results and Discussion

The aim of this study was to assess the content validity of Grade 9 English language classroom tests. To answer the research questions, description of the Grade 9 English language textbook, content analysis and comparison of Grade 9 English language textbook and sample tests, comparison of the test formats of the textbook and the sample tests, and analyses of subjects' responses were made.

4.1. General description of Grade 9 English language Text book

The Grade 9 English language Textbook was published in 1996 and revised in 2005. The text book is divided in to units. It contains fifteen units: one 'special' unit which deals with basic speaking and writing skills, and fourteen topic based units.

As it is stated in the Grade 9 English language Teachers Guide, the topics have been carefully chosen with regard to needs and interests of Grade 9 students. Each of the topic based units is divided in sections as Reading, Vocabulary, Grammar, Speaking, Listening, and Writing. These are the main aspects of English language teaching in the Grade 9 course. The teachers guide also states the following about the grade 9 English language Textbook.

As the course is graded, teachers are advised to work through the material in the students' book systematically, starting with the special unit: Basic writing and speaking skills and continuing from Unit one to Unit fourteen. Similarly, it is also advisable to work through each unit section by section. This is because the work in a given section often follows on logically from the work covered in previous sections.

This means it is difficult to join one section of the text book by jumping the section that precedes it.

The Grade 9 teacher's Guide also recommends that each Unit should take about two to two and half weeks to complete. A single period takes forty minutes. To cover the contents of the Grade 9 English language textbook, the Grade 9 Teachers Guide advises to complete seven and half units of the book by the end of the first semester. However, for this study the contents of the first six units, including the special unit were analyzed as all teachers have reported that they have covered only the first six units by the end of the first semester.

As specified in the Grade 9 English language syllabus, the syllabus is the base for the development of Grade 9 English language text book. The syllabus also states that in each unit, wider opportunities are provided to develop the four language skills and attention is also paid to develop the students' ability in manipulating the language patterns and increasing their vocabulary.

4.2 Content Analysis of Grade 9 English Language Textbook and Sample Test Papers

As it has been mentioned before, reading, vocabulary, grammar, speaking, listening and writing are the main language aspects (content areas) in Grade 9 English language textbook. Then, the analyses were made based on these content areas.

4.2.1. The List of Reading Tasks in the Textbook with the Allotted Time and the Frequencies of Test Items in Each Task

Different reading skills (both macro and micro) are considered in the text book. Reading for detailed information is the dominant task that teachers spend much (62.9%) of their time. On the other hand, the pre- reading questions, which are one of the macro skills, take the least (5.5%) of the total time allotted for reading tasks in the textbook.

In the collected sample test papers reading for detailed information takes the highest (46.1%) share. Guessing meanings of new words, which is believed to encourage students to guess the meanings of difficult words as part of their reading strategy, is given the second (30.7%) place in emphasis from the different tasks.

It is impossible that all the contents of the textbook can not appear in a test. But, sampling should be done reasonably. The tasks which receive focus in the practice session should also get emphasized in the classroom tests. As the emphasized tasks in the time allotment-reading for detail information and deducing meanings of new words-have received attention in the tests, it is possible to say that, the test items in reading are relatively according to the time allotted to them in the textbook. This, then, can help to achieve content validity in this content area. The list of reading tasks in the textbook with the allotted time and the frequencies of test items in each task is presented in Table 1.

Table 1. The List of Reading Tasks in the Textbook with the Allotted

Time and the Frequencies of Test Items in Each Task

| No. | List of reading tasks | Tasks in the textbook | | Items in Tests | |
|-----|----------------------------------|----------------------------|-----------------------|----------------|-------------|
| | | Time allotted In Minute | Percentage of Time | Frequency | Percentage |
| 1 | Pre-reading questions | 30 | 5.5 | - | - |
| 2 | Reading for specific information | 50 | 9.2 | 4 | 15.3 |
| 3 | Deducing meanings of new words | 120 | 22.2 | 8 | 30.7 |
| 4 | Reading for detailed information | 340 | 62.9 | 12 | 46.1 |
| 5 | Reading for gist | - | - | 2 | 7.6 |
| | Total | 540(13.5 Periods) | 99.8 | 26 | 99.7 |

4.2.2. The List of Vocabulary Tasks in the Textbook with the Allotted Time and the Frequencies of Test Items in Each Task

The highest (35.7%) time is given to guessing the meanings of new words related to the reading text. The verb to take and to last, the meanings of make and using feel + adjective are the least (4.7% each) emphasized tasks in time allotment. The other vocabulary tasks which are given separate sections in the text book are fairly given similar emphasis.

In the collected test papers, guessing the meanings of words using contextual clues is also given prominence (70.2%). 'To feel' + adjective is given the least(2.1%) emphasis.

It is good practice that guessing the meanings of new words receives the first emphasis in tests as it gets the highest time allocation in the text book. However, synonyms and prefixes, which are the second and third emphasized tasks in the time allotment respectively, do not appear in the sample test papers. Such practices may affect the adequacy of sampling in the vocabulary content area. The list of vocabulary tasks in the textbook with the allotted time and the frequencies of test items in each task is presented in Table 2.

Table 2. The List of Vocabulary Tasks in the Textbook with the Allotted Time and the Frequencies of Test Items in Each Task

| NO | List of vocabulary tasks | Tasks in the textbook | | Item in Tests | |
|----|--|----------------------------|-----------------------|---------------|-------------|
| | | Time allotted In minute | Percentage Of time | Frequency | Percentage |
| 1 | Guessing meanings of words related to the text | 150 | 35.7 | 33 | 70.2 |
| 2 | Verbs 'to take' and 'to last' | 20 | 4.7 | - | - |
| 3 | The meanings of 'Make' | 20 | 4.7 | 3 | 6.3 |
| 4 | Synonyms | 50 | 11.9 | - | - |
| 5 | Prefixes | 40 | 9.5 | - | - |
| 6 | The meanings of 'Change' | 40 | 9.5 | 3 | 6.3 |
| 7 | Word family | 40 | 9.5 | 4 | 8.5 |
| 8 | Using 'Feel' + Adjective | 20 | 4.7 | 1 | 2.1 |
| 9 | 'Steal' and 'Rob' | 40 | 9.5 | 3 | 6.3 |
| | Total | 420(10.5Periods) | 99.7 | 47 | 99.7 |

4.2.3 The List of Grammar Tasks in the Textbook with the Allotted Time and the Frequencies of Test Items in each Task

From the different kinds of grammar tasks in the English language textbook, the passive and reported speech are given equal focus and relatively receive the highest(17.3% each) time allotment. Conditional sentence is the second (15.2%) emphasized area. On the other hand, verbs followed by-'ing', ways of expressing purpose, used to, drawing conclusions using 'must be', and 'for' and 'since' are the least(4.3% each) emphasized tasks.

Conditional sentences, the passive and verbs followed by- ing are the first three emphasized test items from the sample tests with the time allotment of 16.4%, 12% and 12% respectively. Drawing conclusions using 'must be', and 'for and since' are the least emphasized items in the sample test papers with 2.1% coverage each.

Some of the most emphasized and the least emphasized items in the grammar tests are based on the emphasis which these tasks are given in the time allotment. It is good that these tasks appear in the sample tests in accordance with the emphasis the tasks have been treated in the textbook. But, reported speech which is given the first emphasis in the time allotment is not considered

well in the tests, and verbs followed by -ing, which is not given emphasis in the time allotment, is given prominence in the sample tests. The list of grammar tasks in the textbook with the allotted time and the frequencies of test items in each task is presented in Table 3.

Table 3. The List of Grammar Tasks in the Textbook with the Allotted Time and the Frequencies of Test Items in each Task

| No | Lists of grammar tasks | Tasks in textbooks | | Items in Tests | |
|----|-------------------------------------|------------------------|-----------------------|----------------|-------------|
| | | Time In minutes | Percentage of time | Frequency | Percentage |
| 1 | Verbs followed by infinitive | 80 | 8.6 | 9 | 9.8 |
| 2 | Verbs followed by '-ing' | 40 | 4.3 | 11 | 12 |
| 3 | Ways of expressing purpose | 40 | 4.3 | 6 | 6.5 |
| 4 | Conditional sentences | 140 | 15.2 | 15 | 16.4 |
| 5 | The passive | 160 | 17.3 | 11 | 12 |
| 6 | Used to | 40 | 4.3 | 5 | 5.4 |
| 7 | Reason clause | 80 | 8.6 | 5 | 5.4 |
| 8 | Using so-that to express result | 100 | 10.8 | 5 | 5.4 |
| 9 | Reported speech | 160 | 17.3 | 3 | 3.2 |
| 10 | Drawing conclusions using 'Must be' | 40 | 4.3 | 2 | 2.1 |
| 11 | 'For' and 'Since' | 40 | 4.3 | 2 | 2.1 |
| 12 | Coordinate conjunction | - | - | 7 | 7.6 |
| 13 | Possessive adjectives and pronouns | - | - | 7 | 7.6 |
| 14 | Simple present tense | - | - | 3 | 3.2 |
| | Total | 920(23 periods) | 99.3 | 91 | 98.7 |

4.2.4. The List of Speaking Tasks in the Textbook with the Allotted Time and the Frequencies of Test Items in Each task

The textbook specifies speaking tasks which students should be able to understand and express appropriately in a variety of everyday social situations. It also specifies certain academic functions which the students need to master in order to be successful in a learning situation where English is used as the medium of instruction. Relatively, asking for some one's opinion takes the highest (23%) time of the speaking tasks. Obtaining information and introducing one self are the least (3.8% each) emphasized tasks.

In the sample test papers, asking for repetition (25%) and asking to borrow something (25%) are emphasized areas as they take together half (50%) of the total share from the speaking test items.

Asking for someone's opinion (6.2%) is the least emphasized task from the tasks that appear in the sample tests.

Introducing one self, which has been given little emphasis in the textbook, is considered well in the sample tests. On the other hand, asking for some ones opinion which is the first dominant task in time allotment is not given prominence in the tests. This may influence the adequacy of sampling in this content area. The list of speaking tasks in the textbook with the allotted time and the frequencies of test items in each task is presented in table 4.

Table 4. The List of Speaking Tasks in the Textbook with the Allotted Time and the Frequencies of Test Items in Each task

| No | List of speaking tasks | Tasks in the Textbook | | Items in tests | |
|----|---|-----------------------------|-----------------------|----------------|-------------|
| | | Time allotted In minutes | Percentage Of time | Frequency | Percentage |
| 1 | Obtaining information | 20 | 3.8 | - | - |
| 2 | Introducing one self | 20 | 3.8 | 2 | 12.5 |
| 3 | Giving speech | 40 | 7.6 | - | - |
| 4 | Pair and group discussion | 80 | 15.3 | - | - |
| 5 | Asking for repetition | 80 | 15.3 | 4 | 25 |
| 6 | Asking to borrow something | 80 | 15.3 | 4 | 25 |
| 7 | Asking for, giving and refusing information | 80 | 15.3 | 2 | 12.5 |
| 8 | Asking for someone's opinion | 120 | 23 | 1 | 6.2 |
| 9 | Asking for permission | - | - | 3 | 18.7 |
| | Total | 520(13 periods) | 99.4 | 16 | 99.9 |

4.2.5. The List of Listening Tasks in the Textbook with the Allotted Time and the Frequencies of Test Items in each Task

From the two different kinds of tasks practiced in the listening section of the Grade 9 English language textbook, listening for specific information is given much (75%) of the time. It seems that the activities are aimed to train students to listen efficiently to a variety of extended English speech. The other listening task in the textbook is dictation with time coverage of 25%.

In the sample test papers listening is not considered at all. The list of listening tasks in the textbook with the allotted time and the frequency of test items in each task are presented in Table 5.

Table 5. The List of Listening Tasks in the Textbook with the Allotted

Time and the Frequencies of Test Items in each Task

| No | List of listening tasks in the textbook | Time allotted in minutes | Percentage of time |
|----|---|--------------------------|--------------------|
| 1 | Listening for specific information | 240 | 75 |
| 2 | Dictation | 80 | 25 |
| | Total | 320(8 periods) | 100 |

4.2.6. The List of Writing Tasks in the Textbook with the Allotted Time and the Frequencies of Test Items in Each Task

Different writing tasks are included in the textbook. The inclusion of the Variety of tasks in the textbook is aimed to develop the students writing skill. As writing in English is so important for the students academic success in all subjects from Grade 9 on wards, these writing activities will help to master the basic writing skills. From 14 periods given for writing, Reordering sentences (21.4%) and Writing guided composition (19.6%) are given emphasis. Using pronouns and possessive adjectives to avoid repetition is the least (3.5%) emphasized task in time coverage.

From the different witting tasks which have been practiced in the textbook, only three tasks appear in the sample tests. The simple sentences (45.3%) and punctuation (36.3%) are the first and the second emphasized tasks in the sample tests respectively.

The above mentioned tasks which get emphasis in the sample tests are not emphasized in the time allotment as they are given in tests. On the other hand, reordering sentences and guided composition writing which have been given emphasis in the time allotment are not considered at all in the sample tests. The list of writing tasks in the textbook with the allotted time and the frequencies of test items in each task is presented in Table 6.

Table 6. The List of Writing Tasks in the Textbook with the Allotted Time and the Frequencies of Test Items in Each Task

| No | List of writing tasks | Tasks in the Textbook | | Items in Tests | |
|----|--|--------------------------|--------------------|----------------|-------------|
| | | Time allotted in minutes | Percentage of time | Frequency | Percentage |
| 1 | The simple sentences | 40 | 7 | 5 | 45.3 |
| 2 | Use pronouns and possessive adjectives to avoid repetition | 20 | 3.5 | - | - |
| 3 | Paragraphs | 40 | 7.1 | 2 | 18 |
| 4 | Reordering sentences | 120 | 21.4 | - | - |
| 5 | Guided composition writing | 110 | 19.6 | - | - |
| 6 | Note making | 50 | 8.9 | - | - |
| 7 | Writing instructions | 80 | 14.2 | - | - |
| 8 | Completing summary | 40 | 7.1 | - | - |
| 9 | Punctuation | 60 | 10.7 | 4 | 36.3 |
| | Total | 560(14 periods) | 99.5 | 11 | 99.5 |

4.2.7. A Comparison between the Summary of the Time Coverage Given for the Main Content Areas and the Summary of Frequencies of Tasks in the Sample Test Paper

In the Grade 9 English language textbook, grammar and writing are the first and second emphasized areas with 28% and 17% time allotment respectively. Listening is given the least emphasis with time coverage of 9.7%.

In the frequency of sample test papers, grammar (47.6%) and vocabulary (24.6%) are the first and second emphasized areas respectively. Listening is not totally considered in the sample test papers.

Grammar is the first emphasized content area both in time allotted to it in the textbook and the frequency of test items that appeared in the sample tests. Content area, like listening, which is given reasonable time coverage in the textbook, is not considered in the sample tests. However, this content area should have been considered in the sample tests as the emphasis it receives in the time allotment. A comparison between the summary of the time coverage given for the main

content areas and the summary of frequency of tasks in the sample test papers is presented in table 7.

**Table 7. A Comparison between the Summary of the Time Coverage Given
for the Main Content Areas and the Summary of Frequencies of Tasks
in the Sample Test Papers**

| No | Main content Areas | Time allotted in Periods | Percentage | Frequency of tasks in the Sample tests | Percentage |
|----|--------------------|--------------------------|-------------|--|-------------|
| 1 | Reading | 13.5 Periods | 16.4 | 26 | 13.6 |
| 2 | Vocabulary | 10.5 periods | 12.8 | 47 | 24.6 |
| 3 | Grammar | 23 periods | 28 | 91 | 47.6 |
| 4 | Speaking | 13 periods | 15.8 | 16 | 8.3 |
| 5 | Listening | 8 periods | 9.7 | 0 | 0 |
| 6 | Writing | 14 periods | 17 | 11 | 5.7 |
| | Total | 82 periods | 99.7 | 191 | 99.8 |

4.2.8. Ranking the Six Main Content Areas in their Time Allotment and Frequencies of Test Items

In order to determine the degree of relationship that exists between the time allotted for the different tasks of the textbook and the frequency of test items that appear in the sample test papers, the Spearman's Rank Order Correlation Coefficient statistical method is employed.

Table 8. Ranking the Six Main Content Areas in their Time Allotment and Frequencies of Test Items

| Main content areas | Ranking based on time allotted | Ranking based on the frequency Of tasks in the sample test paper | D | D ² |
|--------------------|--------------------------------|--|---|----------------|
| Reading | 3 | 3 | 0 | 0 |
| Vocabulary | 5 | 2 | 3 | 9 |
| Grammar | 1 | 1 | 0 | 0 |
| Speaking | 4 | 4 | 0 | 0 |
| Listening | 6 | 6 | 0 | 0 |
| Writing | 2 | 5 | 3 | 9 |

As the ranks of the different content areas is observed in Table 8, the extent of relation ship between the time allotment and frequency of test items can be determined using Spearman's Rank Order Correlation Coefficient as follows:

$$r^s=1-6\sum D^2/N(N^2-1)$$

$$r=1-6(18)/6(6 \times 6-1)$$

$$r=1-108/6(35)$$

$$r=210-108/210$$

$$r=102/210$$

$$=0.48$$

The rank correlation coefficient is 0.48.

Where r^s = rank of correlation coefficient

$\sum D^2$ = the sum of squared difference between ranks

D = difference between ranks of corresponding pairs

N = number of observations

According to Best and Kahn (1993:308) the table value from 0.4 up to 0.6 of Spearman's Rank Order Correlation Coefficient is considered as moderate. Then, the correlation shows that there is moderate correlation between the time allotted for the tasks of the textbook and the frequency of tasks that appeared in the sample test papers. As validity is not a matter of all or none, this moderate validity in terms of coverage informs one that teachers gave share for some components of the test items according to the focus they gave while teaching (presenting lessons). This was evidenced by content areas like grammar. This content area received more attention while teachers taught the classroom lessons and this was also reflected in the number of items that appeared in the tests.

Though the rank correlation coefficient was positive, one could observe disparities among the different skills and subs-skills treated in the test. For example, listening, which was given a

reasonable coverage in the textbook, was completely ignored in the sample tests. The absence of listening skill might affect the content validity of the test. This test could not be complete to be used as evidence of the language ability for a certain class.

Furthermore, there were differences between the expected and observed test item frequencies of some content areas. For example, writing which was given the second rank in the time allotment was given the fifth rank in the frequency of tasks that appeared in the test. On the other hand, vocabulary which was given the fifth rank in the practice session was the second emphasized area in the sample tests. As writing was the content area which was given more time allotment than vocabulary, more test items in writing than vocabulary should have appeared in the sample tests. However, this was not reflected in the sample tests. This can be attributed as, Hughes (1989) said, teachers focus on what is easy to test instead of what is important to test. Then, such practice of teachers can affect the content validity of tests in terms of coverage. This, in turn, could make the test resulted in harmful wash back effect.

It is obvious that all the content areas which were in the textbook could not appear in the tests, but it was reasonable to give priority according to the emphasis which the task received during instruction. However, the tasks like reordering sentences, guided composition writing and writing instruction, which are important tasks for writing skill and have been given due consideration in the text book, were not considered at all in the classroom tests. On the other hand, tasks like the simple present tense, which was not treated in the Grade 9 English language textbook, appeared in the sample tests. In this regard, Henning (1987) warned us that inappropriate selection of content is one of the causes for the threat to content validity. To minimize this threat, Weir (1993:5) advises us, "classroom tests should not be divorced from the teaching that precedes it".

Disparity in the classroom tests did not occur only across the different content areas, there was also improper allocation among the tasks with in the content areas. It is important to decide that the weight given for each task in each content area should be according to the emphasis given in the instruction for that specific task. However, this was not implemented in the collected test papers. For example, the tasks under grammar section like, reported speech was the first emphasized task in the textbook; however, this task was not given reasonable coverage in the tests. Only 3.2 % coverage was given while other tasks which had the same time allotment in the

practice session, like the passive received around 16% of coverage. In the above, it has been noted that grammar received emphasis both in the number of items and to the emphasis given in the teaching learning process. However, this in itself did not ensure content validity. The test would have content validity only if it included a proper sample of the relevant structure. Then, this improper allocation again might affect the content validity of classroom tests as “a test in which major areas identified in the specification are under-represented- or not represented at all- is unlikely to be accurate”(Hughes,1989:22).

It has been mentioned that the fundamental task of testing and measurement in education is to provide information to make decision (Popham, 2003) . Teachers are expected to make use of test results to improve their classroom instruction and to have a better understanding of their students test performance. For example, in this research context important content area of the text book, like listening, was not presented in the test. Then, this could lower the content validity of the test in content wise. As a result, the decisions that could be made based on this test could be misleading. In this regard Grondlund (1982:126) notes “ we are interested in content validity when we want to use test performance as evidence of performance in a large domain of situations”. This means, if one wants to talk about the English language ability of grade 9 students based on their first semester result, the information could not be complete as the test did not assess the students listening ability. Furthermore, Bachman (1990:62) notes, “the performance of students on achievement tests can provide an indication of the extent to which the expected objectives of the program are being attained, and thus pinpoint areas of deficiency”. As already mentioned, the sample tests did not represent all content areas, thus, students performance on sample tests do not determine the attainment of the expected goals.

Furthermore, a test that lacks content validity in turn can make tests result in harmful wash back effect as areas which are ignored in testing are likely to be ignored in the teaching learning process (Hughes, 2008). This means, in this research context, again content areas like listening and writing were less likely to be treated well in the teaching and learning process.

As it has been reported above, grammar received big share in the time allotment, which means teachers gave more time to teach grammar than any other content area in the textbook. As a result, this content area received big share in the collected test papers. It was a good practice for

classroom tests as these tests should be based on what has been taught. This could help to achieve content validity in terms of content coverage.

It should not be forgotten that the Grade 9 English language syllabus and teachers guide recommends communicative approach to be used in the teaching learning process. However, when one observes the test items, they are all indirect selective types. Then, this mismatch between what is expected as a final goal in the language learning and the testing system affects students learning as well as the general content validity of the test. In this regard, Heaton (1991:13) warns test makers, “in every case, it is important that a test reflects the actual teaching and the course being followed”.

Then, the indirect tests, which were used in the classroom tests to measure the productive skills of students, were not appropriate testing approaches for communicative language learning. This, in turn, lowers the content validity of that classroom test. For classroom tests, to achieve content validity, direct tests should be used (Brown, 2004).

4.3. Analyses of the Formats of Grade 9 English Language Textbook and the Sample Test Papers

Different kinds of test formats are used both in the Grade 9 English language textbook and sample test papers.

4.3.1 Frequencies of Test Formats in the Textbook and in the Sample Tests

The test formats, write (22.2%), oral practice (18.5%) and change (14.8%) are the first, the Second and the third frequent used test formats used in the Grade 9 English language textbook respectively. On the other hand, reordering (0.09%), punctuate (1.8%) and true false (2.7%) are the least common test formats in the English language Grade 9 textbook. The text book also employs a variety of subjective and objective formats.

Multiple choose (83.2%) is the most dominantly used test format in the sample test papers. Fill-in (7.8%) is the second emphasized test format. The other five test formats used in the tests constitute only 10% of the whole formats used in the tests. Table 10 presents the frequencies of the test formats in the sample tests. Table 9 presents the frequencies of test formats in the textbook and sample tests.

Table 9. Frequencies of Test Formats in the Textbook and Sample Tests

| No | List of Test format | In the Textbook | | In The Sample Tests | |
|----|---------------------|-----------------|------------|---------------------|------------|
| | | Frequency | Percentage | Frequency | Percentage |
| 1 | Oral practice | 20 | 18.5 | - | - |
| 2 | Write | 24 | 22.2 | - | - |
| 3 | Multiple Choose | 4 | 3.7 | 159 | 83.2 |
| 4 | True/false | 3 | 2.7 | 3 | 1.5 |
| 5 | Short answer | 15 | 13.8 | 1 | 0.06 |
| 6 | Fill-in | 9 | 8.3 | 15 | 7.8 |
| 7 | Matching | 9 | 8.3 | 5 | 2.6 |
| 8 | Defining | 5 | 4.6 | - | - |
| 9 | Reordering | 1 | 0.09 | - | - |
| 10 | Punctuate | 2 | 1.8 | 2 | 1 |
| 11 | Change | 16 | 14.8 | 6 | 3.1 |
| | Total | 108 | 99 | 191 | 99.2 |

4.3.2. Comparison of the Test Formats in the Textbook and Sample Test Papers

The format write (22.2%) and oral practice (18.5%) are dominantly used test format in the textbook, but in the case of sample test papers Multiple choose format (83.2%) is employed most dominantly. Formats which are used dominantly in the text book, like writing is not completely included in the sample tests. Table 10 presents the comparison of the test formats in the textbook and sample test papers.

Table 10. Comparison of the Test Formats in the Textbook and Sample Test Papers

| No | Test Formats | | | | | | | |
|----|------------------|-----------|------------|------|--------------------|-----------|------------|------|
| | In the text book | | | | In the test papers | | | |
| | | Frequency | Percentage | Rank | | Frequency | Percentage | Rank |
| 1 | Oral practice | 20 | 18.5 | 2 | Oral practice | 0 | 0 | 8 |
| 2 | Write | 24 | 22.2 | 1 | Write | 0 | 0 | 8 |
| 3 | Choose | 4 | 3.7 | 8 | Choose | 159 | 83.2 | 1 |
| 4 | True/false | 3 | 2.7 | 9 | True/false | 3 | 1.5 | 5 |
| 5 | Short answer | 15 | 13.8 | 4 | Short answer | 1 | 0.06 | 7 |
| 6 | Fill-in | 9 | 8.3 | 5 | Fill-in | 15 | 7.8 | 2 |
| 7 | Matching | 9 | 8.3 | 5 | Matching | 5 | 2.6 | 4 |
| 8 | Defining | 5 | 4.6 | 7 | Defining | 0 | 0 | 8 |
| 9 | Reordering | 1 | 0.09 | 11 | Reordering | 0 | 0 | 8 |
| 10 | Punctuate | 2 | 1.8 | 10 | Punctuate | 2 | 1 | 6 |
| 11 | Change | 16 | 14.8 | 3 | Change | 6 | 3.1 | 3 |
| | Total | 108 | 99 | - | | 191 | 99.2 | - |

The rank '0' in the sample test papers refers to the test formats that do not appear in the tests

Then, in order to determine the degree of relationship between the formats used in the textbook and sample test papers the Spearman's Rank Order Correlation Coefficient was employed.

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

$$r_s = \frac{1 - 6(195)}{11(11 \times 11 - 1)}$$

$$= \frac{1 - 1170}{11(121 - 1)}$$

$$= \frac{1 - 1170}{1320}$$

$$= \frac{1320 - 1170}{1320}$$

$$= \frac{150}{1320}$$

$$= 0.1$$

Best and Kahn (1993:308) has said the rank correlation coefficient from .00 to 0.2 shows negligible relationship between the paired ranks. Then, the rank correlation coefficient (0.1) tells that there was no almost any relationship between the formats used in the textbook and the test papers. This means, the rank order correlation coefficient shows significant difference between the formats used in the textbook and sample test papers. The formats which were given emphasis in the text book were not considered at all in the sample test papers. This included the test formats, like writing and oral practice. Further more, test format like short answer, which was given emphasis in the textbook, was given very little emphasis in the sample tests.

As it has been indicated in Table 10, the textbook employed both objective and subjective formats. This means teachers used both kinds of formats in the classroom teaching. When one observes the formats which were employed in the sample test papers, they were only objective formats. The tests were found to be weak in reflecting the emphases which were given in the textbook's format. However, achievement and progress tests are expected to use test formats which have been practiced before. Supplementing this point Atkins, et al. (1996) recommended that progress and achievement tests should be in the way it has been taught. This means, the different test formats in the text book should be reflected in a reasonable proportion in sample test papers. Moreover, as Alderson, et al. (1995), "it is likely that particular test methods will lend themselves to testing some abilities, and not be so good at testing others". In this

connection, Heaton (1990) also encouraged the use of subjective formats for testing language skills and certain area of language. This means, as the dominant test format which was used in the sample test papers was multiple-choice, it might not be appropriate to test some items which were found in the collected test papers. If students are given selective items too frequently in classroom tests, they will not learn how to use language. In relation with this, Heaton (1990:33) advocates the following, “always remember that most good tests contain both objective and subjective types of items”. That is why scholars in the field of testing, like Alderson, et al. (1995) and Weir (1990) assures to use a variety of test formats, even to test a single skill.

On the other hand scholars, like Brown (2004) has theorized that direct testing should be used for performance skills for a test to achieve content validity. As these classroom tests dominantly used selective type test items, it seemed impossible to test skills like speaking and writing directly. Then, if these skills were not tested directly, it could be difficult to achieve content validity in those areas. This, in turn, might lower the content validity of classroom tests.

4.4. Responses of Subjects

4.4.1. Responses of Teachers

The questionnaire for teachers had three parts (see appendix-5). The first part dealt with the background information such as, the number of students in the class, the number of sections, the area they covered for the first semester and teachers’ awareness about language testing. The second part of the questionnaire dealt with the content area of Grade 9 English language tests. The third part dealt with teachers focus on teaching and testing on classroom tests.

4.4.1.1. Teachers’ Responses to Limitation, Coverage and Proportion of Content

Regarding limitation of content (item 1), the majority of the respondents, 18 reported that they often limited their tests to the material covered in the classroom. This was also evidenced by the collected test papers. Except two items in the grammar part, all the items were from the contents of Grade 9 English language textbook that have been covered in the first semester. On the other hand, Only 2 of the respondents reported that they sometimes limited their test to the material

covered in the classroom. These two teachers responded that a language test should not be only from the materials covered in the classroom.

As classroom progress and achievement tests are based on what has been taught, the majority of teachers have done this. Though limiting the content area to the material covered in the class is not the only criteria to achieve content validity, what teachers have done is a step to achieve the goal. This practice could contribute for a test to achieve content validity in terms of content coverage.

Regarding adequacy of coverage (item2-7), 18(grammar), 15(vocabulary), 13(reading), 8(speaking) and 12(writing) of the 20 respondents reported that their tests often had adequate coverage of the content areas in the bracket. This data indicates that the majority of teachers believed that their tests had adequate coverage of the above mentioned skills. However, this response of teachers was not reflected in the data on the collected test papers. On the other hand, all of the 20 respondents reported that their tests never had adequate coverage of listening skill. This was completely evidenced by the data from the collected test papers. Teachers also reported that it was difficult to include listening tests as testing materials were not available.

Regarding the proportion that teachers gave to the different content areas in their tests (item-8), 10 of the respondents expressed they often sampled their tests as they were treated in the classroom or textbook. Other 8 of the respondents reported that they sometimes sampled tests according to the emphasis given in the text book. These teachers reported that all the content areas were not equally easy to test. They expressed that if the content area was comfortable for testing, they focused on that area to have test items.

On the other hand, 19 of the 20 teachers reported (item-11) that they teach (present) all the major and sub language skills in their class. Then, the content areas which have been treated in the classroom should have appeared in the tests.

Responding to the criteria to decide on item content (Item-12), the Majority of the teachers reported that their guide line to decide the number of items on the different content areas depended on the time spent to teach the content area. According to the collected sample test items, the responses of teachers worked for only grammar items. If teachers really focused on the time spent to teach while deciding the number of items, their tests could have achieved high

content validity. However, the theoretical explanation of these teachers did not fit with their practical consideration. This was because teachers did not consider this justification in their classroom tests as content areas which were given little time allotment have got emphasis in tests, vocabulary is a case in point. On the other hand, some of the respondents reported that they depended on the easiness to prepare objective tests. The response of these teachers went in line with what has appeared in the actual classroom sample tests. For example, teachers had many items from vocabulary than writing as vocabulary items might be easy for them to test.

From the above points, it can be said that teachers did not have strict rule and procedure in constructing tests. The contents of classroom tests were not only decided on the base of emphasis which teachers spent while teaching that content but also on the suitability of the content area in testing. In this regard, Hughes (1989) said most of the time the content of the test is determined by what is easy to test rather than what is important to test. However, when teachers focus on what is easy to test by ignoring the important part of the text book, their practice may affect content validity. Then, this in turn makes tests to have harmful wash back effect. Table 11 presents teachers response to limitation, coverage and proportion of item content.

Table 11. Teachers' Responses to Limitation, coverage and Proportion of Content

| Statement | Scales | | | | | |
|--|--------|---------|-----------|---------|-------|---------|
| | Often | | Sometimes | | Never | |
| | No | Percent | No | Percent | No | Percent |
| 1. Do you limit your test to the material covered in the class room? | 18 | 90 | 2 | 10 | - | - |
| 2. Do your tests have adequate coverage of grammar items? | 18 | 90 | 2 | 10 | - | - |
| 3. Do your tests have adequate coverage of vocabulary items? | 15 | 75 | 5 | 25 | - | - |
| 4. Do your tests have adequate coverage of reading items? | 13 | 65 | 5 | 25 | 2 | 10 |
| 5. Do your tests have adequate coverage of listening items? | - | - | - | - | 20 | 100 |
| 6. Do your tests have adequate coverage of speaking items? | 8 | 40 | 6 | 30 | 6 | 30 |
| 7. Do your tests have adequate coverage of writing items? | 12 | 60 | 6 | 30 | 2 | 10 |
| 8. Do your tests maintain proportional coverage? | 10 | 50 | 8 | 40 | 2 | 20 |

4.4.1.2. Teachers' Responses to Variety and Proportion of Test Formats

Concerning the variety of formats (item 9), ten of the respondents expressed that they often employed both subjective and objective formats. 7 of the respondents reported that they sometimes employed both subjective and objective formats. These group of teachers reported in the open ended part that subjective formats were difficult to mark and were not employed very much.

When one observes the sample test papers, all the items were in objective format. However, the content analysis of grade 9 English language text book showed that both objective and subjective formats were employed. As it has been explained before, this disparity of test format between the test formats which was practiced in the textbook and the test formats which was used in the sample test papers could have influence on the students' scores as well as on the content validity of the test.

Regarding the proportion of the test formats used in the tests (item-10), 9 of the respondents reported that they often employed test formats in their tests in proportion with the formats which were used in the concerned textbook. On the other hand, 6 and 5 respondents reported respectively they sometimes and never employed formats in their tests in proportion with the formats practiced in the classroom.

Having reported the above information, teachers dominantly used multiple choice-format in their collected tests. However, this format was not the dominantly employed format in the concerned textbook. The negligible correlation ship between the test formats used in the test items and the material concerned was evidence that the test formats in the sample tests were not in proportion with the test formats employed in the textbook.

Reporting their views concerning the criteria to choose a certain format (item-13) the majority of teachers reported that the test formats which were appropriate to achieve the objectives of the test item were selected. But, when one examines the test-formats used in the sample test papers, they were dominantly multiple choices. Teachers employed multiple choice formats to assess the speaking and writing skills of students. However, different testing experts agree that selective item formats cannot be used to achieve the objective of the writing and speaking tasks. On the

other hand, some of the teachers reported that convenience for marking was their sole criteria to choose a certain sample.

From the above points it is possible to conclude that teachers seemed to have awareness that a variety of formats should be used in their test. However, they did not implement in the way they have reported in the questionnaire. From this again it is possible to say that teachers either lacked practical knowledge or they lacked concern on implementing the right format. In line with this, Popham(2003) advised language teachers that they should accept not only the idea that testing can help teaching, but also to act on that idea. Table 12 presents teachers responses to variety and proportion of test format.

Table 12. Teachers’ Responses to Variety and Proportion of Test Format

| Statement | Scales | | | | | |
|---|--------|---------|-----------|---------|-------|---------|
| | Often | | Sometimes | | Never | |
| | No | Percent | No | percent | No | Percent |
| 9.Do your test employ both objective and subjective format? | 10 | 50 | 7 | 35 | 3 | 15 |
| 10. Do your tests employ test formats proportionally? | 9 | 45 | 6 | 30 | 5 | 25 |

4.4.1.3. Teachers’ Responses to the Emphasis they Give while Teaching and Testing

Concerning the emphasis in teaching (item-1 of part three), six, four and another four teachers reported that they ranked first grammar, reading and writing respectively. These emphases were also reflected relatively in the time allotted to the different language areas in the textbook. According to the time allotment, grammar, writing, and reading were the first, the second and third emphasized areas respectively. However, teachers did not decide the number of items in the classroom tests according to the emphasis they gave while teaching. As it has been analyzed in the collected test papers, vocabulary was the second emphasized content area, but it was not among the emphasized content areas in time allotment. Then, this disproportion allocation of the content areas might contribute to lowering the content validity of the classroom tests.

As regards the emphasis teachers gave in testing (item-2 of part three), seven, four and another four teachers reported that they gave emphasis in testing to grammar, reading and vocabulary respectively. In the case of grammar and reading, these content areas also received the first and

the third rank respectively in the collected test papers. Vocabulary was given less emphasis while teaching, but teachers reported that it was given prominence in testing. This was also evidenced by the collected test papers. However, the content areas like writing which was given emphasis in the textbook was not considered as it should be. Such content areas could be ignored by students in the teaching learning process, or students might not be motivated to learn these areas as they did not appear in tests.

In reasoning out their focus for the emphasized areas, many of the teachers reported that the content area which has been focused in the textbook could be given prominence. As to them, grammar was the dominant content area treated in the text and received much focus. In the case of the focus for testing, majority of teachers reported that the area which received emphasis in the teaching learning should also receive focus. Some of the teachers also reported that as their students could not write in English well, they did not want to include writing tests.

From the above points, it is possible to deduce that English language teachers seemed to have simple theoretical ground that class room tests should be based on what has been taught before. This was evidenced in the case of grammar. It is a step for classroom tests to be content valid. But, they have reported the reason for their improper coverage of writing items were students' inability to write in English. Scholars in the field of testing, however; recommended that test makers should base on testing the skill that they want to foster. Further more, there seemed a problem behind teachers practice. The first problem was that they had not implemented what they responded in the questionnaire-vocabulary was not emphasized in the teaching learning process. The other problem was that the teachers focus on teaching and testing grammar put under question the communicative nature of the textbook. One can deduce from this that focus on testing grammar might be because of its easiness to test. Table 13 presents teachers response to the emphasis they give while teaching and testing.

Table 13. Teachers' Responses to the Emphasis they Give while Teaching and Testing

| N o | Reading | | | | | | Vocabulary | | | | | | Grammar | | | | | | Speaking | | | | | | Listening | | | | | | Writing | | | | | |
|--------|---------|---|---|---|---|---|------------|---|---|---|---|---|---------|---|---|---|---|---|----------|---|---|---|---|---|-----------|---|---|---|---|---|---------|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 1 | 2 | 3 | 4 | 5 | 6 | 1 | 2 | 3 | 4 | 5 | 6 | 1 | 2 | 3 | 4 | 5 | 6 | 1 | 2 | 3 | 4 | 5 | 6 | 1 | 2 | 3 | 4 | 5 | 6 |
| 1 | 4 | 5 | 4 | 2 | 4 | 2 | 3 | 4 | 5 | 3 | 3 | 2 | 6 | 4 | 4 | 2 | 2 | 2 | 3 | 3 | 2 | 5 | 3 | 3 | - | 1 | 2 | 4 | 5 | 8 | 4 | 3 | 3 | 4 | 4 | 2 |
| 2 | 4 | 3 | 5 | 3 | 3 | 2 | 4 | 4 | 3 | 4 | 2 | 3 | 7 | 4 | 3 | 2 | 2 | 2 | 2 | 3 | 4 | 4 | 4 | 3 | - | 3 | 2 | 5 | 4 | 6 | 3 | 3 | 3 | 2 | 5 | 4 |

Number 1&2down refers to the question No.

Number1-6 across refers to the rank teachers give to different content areas

4.4.2. Students' Responses

The questionnaire for students had three parts (see appendix-5). The first part dealt with background information such as the name of the school and the no of students in a section the second part of the questionnaire dealt with the content and format of grade 9 English language tests. The third part dealt with the teachers focus on classroom teaching and testing.

4.4.2.1. Students' Responses to the Match, Coverage and Proportion of Item content

Concerning the match between the contents of the test and the material concerned (item-1), from the total of 118 repondents, 90(76.2%) of the respondents (students) reported that their tests often matched with the material concerned. This information was supported by the content analysis of the collected test papers and teachers response from the questionnaire. This was a good trend for classroom tests to base on the material covered in the classroom as it is a step to contribute for the content validity of the test.

Regarding coverage of tests (item2-7), the majority 90(76.2%), 79(66.9%), 72(61%) of the respondents reported that their classroom tests often had adequate coverage of grammar, vocabulary and reading respectively. On the other hand, 60 (50.8%) and 50 (42.3%) of the respondents reported that their classroom tests sometimes had adequate coverage of speaking and writing respectively. The other dominant number, 78(66.1%) of respondents reported that their classroom tests never had adequate coverage of listening.

The data collected from students concerning, grammar, vocabulary, reading and listening supported the information gathered from teachers. In the sample test papers, similar data was also collected in terms of the above four content areas. However, the data gathered from students in terms of speaking and writing contradicted with the data gathered from teachers. From this data one can deduce that students were not interested in speaking and writing classroom sample tests.

Regarding the proportion of the test items to the textbook concerned (item-8), the majority of the students, 60 (50.8%) reported that their classroom tests were sometimes according to the proportion which was given to the content areas of the textbook or to the focus which was given to the classroom instruction. On the other hand, 40 (33.8%) of students reported that their

classroom tests were often in proportion to the material covered in the classroom. The above two responses of students were also a good evidence for the information which were collected from teachers and classroom tests. Though teachers responded that their tests were in proportion to the material taught in the classroom, the collected sample tests did not reflect that. Table14 summarizes the students’ response to the match, coverage and proportion of item content.

Table 14. Students’ Responses to the Match, Coverage and Proportion of Item Content

| Statements | Often | | Sometimes | | Never | |
|--|-------|---------|-----------|---------|-------|---------|
| | No. | percent | No. | percent | No. | Percent |
| 1. Do your tests match with the contents of the textbook? | 90 | 76.2 | 20 | 16.9 | 8 | 6.7 |
| 2. Do your tests have adequate coverage of grammar items? | 90 | 76.2 | 28 | 23.7 | - | - |
| 3. Do your tests have adequate coverage of vocabulary items? | 79 | 66.9 | 32 | 27.1 | 7 | 5.9 |
| 4. Do your tests have adequate coverage of reading items? | 72 | 61 | 27 | 22.8 | 19 | 16.1 |
| 5. Do your tests have adequate coverage of speaking items? | 28 | 23.7 | 60 | 50.8 | 30 | 25.4 |
| 6. Do your tests have adequate coverage of listening items? | 10 | 8.4 | 30 | 25.4 | 78 | 66.1 |
| 7. Do your tests have adequate coverage of writing items? | 36 | 30.5 | 50 | 42.3 | 32 | 27.1 |
| 8. Do the items in the tests in proportion to the content areas of the text? | 40 | 33.8 | 60 | 50.8 | 18 | 15.2 |

4.4.2.2. Students’ Responses to Variety and Proportion of Test Formats

When we see the variety of formats used in the test items (item-9), the majority of the students 78(66.1%) reported that their classroom tests often employed a variety of(both subjective and objective) test formats. This information is almost similar to the information which was gathered from the sample teachers. However, this information of teachers and students was not evidenced by the formats used in the collected test papers. Although it was not in the expected way, varieties of test formats were used in the collected test papers- but all the formats are objective types.

Concerning the proportion of test formats (item-10), 50(42.3%) of the respondents reported that their test sometimes employed the test formats in proportion with the type of formats practiced in the classroom. Though this response of students negated with the responses of teachers, it almost agreed with the information collected from the analysis of the sample tests. This can inform one that students were dissatisfied with the improper balance of the formats used in the textbook. Table 14 presents students’ response’ to variety and proportion of test formats.

Table 15. Students' Responses to Variety and Proportion of Test Formats

| Statements | Often | | Sometimes | | Never | |
|---|-------|---------|-----------|---------|-------|---------|
| | No. | percent | No. | Percent | No. | Percent |
| 9.Variety of test formats in the tests | 78 | 66.1 | 33 | 27.9 | 7 | 5.9 |
| 10.proportion of formats in tests in relation with the textbook | 35 | 29.6 | 50 | 42.3 | 33 | 27.9 |

4.4.2.3. Students Response to the Emphasis Teachers Give in Teaching and Testing

Regarding the emphasis teachers give in teaching (item-11), students reported that grammar, reading and writing received the first three emphasized ranks respectively(see Table 16). Listening received the least emphasis when teachers taught English.

In the case of grammar, teachers' response and the students' response remained the same. The two data sources were also supported by the time allotment given to grammar. This agreement was a good evidence for the coverage that grammar received in the collected sample tests.

To the focus that teachers gave to tests (item-12), students reported that grammar, reading and vocabulary were the first the second and third emphasized areas respectively(See Table 16). This data exactly fitted with the teachers' response from the questionnaire. These three areas were also emphasized in the collected test papers. The difference was that in the collected test papers vocabulary was the second emphasized area next to grammar.

Writing which was given the second emphasis according to the time allotment was not given due emphasis in the sample tests. The emphasis that writing received in teaching was proved from different sources. Both teachers and students reported that it was given consideration in the teaching learning process. In the time allotment, writing was given the second emphasis next to grammar. However this emphasis was not reflected in the collected test papers. From this again one can deduce that test contents of classroom tests were not decided based on the emphasis the contents received in the teaching learning process. As Hughes (1989) said, it might have been decided based on what is easy to test. This, then, could lower the content validity of the classroom tests. Table 16 presents students response to the emphasis teachers gave in teaching and testing.

Table 16. Students' Responses to the Emphasis Teachers Gave in Teaching and Testing

| N o | Reading | | | | | | Vocabulary | | | | | | Grammar | | | | | | Speaking | | | | | | Listening | | | | | | Writing | | | | | |
|--------|---------|---|---|---|---|---|------------|---|---|---|---|---|---------|---|---|---|---|---|----------|---|---|---|---|---|-----------|---|---|---|---|---|---------|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 1 | 2 | 3 | 4 | 5 | 6 | 1 | 2 | 3 | 4 | 5 | 6 | 1 | 2 | 3 | 4 | 5 | 6 | 1 | 2 | 3 | 4 | 5 | 6 | 1 | 2 | 3 | 4 | 5 | 6 |
| 1 | 2 | 2 | 3 | 1 | 3 | 2 | 1 | 1 | 3 | 1 | 1 | 2 | 5 | 2 | 1 | 9 | 1 | 5 | 1 | 2 | 1 | 1 | 2 | 2 | - | 6 | 1 | 2 | 4 | 3 | 1 | 2 | 2 | 2 | 2 | 1 |
| | 1 | 3 | 0 | 8 | | 3 | 8 | 6 | 4 | 8 | 2 | 0 | 0 | 8 | 3 | | 3 | 0 | 2 | 9 | 6 | 5 | 6 | | | 1 | 7 | 0 | 4 | 9 | 3 | 1 | 0 | 5 | 0 | |
| 2 | 2 | 2 | 2 | 2 | 1 | 1 | 2 | 3 | 1 | 1 | 1 | 1 | 5 | 2 | 1 | 1 | 1 | 5 | 8 | 1 | 2 | 2 | 3 | - | - | 1 | 2 | 3 | 3 | 1 | 2 | 2 | 1 | 1 | 2 | |
| | 2 | 1 | 5 | 1 | 5 | 4 | 1 | 8 | 9 | 3 | 3 | 4 | 2 | 7 | 7 | 1 | 0 | | | 9 | 4 | 8 | 4 | | | 8 | 8 | 7 | 5 | 8 | 4 | 0 | 5 | 5 | 6 | |

Number 1 and 2 down refers to the question No. in the questionnaire.

Number 1-6 across refers to the rank that students give to the different content areas

Chapter Five: Summary, Conclusion, and Recommendation

In this chapter, the summary of the findings, the conclusions drawn based up on the findings and recommendations were provided.

5.1. Summary of the Findings

In this study an attempt was made to assess the content validity of Grade 9 English language teacher-made classroom tests of Bahir Dar town.

To achieve this purpose the following two basic questions were set to guide the researcher.

1. Are all the major and sub-language skills of Grade 9 English language textbook proportionally included in Grade 9 English language classroom tests?
2. Do the formats in the classroom tests adequately reflect the formats in the Grade 9 English language textbook?

Comparison between the time allocation for the different contents of the text book and the frequency of the test items in each content areas, and comparison between test formats of the textbook and the sample test items were made. Subjects' responses were also analyzed. The results of the findings were summarized as follows:

1. The relationship between the time allotment to the contents of text book and the frequencies of sample test papers
 - The sum of the time allotted to the different content areas of the text book and the frequencies of test items maintained moderate (0.48) correlation.
 - Though the syllabus and teachers guide of Grade 9 English recommend communicative approach, teachers focused on teaching grammar. As a result, grammar items are the dominant test items.
 - Listening was not considered at all.

- Improper allocation of test items among content areas. Writing, which was given more time in the textbook received little share in the tests and vocabulary, which was given less time than grammar had big share in the sample tests.
- Contents of the sample tests were determined by easiness to test.

2. The relationship between the formats of the text book and sample tests

- The formats were hardly correlated (0.12).
- Multiple -choice format was the dominant format in the tests.
- The dominant test formats in the textbook, like write and short answer were completely ignored.
- All the test formats in the tests were objective items.
- Productive skills were tested indirectly.
- Test formats were determined by easiness to mark.

5.2. Conclusion

As noted before, comparison between the time allocation for the different contents of the text book and the frequencies of the test items in each content area, and comparison between frequencies of test formats of the textbook and the sample test items were made. Then,

1. Though the time allotted to the whole content areas correlates moderately(0.48) to the items which appear in the sample tests, disparities have been observed among the different test items in relation to the time allotment for different content areas. Some content areas, which have been given coverage in the textbook, were not completely considered. On the other hand some other content areas were not represented proportionally. But, testing scholars, like Heaton (1991) advised test makers to relate the different areas covered in the test to the length of time spent on teaching those areas in class.
2. The test formats which are employed in the sample tests also do not proportionally reflect the formats which are found in the textbook. Though the textbook employed objective and subjective test formats, the sample tests employed only objective formats, specifically multiple choice format, which is not the dominant test format in the textbook. As a result, the correlation of test formats is found to be negligible (0.1). However, testing scholars, like Atkins, et al. (1996) recommended that achievement tests should be tested in the way they have been practiced in the classroom.

Then, from the above findings, it can be concluded that the grade 9 English language teacher made classroom tests relatively maintain low content validity.

5.3. Recommendations

In the light of the findings obtained from the study, the following recommendation can be drawn:

- Teachers should check their tests whether they give weight for the content areas according to the emphasis they gave in teaching those areas or not.
- Teachers should check whether they use variety of test formats as they were practiced in the classroom or not.
- Concerned authorities (Zone education office) should provide teachers with in-service training on language testing, specifically progress and achievement test construction.
- Schools should provide materials for testing listening.
- Textbook and syllabus writers should limit time for the language skills as it could inform teachers where to relatively emphasis.

Bibliography

- Airasian, P.W.1991. **Classroom Assessment**. NY: McGraw-Hill
- Alderson, J.C., Clapham, C. and Wall, D. 1995. **Language Test Construction and Evaluation**. Cambridge: Cambridge University Press
- Anastasi, A. 1988. **Psychological Testing**. New York: Macmillan Publishing Company
- Anderson, G. 1990. **Fundamentals of Educational Research**. London: Longman
- Argaw Sileshi. 2004. 'The Content Validity of Grade 10 National English Language Examination'. Unpublished MA. Thesis. Addis Ababa: A.A.U
- Asmare Miheret. 2008. 'An assessment of Awassa Health Science College English Language Tests'. Unpublished MA. Thesis. Addis Ababa: AAU
- Atkins, J., Hailom Banteyirga and Nuru Mohammed. 1996. **Skills Development Methodology**. part 2. Addis Ababa: Addis Ababa University.
- Bachaman, L.F .1990. **Fundamental Considerations in Language Testing**. Oxford: Oxford University Press.
- Bachman, L. F and Palmer,A.S.1996. **Language Testing in Practice**. Oxford: OUP
- Baker, D. 1991. **Language Testing: A Critical Survey and Practical Guide**. London: Edward Arnold.
- Best, J. and Kahn, J. 1993. **Research in Education**.(7th.ed.). New Delhi: Prentice Hall.
- Brown, J. D.1996. **Testing in Language Programs**. Upper Saddle River, NJ: Prentice Hall Regents
- _____ 2004. **Language Assessment: Principles and Classroom Practices**. Longman: Longman com.

- Chapelle, C. A. 1999. Validity in Language Assessment. **Annual review of Applied Linguists**. Vol.19, 254-272. New York: CUP.
- Davies, A. and Upshur, T. 1993. Language Testing. Vol.10, No. 3. Great Britain.
- Gronlund, N.E. 1982. **Constructing Achievement Tests**. University of Illinois, Prentice Hall
- Harmer, J.2009. **How to Teach English(4th ed.)**.London: Pearson Longman.
- Harrison, A. 1983. **A Language Testing Handbook**. London: Macmillan.
- Haynes, S,N., Richard, D. S. and Kubany,E.S.1995. Content Validity in Psychological Assessment: **A Functional Approach to Concepts and Methods**. Vol. 7, No.3, 238-247.
- Heaton, J. B.1990. **Classroom Testing**. London: Longman Group Uk. Ltd.
- _____ 1991. **Writing English Language Tests**. A Practical Guide for Teachers of English as a Second or Foreign Language. London :Longman Group UK. Ltd.
- Henning, G.1987. **A Guide to Language Testing**: Cambridge, MA.: Newbury House.
- Hughes, A. 1989. **Testing for Language Teachers**. Cambridge: Cambridge university Press
- _____ 2008. **Testing for Language Teachers (8th ed.)**. Cambridge: CUP
- Kassaw Baye .2006. ‘An Exploration of the Content Validity of Grade 12 National English Language Examination’.Unpublished MA Thesis. Addis Ababa: Addis Ababa University
- Kifle Kebede. 1995. ‘An Assessment of the Content Validity of High School English Language Classroom Tests in Relation to the Textbook concerned’. Unpublished MA. Thesis. Addis Ababa: Addis Ababa University
- Madsen, H.S. 1983. **Techniques in Testing**. London: Longman
- McNacmara, T. 2000. **Language Testing**. Oxford: OUP
- Ministry of Education. 2005. **English for Ethiopia Grade 9 Syllabus**. Addis Ababa: Mega Publishing Enterprise.

- Ministry of Education. 2005. **English for Ethiopia Grade 9 Students' Book**. Addis Ababa: Mega publishing Enterprise.
- Ministry of Education. 2005. **English for Ethiopia Grade 9 Teacher's Book**. Addis Ababa: Mega Publishing Enterprise.
- Oller,J.1989. **Language Tests at School**. London: Longman.
- Popham, W. J. 2003. **Test Better Teach Better: The Institutional Role of Assessment**. USA: Macmillan
- Sarantakos ,S. 2005. **Social Research**. New York: Pal Grave Macmillan.
- Sireci, S.G.2007. On Validity Theory and Test Validation: *Educational researcher* Vol.36, No.8pp477-481
- Tawio, A.1998. **Fundamentals of Classroom Testing**. Delhi: Vicas Publishing Company.
- Weir, C. J. 1990. **Communicative Language Testing**. New York: Prentice Hall, Inc.
- _____ 1993. **Understanding and Developing Language Tests**. New York: Prentice Hall.
- _____ 2005. **Language Testing and Validation**. Basingstoke: Pal Grave Macmillan.

Appendix- 1

Look this tasks carefully at your grade 9 English language textbook and decide the time that each task could take while practicing it. You can put the time in minutes in front of each task.

The time for the unit coverage is given. You can use it as a reference (starting point) to decide the time for each task in each unit.

| Content Areas | Time given In minutes | Content Areas | Time given |
|---|----------------------------------|-------------------------------------|-----------------------|
| Special unit(5 periods) | | Unit 2(continuation) | |
| A. Tasks in Speaking | | Tasks in Vocabulary | |
| -obtaining information from a speech | 20 | -Synonyms | 40 |
| -introduction | 20 | -Prefixes | 40 |
| -Giving speech | 40 | Tasks in Grammar | |
| B. Tasks in Writing | | -Conditional sentences | 120 |
| -paragraphs | 40 | Tasks in Speaking | |
| -The simple sentences | 20 | -Asking for repetition | 80 |
| -Punctuating simple sentences | 20 | Tasks in Listening | |
| -using 'but' and 'and' to join sentences | 20 | -Listening for specific information | 40 |
| -using pronouns and possessive adjectives to avoid repetition | 20 | Tasks in Writing | |
| Unit- 1(13 periods) | | -Writing instructions | 80 |
| Tasks in Reading | | Unit 3 (13 Periods) | |
| -pre reading questions | 10 | Tasks in Reading | |
| -reading for specific information | 10 | -reading for specific information | 10 |
| -reading for detailed information | 60 | -guessing meanings of words | 20 |
| Tasks in vocabulary | | -reading for detailed information | 60 |
| Guessing meanings of words related the text | 40 | Tasks in Vocabulary | |
| Verbs 'to take' and 'to last' | 20 | -guessing meanings of words | 40 |
| -the meanings of 'to make' | 20 | -the meanings of change | 40 |
| Tasks in Grammar | | Tasks in Grammar | |
| -verbs followed by infinitive | 80 | -the passive | 80 |
| -verbs followed by -ing | 40 | -used to | 40 |
| -ways of expressing purpose | 40 | -Conditional sentences | 20 |
| Tasks in speaking | | Tasks in Speaking | |
| -Pair and group discussion | 80 | -asking to borrow something | 80 |
| Tasks in Listening | | Tasks in Listening | |
| -Listening for specific information | 40 | -listening to specific information | 40 |
| Tasks in Writing | | Tasks in Writing | |
| -composition writing | 30 | -Completing summary | 40 |
| -note making | 10 | -punctuation | 40 |
| -reordering sentences | 40 | | |
| Unit -2 (13 Periods) | | | |
| Tasks in Reading | | | |
| -pre-reading activity | 10 | | |
| Reading for specific information | 10 | | |
| Deducing meanings of new words | 30 | | |
| Reading for detailed information | 80 | | |

| Content Areas | Time given | Continued from unit 5 | Time given |
|---|-------------------|--|-------------------|
| Unit-4(13periods) | | -The passive | 40 |
| Tasks in Reading | | -Reason clause | 40 |
| -Reading for specific information | 10 | -Result clause | 40 |
| -Guessing meanings of words | 30 | Tasks in speaking | |
| -Reading for detailed information | 60 | -Asking for some ones opinion | 40 |
| Tasks in Vocabulary | | Tasks in Listening | |
| -Guessing meanings of words related to the text | 40 | -Listening for specific information | 40 |
| Tasks in Grammar | | Tasks in writing | |
| -The passive | 40 | -Reordering sentences | 40 |
| -Reason clause | 40 | Unit -6 (12 periods) | |
| -Using so that to express result | 60 | Tasks in Reading | |
| Tasks in Speaking | | -Pre reading activity | 5 |
| -Asking for, giving and refusing information | 80 | -Reading for specific information | 5 |
| Tasks in Listening | | -Guessing meanings of new words | 10 |
| -Listening for specific information | 40 | -Reading for detailed information | 40 |
| -Dictation | 40 | Tasks in Vocabulary | |
| Tasks in Writing | | -Guessing meanings of words related to the text | 10 |
| -Guided composition writing | 40 | -Synonyms | 10 |
| -Note making | 40 | -Steal and rob | 40 |
| Unit -5 (13 Periods) | | Tasks in Grammar | |
| Tasks in Reading | | -Reported speech | 80 |
| Pre reading activity | 5 | -For and since | 40 |
| Reading for specific information | 5 | Tasks in Speaking | |
| Guessing meanings of new words | 30 | -Asking for someone's opinion and giving opinion | 80 |
| Reading for detailed information | 40 | Tasks in Listening | |
| Tasks in Vocabulary | | -Listening for specific information | 40 |
| Guessing meanings of words related to the text | 20 | -Dictation | 40 |
| Word family | 40 | Tasks in Writing | |
| Using feel + adjective | 20 | -Guided composition | 40 |
| Tasks in Grammar | | -Reordering sentences | 40 |
| Reported speech | 80 | | |
| Drawing conclusion using must be | 40 | | |

Appendix- 2

Inter Rater Reliability among Raters for the Time Allotted for the Different Content Areas

Time given in periods by the raters for the different content areas

| Raters | Teaching experience | Content areas and the time given in periods | | | | | |
|----------|------------------------|---|------------|---------|----------|-----------|---------|
| | | Grammar | vocabulary | Reading | Speaking | Listening | Writing |
| Rater- A | BA 17 years in English | 21 | 12 | 14 | 13 | 8 | 15 |
| Rater- B | BA 20 years in English | 24 | 11 | 13 | 14 | 8 | 16 |
| Rater- C | BA 24 years in English | 22 | 12 | 13 | 11 | 9 | 14 |
| Rater- D | MA 15 years in TEFL | 23 | 10 | 13 | 14 | 9 | 13 |

Ranks given by the raters for the different content areas

| Content areas | Ranks by raters | | | | | Sum |
|---------------|-----------------|----------|----------|----------|----|-----|
| | Rater- A | Rater- B | Rater- C | Rater- D | | |
| Grammar | 1 | 1 | 1 | 1 | 4 | |
| Vocabulary | 5 | 5 | 4 | 5 | 19 | |
| Reading | 3 | 4 | 3 | 3 | 13 | |
| Speaking | 4 | 3 | 5 | 2 | 14 | |
| Listening | 6 | 6 | 6 | 6 | 24 | |
| Writing | 2 | 2 | 2 | 3 | 9 | |
| Sum | 21 | 21 | 21 | 20 | 83 | |

Having the above information crobach alpha was used to compute inter rater reliability.

The formula is $\alpha = (K/K-1)(1 - \sum S^2_j / S^2_x)$

Where $\sum S^2_j =$ sum of the variance for the ranks of the raters for each content areas.

$S^2_x =$ Variance of the rank for the content areas

K = the number of raters

Then, $\sum S^2_j = 13.9$

$S^2_x = 50.8$

K= 4

$\alpha = (4/4-1)(1-13.4/50.18)$

$\alpha = 0.9$

when the computed result approaches to one it indicates agreement between judges(raters).

Then, alpha coefficient of 0.9 indicates high agreement among raters.

Appendix -3A

Tana Haik General Secondary School English Mid- Exam for grade 9th year 2003

1. Choose the correct answer from the given alternatives

1. Which one of the following is a complete sentence?

- A) He has B) The tiger killed C) The flower is yellow D) the boys

2. My father is Doctor _____ my mother is a nurse.

- a) but B) and C) or D) because

3. I called for help _____ nobody Game

- A) but B) and C) or D) because

4. He enjoys _____ detective stories

- a) to read B) read C) reads D) reading

5. They arranged _____ at the party.

- A) meeting B) meet C) to meet D) met

6. If he _____ 100 birr, he will buy a book

- A) get B) getting C) got D) gets

7. unless you study hard, you _____ the exam

- A) earns B) builds C) forces D) achieves

8. I do not want to do this thing but my sister makes me to do it. The meaning of the underlined word is _____

- A) earns B) builds C) forces D) achieves

9. I don't have any change in my pocket to drink tea. Change in this sentence means _____

- A) Balance of money B) coin C) different D) to replace

10. usually, people go to pharmacy _____ medicine

- A) buy B) to buying C) in order to buy D) bought

11. which of the following correctly punctuated?

- A) where are you going C) Hey, you! Go a way!

- B) my mother asked me a question? D) shut the door

12. Teacher: can you tell me how to use a computer?

Student: Sorry teacher, would you mind _____ that again please?

- A) to repeat B) repeat C) repeating D) repeated

II. Fill the blanks with words from the given list to complete the sentence

Diagnosis, Flapping, Abstain, refuse, isolate

13. if you _____ yourself from penetrating sexual activities, you will protect yourself from HIV/AIDS

14. In the countries where treatments are unavailable most. AIDS patients die within a few years of _____

15. Birds are flapping by _____ their wings

16. do not _____ your parents advice

Appendix- 3B

Tana Haik Secondary School 2003.E.C First Semester English final Examination for Grade 9

I read the following interview made by Ethiopian sports for Hailie G/Silasie and answer the questions that follows based on the interview

E.S: First of all, I would like to thank you personally and on behalf of all Ethiopians living abroad and the sounds of your fans else where, for guarantying me this interview

H.G: the pleasure is all mine

E.S: Looking back at your career, did you ever imagine that you are goig to be successful?

H.G: well, I wanted to be successful and follow in the foot steps of Abebe Bikila and Miruts Yifter, but I never imagine that I would become this successful breaking world records.

E.S: of all your victories, which one do you considere to tbe the most important one and why?

H.G: it is without a doubt my victory in the 10,000 meters, at the Atlanta Olympic Games in 1996, because it was my first Olympic medal

E.S: who were your idols when you were growing up?

H.G: When I was young, Abebe Bikila and Miruts Yifter were my idols

E.S: I was told that you train in both the Nether Lands and Ethiopia. If so how are the two trainings different?

H.G: that is not true. I do all my trainings in Ethiopia sometimes, when I have a few days to spare before a competition there, I do some trainings but that is about it. Some people might think that I actually live and train in the Netherlands because I was given a residence permit in that country but I am still an Ethiopian and Ethiopia is my home forever

E.S: what other sports do you like besides track and fields?

H.G: I enjoy watching boxing and I am also beginning to enjoy Soccer now although I didn't have any interest in the sport before

(Adapter from improve your English. A course for Ethiopian teachers face-to-face training material

1. career (in line 4) means:

2. some one who carry something 2. Profession 3. Sport 4. Runner

3. besides (in line 19) means ___ 1. Near 2. Support 3. In addition 4. Track

4. victory (in line 8) means _____

1. Success 2. Competition 3. Olympic game 4. Runner

5. imagine (in line 4) means _____ 1. Think 2. Hope 3. Certain 4. None

6. in what year Hailie G/Silasie got his first medal?

1.1997 2. 1996 3. 1999 4. 1998

7. Who were Hailie G/Silasie's idols?

1. Kenenisa & sileshi 2. Doctor G/M/Kostore 3. Abebe and Miruts

4. Mamo Woldie and Gebreamlak

7. Haile G/silasie was given a residence permit by _____

1) Ethiopia 2. Awasa 3. Netherland 4. Olympic

II. Choose the Correct Answer

8. which (one of the following) is a complete sentences?

1.The girl 2. She smiled picked 3. The dog chased 4. None of the above

9. she is very poor _____ she is always happy 1. Also 2.And 3. But 4. So

10. Ministry of education started _____ new text books to schools

1. To distributed 2. Distribute 3. Distributing 4. 1and 3 5. All

11. Many tourists come to Ethiopia _____ historical places

1. to visiting 2. I order to visit 3. Visiting 4. 1&2

12. You will be successful if you _____ hard. 1. Works 2. Work

3. Working 4. Worked

13. Pictures were "used to" to represent things used to in this sentence expresses

1. Habit 2. Purpose 3. Action 4. All

14. my brothe used to smoke cigarette mean

1. He is smoking cigarette now 2. He is not smoking cigarette now

3. he dislikes smoking cigarette now 4. It I not known

15. the evolution of the students can be made through examination. The under lined word is _____ 1.

Noun 2. Verb 3. Adjective 4. Adverb

16. senait always wears precious ornaments and quality dresses. When we deduce this sentences

1. Senait likes previous things 3. Senait must be rich or from a rich family

2. Senait borrows from her friends 4. Senait does not like ordinary dresses

17. most of the residences of Bahir Dar are complaining _____ the cost of every item is increasing from time to time

1. because 2. Since 3. As 4. All

18. Hiwot= what do you think will happen to child prostitution in Ethiopia? Alazar =I _____ that they will face many problems

1. think 2. Expect 3. Guess 4. All

19. which of the following is correct?

1. is it ok if I ask you're a question ? 2. Is it all right if I ask you questions?
3. would you mind if I ask you a question? 4. All

20. The box is ----- that we would not move it?

1. Too heavy 2.so heavy 3. Very heavy 4. All

21. The coffee is too hot to drink. This means

1. we can drink the coffee 2. We can't drink the coffee
3. It is very nice to drink 4. It is better to drink the coffee

22. which one is a reported speech/sentence?

1. I am very hungry now,'said Aster 2. She is very hungry now
3. she said that she was very hungry then

23. what is your conclusion?

Hana: All the girls like him

You: _____

1. He must be beautiful 2. He must be handsome 3. He likes girls 4. None

24. when we change active sentences into passive ones, the main verb in the passive changes to _____
1. V2 2. V3 3. V1 4. All

25. The two men went across the compound. The underlined word is replaced by

1. He and She 2. Them 3. Their 4. They

III. Fill the blanks with the words listed below:

Borrow, focus, robbed, communicate, successful

26. now a days people _____ with each other through e-mail

27. may I _____ your pen, please?

28. our teacher advised us to _____ our attentions on our lessons

29. the burglars stopped the bus and _____ many people last Monday

30. Ayele was a _____ businessman.

Appendix-3C

GHION SECONDARY AND PREPARATORY SCHOOL FIRST SEMESTER ENGLISH MID-EXAMINATION FOR GRADE 9 IN 2003 E.C

Choose the correct answer for the following questions

1. A: how do you do Almza?

B: ----- The response of 'B' will be _____

- A) I am fine C) how are you
B) Good morning D) how do you do

2. A: I am happy to meet you

B: _____

- A) Me too B) Good by C. I' M fine D) Hello

3. She has finished ----- her home work

- A) Do B) to do C) doing D. none

4. One is correct

- A) What beautiful girl she is? C) What is your father name!
B) Open the box, please? D) Clean the black board?

5. She made hen sister go to the liberty. Made means

- A) Cause of something C) achieve
B) Earn D) forced

6. His father likes Alemu. My sister wants to marry Alemu

The possessive adjective from of Alemu is _____

- A) He B) his C) him D) her

7. Many people visit Ethiopia. -----think that Ethiopia is very beautiful country.

- A) Their B) them C. they D) all

8. They arranged -----out side the school compound

- A) Meet B) Meets C) Meeting D) to meet

9. Kebede likes -----to radio

- A) Listening B) listen C) to listen D) A and B

10. The boy who was shinning the shoes of the people resign his job by himself. The underlined word is -

- A) Join B) Stop working C) start working D) get more money

Fill the black/given short answer

11. The dog's meat was taken by the bird. The pronoun that can replace the underlined word is -----

12. My mother gave me some books
 1 2 3 4 5 6

The direct object is expressed by number _____

13. He refused -----on the bed/sleep. The appropriate form of sleep that should be put in the black space is -----

14. Would you mind show us how to do the assignment?

- 1 2 3 4

From the above question the mistake is expressed by number -----

15. What is the difference between 'to' and 'in order to' in the expression of purpose?

Appendix- 3D

Ghion Secondary School 2003 E.C First Semester Final English Examination For Grade 9 9

I. Join the following statements and responses

“A”

1. Avoid, stop
2. Complication, health problem
3. Not understand, secret
4. Stop working because you don't want to do the job
5. Discriminate exclude from a common right or interest

“B”

- A. Resign
- B. Isolate
- C. Abstain
- D. Mystery
- E. Misery
- F. Sero

II. Choose the answer that BEST completes the blank space

6. If the room -----, there will not be pieces of paper on the floor.

- A) Cleans B) Clean C. is clean D. is cleaned

7. Don't go -----interested in the journey

- A) Unless you are not B. if you are not C. If you shall be D. if you are

8. A: I am tired of sitting for long

B: why not go to the cinema for a change, 'change' means -----

- A) To remove B) . Something different C) to alter D). become different

9. Many houses and animals -----by the earth quake

- A) Have destroyed C) have been destroying
B) Have destroying D) have been destroyed

10. Teacher: Turn to page 60 students

Student: -----?

Teacher: open you book on page 60.

- A) Have we a test C. what do you think will be the use
B) Is it all right D. sorry, what did you say

11. Many books ----- in the USA

- A) Publish B Are being published C) are publish D) published

12. I -----run five kilometers an hour when I was a boy

- A) Was using to B) Used for C. did n't use D. used to

13. A: Excuse me -----?

B: Yes, of course. Here you are

- A) Could you borrow me your pen, please C) Please borrow me the pen
B) Could you lend me your pen please D) Could you lend me the pen

14. My bike ----- yesterday

- A) Was stole B) was stealing C) was stolen D) stolen

15. Have you managed -----the problem?
 A) Solving B) Solve C) to solve D. solved
16. A: Why did you read fictions? B: -----
 A) Since I had no time C) Because to improve my reading skill
 A) To improve my reading skills D) usually at home
17. Enjoy-----; I will be back soon
 A. To watch B. watch C. watched D. watching
18. Unless he -----, he won't pass the exam.
 A) Study B) Studies C) is study D) studying
19. Water is a -----of hydrogen and oxygen
 A) Record B) discovery C) device D) combination
20. I am studying hard ----- I will take a test tomorrow
 A) If B) in order to C) because D) unless
21. "The coffee is too strong for me to drink" means:
 A) The coffee is very strong, so I can drink it C) The coffee is not as such strong to drink
 B) The coffee is so strong that I can drink it D) the coffee is not as such strong that I can't drink
22. Student: may I go out for a second, teacher?
 Teacher: -----, but do not stay longer
 A) No, I'm sorry B) Yes, you may C) Sit still, please D) No I'm afraid
23. To be -----you must save some money in the bank
 A) Economy B) economically C) economical D) economize
24. Why do you feel -----? Have you lost your money?
 B) Sadly B) sad C) sadden D) saddening
25. In the sentence 'He told me a story' the in direct object is:
 A) Me B) a story C) he D) told

III. Change the following sentences into passive voice

26. Coffee (drink) in Ethiopia -----
 27. The bridge (repair) last year -----
 28. Chinese (speak) in China -----

I. Change the following sentences into Reported speech (after a few days)

29. "I will take a test tomorrow." He said -----
 30. "the sun rises in the east," said Almaz -----

Appendix 3E

2003 E.C Fasilo General Secondary School First Semester Mid-Semester
English Test for Grade 9 Regular students

II. Choose the correct answer to fill the black spaces

1. Solomon and Yared agreed _____ the profit equally
A) Divided B) to divide C) dividing D) divides
2. Senant enjoys _____ the paino during her spare time
A) Plays B) played C)to play D) playing
3. Many students are working hard on their education ____ their goal
A) To achieve B) for achieve C) achieving D) achieve
4. You finish this work, you will not go home
A) if B) when C) unless D) otherwise
5. The teacher will explain the lesson if _____
A) ask him B) you ask him C) unless you ask D) asking him
6. Aster has just finished _____ the new novel
A) to read B) read C) for reading D) reading
7. My friend, Solomon _____ French fluently
A) spok B) spoke C)speaks D) speak
8. The journey to the airport _____ about half an hour
A) takes B) took C) finishing D) starting
9. I tried to reach the top shelf _____ i am not tall enough
A) and B) but C) or D) because

III. Choose the meanings of the underlined worlds

10. We will be free from AIDS if we abstain from sexual activity
A) Practice B) use rarely C) danger D) avoid
11. She refused to accept that there was a problem
A) Said no B) received C) took D) believed
12. The car was completely crushed against a truck
A) Grind B) damaged C) pressed D)forced Conversation

IV. Completed the dialogue by choosing the correct expression

13. _____ repeat what you have just said?
A) You can B) Could you C) Excuse me D) you could
14. _____ I did n't get what said
A) Could you B) would you mind C) you can D) I'm sorry
15. A. could I borrow your pen, please?
B. _____ (polite response)
A) Yes of course Here it is B) Why not you buy
C) Ask your father to buy you D) it is my own

Appendix- 3F

2003 E.C First Semester Fasilo Secondary School English Final exam for grade 9

1. Read the passage and answer the questions that follow

There was once a rich merchant whose wife fell ill, so he went to get a doctor. The doctor willingly comes to see the woman and to do what he could for her, but before he went into the house, he stopped for a short time. He knew that the merchant was famous for not paying. He therefore said to the man, " But if I do cure your wife, I am afraid you may not pay me/" " my dear doctor," answered the merchant," here I have thousand pieces of gold. Whether you cure my wife or whether you kill her, I will give you all this. The doctor was very happy and went into the house. When he reached the woman's beside it was soon clear to him that he could do little, she was very ill. He gave her medicine, but she soon died. He told the merchant that he was very sorry and asked for the money which he had been promised.

" did you kill my wife?" asked the man

" certainly not," answered the doctor

"that was impossible," answered the doctor

"well, you have not killed my wife; you have not cured my wife. Therefore, I will not pay you."

1. The women who fell ill was _____
 - a) The wife of the rich merchant c) the wife of the doctor
 - b) The grand mother of the merchant D) the relative of the merchant
2. Before the doctor went into the house, he _____ for a short time
 - A) Become angry C) said something
 - B) Stooped D) could do little
3. The doctor was very happy when _____
 - A) He want into the house C) the merchant promised him to pay him gold
 - B) He knew the women would die D) he reached the woman's beside
4. At last the rich merchant
 - A) Fell sorry for the death C) Refused to pay the money
 - B) Gave the doctor the money D) become happy
5. According to the passage, which one is wrong?,
 - A) The merchant did n't keep his promise
 - B) The doctor really cured the woman
 - C) The merchant was rich D) the merchant was not famous for paying
6. The underlined world 'willingly' means _____
 - A) Voluntarily B) Sadly C) angrily D) exactly
7. you may not pay me. (paragraph 1) " me" refers to _____
 - A) The merchant B) the doctor C) the woman D) the nurse

24. Alemu: DO you mind lending me your car?
 Dawit: _____. I do n't need it
 A) Certainly B) No, of course not
 C) Yes, of course D) yes, I do mind
25. This problem is _____ difficult _____ solve it.
 A) Sothat C) suchthat
 B) Totoo D) too.....t
26. Bekalu: Do you play football?
 Kirkim: No, but I _____
 A) Am used to C) will used to
 B) Used to D) None of the above
27. If she had studied harder, she would have passed the exam it means _____.
 A) Passed B) is clever C) Failed D) will pass
28. _____ i borrow your pen, please?
 A) Excuse me B) Why C) do D) may
29. Sugar _____ from sugar cane or sugar beet
 A) Was made B) will be made C) is made D) was being made
30. She opened the windows _____ she needed fresh air.
 A) Because B) Since C) as D) all are possible answers

Appendix- 3G

Bahir Dar Zuria Secondary School First Semester English mid- exam for grade 9

Choose the best answer for the following questions

1. If she -----, I -----invite her
A) Will comes B) comes/will C) come/will D) will/come
2. I called for help---no body came
A) And B) But C) or D) none
3. If you see any AIDS victim, -----him
A) Helped B) helps C) help D) helping
4. Both he-----she wore shirts
A) And B) but C) for D) none
5. The burglar did not make a sound ----were experts
A) He B) She C) they D) all
6. He makes 500 birr a month. What is the meaning of mmake in this sentence?
A) To order B) to complete C) to earn D) none
7. She enjoys -----foot ball
A) Playing B) Play C) to play D) all
8. I should study hard -----better marks
A) To score B) in order to score C) score D) A&B
9. This is my sister -----name is Aster
A) She B) Hers C) her D) all
10. If your boil water at 100 -----into vapor
A) Will change B)changes C) changing D) none

II-Write true for the correct and false for the wrong sentences

11. A paragraph deals with one main idea
12. A paragraph always begins on a new line
- 13.'The dog bit' is a complete sentence

III Write out the following sentences with correct punctuation

14. Daniel and Ali went to school
15. Where are you going

Appendix- 3H

Bhir Dar Zuria G.S.S.S. Department of English 2003E.C First Semester Final Exam Grade 9th

I read the following passage easefully and answer the questions that follow it

The dignity (respect) given to labour (work) in the western countries is by far better than it is given in our country, Ethiopia. In the west, students work hard to get money for their studies. There are individuals, who are now at the top, but who have washed plates at hotels, have picked garbage (dirty things), have shoveled snow from the streets and have washed the cars of others. There is no shame involved in doing these kinds of works in the western. On the other hand, here is Ethiopia we feel ashamed to do even our own work such as sweeping our own rooms or even cleaning our own vehicles (cars). Some jobs like picking up garbage or cleaning our own pathways are beyond our imagination. The worst part is that some people tend to look down on others who have involved in manual jobs.

The honest labour of men in all forms of welfare schemes and activities, has dignity of its own. Those who don't understand this dignity and who avoid participating in any labour or detest the idea of hard work, must beg or steal to stay alive.

Therefore, we Ethiopians should change this wrong attitude and never despise any labour howsoever lowly paid

1. In the western countries people do not feel ashamed of doing such jobs as
 - A) Washing plates at hotels B) picking up garbage
 - B) Washing the cars of others D) all are answers
2. What is the worst part some Ethiopians tend to do with regard to giving respect to work?
 - A. They do not wash the cars of others B) they are ashamed to do their own work
 - C) They are ashamed to do their own work
 - D) they support others who have involved in manual jobs
3. The writer advises ---- not to despise manual jobs and not to be ashamed of doing them
 - A) Europeans B) Ethiopians C) students in the west D) people in the west
4. According to the passage, people who... must beg or steal to stay alive
 - A) Dislike the idea of hard work C) believe in the dignity of labour
 - B) Participate in manual jobs D) respect manual jobs
5. Which of the following is not true according to the passage?
 - A) People should not participate in manual jobs as manual jobs have low value
 - B) People in the western countries give more respect to work than people in Ethiopia
 - C) The writer to the passage has positive attitude towards labour
 - D) Washing plates, picking up garbage, washing the cars of others, etc.
6. _____ these kinds of works(line 4) refers to
 - A) Disliking manual jobs
 - B) Discouraging other who have involved in manual jobs
 - C) Being ashamed of doing one's own works
 - D) Washing plates, picking up garbage, washing the cars of others, etc.

7. this wrong attitude ... (line 10) refers to:-
 A) Understanding the dignity of labour C) Giving value to labour
 B) Disliking and/or not respecting labour D) Participating in any labour
8. ----- dignity -----(line 1) means a) unnecessary b) taking no care
 c) respect d) disadvantage
- 9) ----- look down on ----- (line 7) means A) give value to
 b) give little no regard to c) encourage d) support
10. ----- despise ----- (line 10) means:- A) participate in
 B) give value to C) dislike D) respect

II. Choose the best word or phrase that can replace the underlined one

11. the most serious handicap for the people on high land areas is the poor quality of the soil. This is because high land areas are exposed to erosion
 a) problem b) importance c) solution D) advantage
12. Ethiopia is a country with a great ethnic, religious, cultural and language diversity. For example, there are about 80 languages in the country
 A) similarity B) variety C) sameness D) disappearance
13. her job is temporary. She has been employed only for six months
 A) Permanent B) interesting C) lasting for a short time D) lasting for ever
14. The main reason for the economic decline of the country is the continuous drought that is the presence of little or no rain at all the country
 A) development B) growth C) increase D) decrease
15. Deaf people can communicate with other people by carefully observing the mouth of the speaker, by using a finger alphabet and by learning a sign language
 a) exchange information b) transmit diseases c) be excluded (isolated) from d) hear or listen
16. Abebe: Do you like to go to the stadium to watch foot ball, match now?
 Tesfaye: yes, like to go.... I have to study for the exam now
 a) but b) so c) and d) or
17. Ato Tesfaye went to the market yesterday and Bought an ox.
 A) him B) his C) he D) B&C
18. Tewodros denied ___ my boy A) to take B) to taking C) taking D) taking
19. Dawit: What is the purpose of studying hard? Almaz:-----
 A) in order to score excellent grade B) To pass the exam
 C) Not to fail in the exam D) A&B E) B&C
20. She continued -----until she was 1. A) work B) working C) to work D) B&C
21. SA: What will you do if a lion -----in your class?
 SA: I will shout A) enters B) entered C) will enter D) was entering
22. Let us say you are learning biology. In the middle of the class, you have missed the idea which the teachers has said. How do you ask for repeating the idea?
 a) teacher, repeat what you said b) could you repeat again, please?
 c) I do not mind if you repeat d) do you mind if repeat?
23. kebede: who killed this dog?
 Tesfaye: I do not know who killed it, how ever, it ----- on the road
 a) found b) was found c) were found d) none

24. if any student disturbs , ----- to the director immediately
 a) he would have reported b) he reported c) report d) he would report
25. he used to play foot ball when he was young. This sentence is nearly to mean -----
 a) he is still young and plays foot ball b) he is now old but the plays foot ball
 c) he is now old, so he does not play foot ball d) he still plays the game
26. SA: ... I have forgotten my pen. Can I ----- your pen?
 SA: Yes, you can a) lend b) borrow c) lent d) borrowed
27. the foot ball match was postponed to the next day -- there was a heavy rain
 a) so b) but c) and d) because
28. he is too old to walk alone. This is to mean
 a) he is old enough to walk b) he can not walk alone
 c) he can walk alone d) he is very old but he can walk
29. SA: do you mind if I use you book?
 SA: (imagine the response is giving). A) yes, I do b) yes, I would
 c) no, I don't d) No, I am afraid
30. if you study very hard, you will be a -----student
 a) succeed b) successful c) success d) successfully
31. which sentence is correctly punctuated?
 a) Alemu said, " The teacher ordered me to come."
 b) " The teacher ordered me to come." Alemu said
 c) " the teacher ordered me to come," Alemu said
 d) " the teacher ordered me to come", Alemu said
32. The thief _____ her necklace yesterday.
 A) stole b) rob c) has robbed d) A&B
33. they have taught in this school.....1999. a) in b) for c) since d) at
34. I haven't seen her Five years. A) since b)in c) for d) at
35. some one... the bank last night. A) stole b) robbed c) has stolen d) robs
36. Is used to force nails into wood or hit things.
 a) Axe b) Needle c) Hammer d) Spade
37. my brother always carrier a lot of small change in his pocket the underlined word means.....
 a) remove b) coins c) look or become different d) all
38. Abebe's cloth is wet. He-----in the rain a) is b) will be c) must be d) be
39. we finished our work ----we couldn't get the reward
 a) to lately to b) too lately c) so lately that d) so lately to
40. Although my mother was working the whole day, she---- tired.
 a) feel b) felt c) didn't feel d) A&B

Appindix-4A

Frequencies of content areas of Tana Haik Mid and final tests
by question number and content areas

| Quest ion No. | Mid Test | | | | | | Final | | | | | |
|---------------------|---------------|---|---|---|---|-----------------|---------------|---|----|---|---|---|
| | Content areas | | | | | | Content areas | | | | | |
| | R | V | G | S | L | W | R | V | G | S | L | W |
| 1 | | | | | | X | X | | | | | |
| 2 | | | X | | | | X | | | | | |
| 3 | | | X | | | | X | | | | | |
| 4 | | | X | | | | X | | | | | |
| 5 | | | X | | | | X | | | | | |
| 6 | | | X | | | | X | | | | | |
| 7 | | | X | | | | X | | | | | |
| 8 | | X | | | | | | | | | | X |
| 9 | | X | | | | | | | X | | | |
| 10 | | | X | | | | | | X | | | |
| 11 | | | | | | X | | | X | | | |
| 12 | | | | X | | | | | X | | | |
| 13 | | X | | | | | | | X | | | |
| 14 | | X | | | | | | | X | | | |
| 15 | | X | | | | | | X | | | | |
| 16 | | X | | | | | | x | | | | |
| 17 | | | | | | | | | X | | | |
| 18 | | | | | | | | | | X | | |
| 19 | | | | | | | | | | | | X |
| 20 | | | | | | | | | X | | | |
| 21 | | | | | | | | | X | | | |
| 22 | | | | | | | | | X | | | |
| 23 | | | | | | | | | X | | | |
| 24 | | | | | | | | | X | | | |
| 25 | | | | | | | | | X | | | |
| 26 | | | | | | | | X | | | | |
| 27 | | | | | | | | X | | | | |
| 28 | | | | | | | | X | | | | |
| 29 | | | | | | | | X | | | | |
| 30 | | | | | | | | x | | | | |
| Total | - | 6 | 7 | 1 | | 2 | 7 | 7 | 13 | 1 | | 2 |
| Total mid = 16 | | | | | | Total final= 30 | | | | | | |

Key

R- reading L- listening
V- vocabulary S- speaking
G- grammar W- writing

Appendix-4B

Frequencies of content areas of Ghion Mid and final tests
by question number and content areas

| Quest -ion No. | Mid Test | | | | | | Final | | | | | |
|----------------------|---------------|---|----|---|---|---|-----------------|----|----|---|---|---|
| | Content areas | | | | | | Content areas | | | | | |
| | R | V | G | S | L | W | R | V | G | S | L | W |
| 1 | | | | X | | | | X | | | | |
| 2 | | | | X | | | | X | | | | |
| 3 | | | X | | | | | X | | | | |
| 4 | | | | | | X | | X | | | | |
| 5 | | X | | | | | | X | | | | |
| 6 | | | X | | | | | | X | | | |
| 7 | | | X | | | | | | X | | | |
| 8 | | | X | | | | | X | | | | |
| 9 | | | X | | | | | | X | | | |
| 10 | | x | | | | | | | | | X | |
| 11 | | | X | | | | | | X | | | |
| 12 | | | X | | | | | | X | | | |
| 13 | | | X | | | | | | | | X | |
| 14 | | | X | | | | | | X | | | |
| 15 | | | X | | | | | X | | | | |
| 16 | | | | | | | | | X | | | |
| 17 | | | | | | | | | X | | | |
| 18 | | | | | | | | | X | | | |
| 19 | | | | | | | | X | | | | |
| 20 | | | | | | | | | X | | | |
| 21 | | | | | | | | | X | | | |
| 22 | | | | | | | | | | | X | |
| 23 | | | | | | | | X | | | | |
| 24 | | | | | | | | X | | | | |
| 25 | | | | | | | | | X | | | |
| 26 | | | | | | | | | X | | | |
| 27 | | | | | | | | | X | | | |
| 28 | | | | | | | | | X | | | |
| 29 | | | | | | | | | X | | | |
| 30 | | | | | | | | | X | | | |
| Total | | 2 | 10 | 2 | | 1 | | 10 | 17 | 3 | | |
| Total mid= 15 | | | | | | | Total Final= 30 | | | | | |

Key

R- reading L- listening
V- vocabulary S- speaking
G- grammar W- writing

Appendix-4D

Frequencies of content areas of Bahir Dar Zuria Mid and final tests by question number and content areas

| Quest -ion No. | Mid Test | | | | | | Final | | | | | |
|----------------------|---------------|---|---|---|---|----------------|---------------|----|----|---|---|---|
| | Content areas | | | | | | Content areas | | | | | |
| | R | V | G | S | L | W | R | V | G | S | L | W |
| 1 | | | X | | | | X | | | | | |
| 2 | | | X | | | | X | | | | | |
| 3 | | | X | | | | X | | | | | |
| 4 | | | X | | | | X | | | | | |
| 5 | | | X | | | | X | | | | | |
| 6 | | x | | | | | X | | | | | |
| 7 | | | X | | | | X | | | | | |
| 8 | | | X | | | | X | | | | | |
| 9 | | | X | | | | X | | | | | |
| 10 | | | X | | | | X | | | | | |
| 11 | | | | | | X | | X | | | | |
| 12 | | | | | | X | | X | | | | |
| 13 | | | | | | X | | X | | | | |
| 14 | | | | | | X | | X | | | | |
| 15 | | | | | | X | | X | | | | |
| 16 | | | | | | | | | x | | | |
| 17 | | | | | | | | | X | | | |
| 18 | | | | | | | | | X | | | |
| 19 | | | | | | | | | X | | | |
| 20 | | | | | | | | | X | | | |
| 21 | | | | | | | | | X | | | |
| 22 | | | | | | | | | | X | | |
| 23 | | | | | | | | | X | | | |
| 24 | | | | | | | | | X | | | |
| 25 | | | | | | | | | X | | | |
| 26 | | | | | | | | | | X | | |
| 27 | | | | | | | | | X | | | |
| 28 | | | | | | | | | X | | | |
| 29 | | | | | | | | | | X | | |
| 30 | | | | | | | | X | | | | |
| 31 | | | | | | | | | | | | X |
| 32 | | | | | | | | X | | | | |
| 33 | | | | | | | | | X | | | |
| 34 | | | | | | | | | X | | | |
| 35 | | | | | | | | X | | | | |
| 36 | | | | | | | | X | | | | |
| 37 | | | | | | | | X | | | | |
| 38 | | | | | | | | | X | | | |
| 39 | | | | | | | | | X | | | |
| 40 | | | | | | | | X | | | | |
| Total | | 1 | 9 | | | 5 | 10 | 11 | 15 | 3 | | 1 |
| Total mid= 15 | | | | | | Total Final=40 | | | | | | |

Key

R- reading

L- listening

V- vocabulary

S- speaking

G- grammar

W- writing

Appendix -5
Addis Ababa University
School of Graduate Studies
Faculty of Language Studies

Questionnaire for Teachers

The main purpose of this questionnaire is to gather data about the opinions of teachers towards the content validity of Grade 9 English Language classroom tests.

The out come of this questionnaire will directly or indirectly be successful with your genuine responses to each item. There fore you are kindly requested to give frank responses.

Thank you in advance for your cooperation!

Part I

General Information

1. The number of students(on average) in one class _____
2. The number of sections you teach _____
3. Which units do you cover for your first semester tests? _____
4. Have you taken any course about English language testing? _____

Part II About the content area of the Grade 9 English Language tests

Direction: Three rating scales (A_C) are given to assess each stated item regarding the content and formats of your English language tests. Please circle the letter that represent your appropriate scale.

Scales: A – Often B - Sometimes C - Never

Note: the word ‘test’ refers to both mid-term and final exams

1. Do you limit your tests to the materials covered in the class room?
A. Often B. Sometimes C. Never

If your answer for question number one is ‘B’ or ‘C’, what is your reason?

2. Do you think that your tests have adequate coverage of grammar test items ?
A. Often B. Sometimes C. Never
3. Do you think that your tests have adequate coverage of vocabulary test items?
A. Often B. Sometimes C. Never
4. Do you think that your tests have adequate coverage of reading test items ?
A. Often B. Sometimes C. Never
5. Do you think that your tests have adequate coverage of listening test items?
A. Often B. Sometimes C. Never
6. Do you think that your tests have adequate coverage of speaking test items?
A. Often B. Sometimes C. Never
7. Do you think that your tests have adequate coverage of writing test items?
A. Often B. Sometimes C. Never

If your answers from question number 2-7 are 'B' or 'C' what are your reasons?

8. Do your tests sample the language elements (Grammar, Vocabulary) and language skills(Reading, Listening, Speaking, Writing) in an appropriate proportion as they are covered in the text book?

- A. Often B. sometimes C. Never

If your answer for question number 8 is 'B' or 'C', what is your reason?

9. Do your tests employ a variety of both subjective and objective formats?

- A. Often B. sometimes C. Never

If your answer for question number 9 is 'B' or 'C', what is your reason?

10. Do you think that your tests adequately mirrored the test formats employed in the textbook?

- A. Often B. sometimes C. Never

If your answer for question number 10 is 'B' or 'C', what is your reason?

11. Do you teach all the major and sub-language skills?

A) Yes B) No

If you say No, what is your reason?

12. What is your guide line to decide the number of items in your tests for each language skills and language areas?

13. What is your guideline to decide on the type of format to use in the tests?

Part III: About major and sub-language skills

Please rank the following content areas according to the emphasis you give while teaching, and giving tests of English. Give '1' for the most emphasized content area, '2', for the next emphasized and so on.

1. The content area you give emphasis to when teaching English:

| | |
|-----------------|----------------|
| Reading_____ | Listening_____ |
| Vocabulary_____ | Speaking_____ |
| Grammar_____ | Writing_____ |

2. The content area you give emphasis while testing:

| | |
|-----------------|----------------|
| Reading_____ | Listening_____ |
| Vocabulary_____ | Speaking_____ |
| Grammar_____ | Writing_____ |

What are your reasons for the first three emphasized content areas for question number 1 and 2

Appendix -6
Addis Ababa University
School of Graduate Studies
Faculty of Language Studies

Questionnaire for Students

This questionnaire is intended to gather evidences on the views of students towards Grade 9 English language tests.

It is hoped that the results of this questionnaire will be successful on the basis of your genuine responses to each item. Therefore, you are kindly requested to give your frank opinions.

Part I General Information

Fill the following blank spaces with the necessary information.

1. Name of Your School _____
2. The Total Number of Students in Your Sections _____

Part II About the contents of the Grade 9 English language tests

Direction: Three rating scales (A-C) are given to assess each item regarding the contents of Grade 9 English Language tests. Please circle the letter with the appropriate scale for your choice.

Scales: A. Often B. Sometimes C. Never

Note: the word 'test' refers to both mid-term and final exams

1. Do you think that items found in the tests match with the text books covered in the classroom?
A. Often B. Sometimes C. Never.
2. Do you think that your tests have adequate coverage of grammar test items ?
A. Often B. Sometimes C. Never
3. Do you think that your tests have adequate coverage of vocabulary test items?
A. Often B. Sometimes C. Never
4. Do you think that your tests have adequate coverage of reading test items?
A. Often B. Sometimes C. Never

5. Do you think that your tests have adequate coverage of speaking test items?
A. Often B. Sometimes C. Never
6. Do you think that your tests have adequate coverage of listening test items?
A. Often B. Sometimes C. Never
7. Do you think that your tests have adequate coverage of writing test items?
A. Often B. Sometimes C. Never
8. Do you think that the language elements (Grammar, Vocabulary) and language skills(Reading, Listening, Speaking, Writing) are in an appropriate proportion in the tests as they are found in your text book?
A. Often B. Sometimes C. Never
9. Do your tests employ a variety of both subjective and objective formats?
A. Often B. sometimes C. Never
10. Do you think that your tests adequately mirrored the test formats employed in the textbook?
A. Often B. sometimes C. Never

Part III About the major and sub- language skills

Direction: Please rank the following content areas according to the emphasis your teacher gives when he/she teaches and gives test of English. Give '1' to the most emphasized content area and '2' to the next most emphasized content area and so on

1. The content area your teacher gives emphasis to when teaching English:

| | |
|-----------------|----------------|
| Grammar_____ | Listening_____ |
| Vocabulary_____ | Speaking_____ |
| Reading_____ | Writing_____ |

2. The content area your teacher gives emphasis to in tests:

| | |
|-----------------|----------------|
| Grammar_____ | Listening_____ |
| Vocabulary_____ | Speaking_____ |
| Reading_____ | Writing_____ |

Appendix-7

አዲስ አበባ ዩኒቨርሲቲ

የድህረ ምረቃ ትምህርት መርሃ ግብር

የውጭ ቋንቋዎች ጥናት ክፍል

የተማሪዎች የጽሁፍ መጠይቅ

የዚህ መጠየቅ ዋና ዓላማ የ9ኛ ክፍል የእንግሊዝኛ ቋንቋ ፈተናዎችን በተመለከተ ተማሪዎች ስላላቸው አመለካከት መረጃዎችን ለመሰብሰብ ታስቦ የተዘጋጀ ነው።

ከመጠይቁ የሚገኘው ውጤት ሊደረግ ለታሰበው ጥናት የሚኖረውን ጠቀሜታ ግምት ውስጥ በማስገባት በመጠይቁ ውስጥ ለተካተቱት ጥያቄዎች ትክክለኛ ምላሻችሁን ትሰጡ ዘንድ በትህትና ትጠየቃላችሁ።

አመሰግናለሁ

ክፍል I. አጠቃላይ መረጃ

1. የት/ቤቱ ስም _____
2. በክፍል ውስጥ ያሉት ተማሪዎች ብዛት _____

ክፍል II. የ9ኛ ክፍል የእንግሊዝኛ ቋንቋ ፈተናዎችን በተመለከተ

መመሪያ:- የ9ኛ ክፍል የእንግሊዝኛ ቋንቋ ፈተናዎችን ይዘትና አቀራረብ በተመለከተ ሶስት መመዘኛዎች (A-C) ለእያንዳንዱ ጥያቄ አይነት ተመድቧል። ስለ ክፍል ፈተናዎች ትክክለኛ መስሎ የሚታያችሁን ምርጫ በመክበብ መልሱ።

1. በፈተናዎች ውስጥ የሚገኙት ጥያቄዎች ከ9ኛ ክፍል የእንግሊዝኛ መማሪያ መጽሀፍ /በክፍል ውስጥ ከሚሰጠው ተምህርት ጋር ይዛመዳል?

| | | |
|----------|-------------|-----------|
| A) ብዙ ጊዜ | B) አንዳንድ ጊዜ | C) አይዛመድም |
|----------|-------------|-----------|
2. የክፍል ፈተናዎች የሰዋሰው እውቀትን የሚገመገሙ ጥያቄዎችን በሚገባ ያካትታሉ?

| | | |
|----------|-------------|-----------|
| A) ብዙ ጊዜ | B) አንዳንድ ጊዜ | C) አያካትቱም |
|----------|-------------|-----------|
3. የክፍል ፈተናዎች የቃላት እውቀትን የሚገመገሙ ጥያቄዎችን በሚገባ ያካትታሉ?

| | | |
|----------|-------------|-----------|
| A) ብዙ ጊዜ | B) አንዳንድ ጊዜ | C) አያካትቱም |
|----------|-------------|-----------|
4. የክፍል ፈተናዎች የንባብ ክህሎትን የሚገመገሙ ጥያቄዎችን በሚገባ ያካትታሉ?

| | | |
|----------|-------------|-----------|
| A) ብዙ ጊዜ | B) አንዳንድ ጊዜ | C) አያካትቱም |
|----------|-------------|-----------|

5. የክፍል ፈተኞች የማዳመጥ ክህሎትን የሚገመገሙ ጥያቄዎችን በሚገባ ያካትታሉ?
 - A) ብዙ ጊዜ
 - B) አንዳንድ ጊዜ
 - C) አያካትቱም
6. የክፍል ፈተኞች የመጻፍ ክህሎትን የሚገመገሙ ጥያቄዎችን በሚገባ ያካትታሉ?
 - A) ብዙ ጊዜ
 - B) አንዳንድ ጊዜ
 - C) አያካትቱም
7. የክፍል ፈተኞች የመናገር ክህሎትን የሚገመገሙ ጥያቄዎችን በሚገባ ያካትታሉ?
 - A) ብዙ ጊዜ
 - B) አንዳንድ ጊዜ
 - C) አያካትቱም
8. የፈተናዎቹ ጥያቄዎች የቋንቋ ክፍሎችን /ሰዋሰው፣ ቃላት/ እና የቋንቋ ክህሎችን /ማንበብ፣ ማዳመጥ፣ መናገርና መጻፍ/ ከማመሪያ መጽሐፋችሁ ላይ በተሰጠው ትኩረት መሰረት በተመጣጠነ መልኩ አካተዋል?
 - A) ብዙ ጊዜ
 - B) አንዳንድ ጊዜ
 - C) አያካትቱም
9. የክፍል ፈተናዎች የተለያዩ የአጠያየቅ ስልት /እውነት/ ሀሰት፣ አዛምድ፣ ምርጫ ወዘተ/ ይጠቀማሉ?
 - A) ብዙ ጊዜ
 - B) አንዳንድ ጊዜ
 - C) አይጠቀሙም
10. በ9ኛ ክፍል የእንግሊዝኛ መጽሐፋችሁ ወይም በክፍል ውስጥ ልምምድ የተደረገባቸው የአጠያየቅ ስልቶች /እውነት/ሀሰት፣ አዛምድ፣ ምርጫ ወዘተ/ በፈተናዎቹ ውስጥ በተመጣጠነ መልኩ ተካተዋል?
 - A) ብዙ ጊዜ
 - B) አንዳንድ ጊዜ
 - C) አያካትቱም

ክፍል III. የተለያዩ የእንግሊዝኛ ቋንቋ ክፍሎችንና ክህሎችን የሚመለከት

መመሪያ:- በእንግሊዝኛ ቋንቋ ትምህርት ክፍል ጊዜ መምህሩ በሚያስተምርበት እና በሚፈትንበት ወቅት የበለጠ ትኩረት ለሚሰጣቸው የቋንቋ ክፍሎችና ክህሎቶች እንደቅደም ተከተላቸው አስቀምጥ/ጪ። ትኩረት ለሚሰጠው የቋንቋ ክፍል ወይም ክህሎት “1” ቀጥሎ ትኩረት ለሚሰጠው “2” ወዘተ በማለት ባለው ክፍት ቦታ ቁጥሮችን ከ1-6 አስቀምጡ።

1. መምህሩ የእንግሊዝኛ ትምህርትን በሚያቀርብበት/ በሚያስተምረበት ጊዜ የበለጠ ትኩረት የሚሰጠው:-


| | |
|-----------|-----------|
| ሰዋሰው----- | ማዳመጥ----- |
| ቃላት----- | ንግግር----- |
| ንባብ----- | ጽሁፍ----- |

2. መምህሩ የእንግሊዝኛ ቋንቋ ትምህርት ፈተናዎችን በሚሰጥበት ወቅት ብዙ ጥያቄዎች የሚያተኩሩት:-

| | |
|-----------|-----------|
| ሰዋሰው----- | ማዳመጥ----- |
| ቃላት----- | ንግግር----- |
| ንባብ----- | ጽሁፍ----- |

Declaration

I, the undersigned, declare that this thesis is my original work and has not been presented for a degree in any other university, and all the sources of materials used for the thesis have been duly acknowledged.

| | |
|--------------------|---|
| Name | Yeshambel Almaw Bitew |
| Signature |  |
| Place | Faculty of Language Studies |
| Date of Submission | May, 2011 |