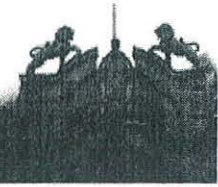


Addis Ababa
University
(Since 1950)



**ASSESSING CONTENT VALIDITY OF PRIMARY
EDUCATION COMPLETION ENGLISH
EXAMINATION OF SNNPRS:
2002 E.C IN FOCUS**

**BY
SAMUEL DOSSA**



**MAY 2011
ADDIS ABABA**

**ASSESSING CONTENT VALIDITY OF PRIMARY EDUCATION
COMPLETION ENGLISH
EXAMINATION OF SNNPRS:
2002 E.C IN FOCUS**

**BY
SAMUEL DOSSA**

**A THESIS SUBMITTED TO THE SCHOOL OF GRADUATE
STUDIES IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF MASTER
OF ARTS IN TEACHING ENGLISH AS A
FOREIGN LANGUAGE (TEFL)**

**MAY 2011
ADDIS ABABA**

**ADDIS ABABA UNIVERSITY
SCHOOL OF GRADUATE STUDIES
DEPARTEMENT OF ENGLISH**

**ASSESSING CONTENT VALIDITY OF PRTIMARY
EDUCATION COMPLETION ENGLISH
EXAMINATION OF SNNPRS:
2002 E.C. IN FOCUS**

**BY
SAMUEL DOSSA**

Approved by


Advisor


Signature


Examiner


Signature

Department head

Signature

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to my advisor, Dr. Abebe G/Tsadik for the close guidance and encouragement he has given me throughout the period of this study.

My appreciation also goes to the test constructors and the director of Teaching-Learning Ass/Core Process in SNNPRSEB for their cooperation by responding to the interview questions and making the necessary materials available for use in this study.

Great thanks also go to my classmates Sirak Mekonnen and Tariku Anega for their appreciable comments and technical support on my work. I am also grateful to Dr. Silassie Cheru for proof reading the first draft.

My indebtedness goes to my brother Zeleke Dossa and my friend Temesgen Engida for their financial support and moral encouragement throughout my graduate study.

Last but not least, I am also grateful to W/ro Haymanot Tekalgn and Rahel Tekalgn for typing and printing the manuscript.

TABLE OF CONTENTS

	Page
Acknowledgements -----	i
Table of Contents -----	ii
List of Tables -----	iv
Abstract -----	v
CHAPTER ONE	
Introduction -----	1
1.1. Background -----	1
1.2. Statement of the Problem -----	4
1.3. Objectives of the Study -----	6
1.3.1. General Objective -----	6
1.3.2. Specific Objectives -----	6
1.4. Significance of the Study -----	6
1.5. Delimitation of the Study -----	7
1.6. Definition of Terms and Acronyms -----	7
CHAPTER TWO	
Review of Related Literature -----	8
2.1 Distinction between Concepts: Test, Assessment and Examination -----	8
2.2 Test Purposes and Test Types -----	10
2.2.1 Achievement Test -----	11
2.2.2 Diagnostic Tests -----	12
2.2.3 Progress Test -----	13
2.2.4 Placement Tests -----	13
2.2.5 Proficiency Tests -----	13
2.2.6 Aptitude Tests -----	14
2.3 Relation of Language Testing to Language Teaching -----	14

2.4 Qualities of a Good Test -----	17
2.4.1. Validity -----	17
2.4.1.1 Content Validity -----	18
2.4.1.2 Criterion Related Validity -----	21
2.4.1.3 Face Validity -----	23
2.4.2 Reliability -----	23
2.4.3 Practicality -----	25
2.4.4 Comprehensiveness -----	25
CHAPTER THREE	
Methodology and Procedure -----	27
3.1 A General Description of Grades 7 and 8 English Textbooks and the 2002 E .C English Exam -----	27
3.2 The Data Sources -----	28
3.3 Data Collecting Procedures -----	28
CHAPTER FOUR	
Data Analysis, Results and Discussions -----	31
4.1. Textbook Analyses -----	31
4.2. Analyses of the 2002 E.C Primary Education Completion English Examination -----	35
4.3 Analysis of Interview Questions -----	42
4.4 Discussion -----	43
CHAPTER FIVE	
Summary of the Findings, Conclusions and Recommendations -----	46
5.1. Summary of the Findings -----	46
5.2. Conclusions -----	47
5.3. Recommendations -----	48
Bibliography -----	49
Appendices	

LIST OF TABLES

	Page
Table 4.1: The Six Major Language Content Areas of the Grades 7 and 8 English Textbooks and the Number of Periods Allotted to them -----	32
Table 4.2: Content Areas of the Grades 7 and 8 English Textbooks and their Frequency of items of Practice -----	34
Table 4.3: Content Areas of the exam and their Frequencies of Items -----	35
Table 4.4: Summary of the major content areas of the exam and their frequency of items -----	37
Table 4.5: The Comparison Between total Frequencies of content areas of grades 7 and 8 English Textbooks and the 2002 E.C Primary Education Completion English Examination -----	40
Table 4.6: Extent of Relationship between the Contents of Grades 7 and 8 English Textbooks and the 2002 E.C Primary Education Completion English Examination -----	41

ABSTRACT

The purpose of this study was to assess whether or not 'content validity' had been achieved between the language contents of the Grades 7 and 8 English Textbooks and the 2002 E.C Primary Education Completion English Examination of South Nations, Nationalities and Peoples Regional State (SNNPRS). To this effect, the study set out specifically to (1) determine whether the 2002 E.C English examination comprises a reasonable representative samples of the language content areas of the two textbooks. (2) find out the degree of the extent of Textbook-Examination relationship (3) examine the allocation of percentages to the content areas of the two textbooks.

Descriptive study was used for it explains the present situation of the problem and attempts to measure the problem exists without questioning why it exists. The major data gathering tool employed in this study was 'content analyses. It was assumed to help get a brief summary of the contents of the textbooks and the exam. This tool was substantiated by unstructured interview which was held with the test constructors and the director of Teaching-Learning Ass/Core Process in South Nations, Nationalities and Peoples Regional State Education Bureau (SNNPRSEB).

The content validity analysis of the contents of the textbooks and the exam was done by employing different steps. First, contents of the exam and the textbooks were classified into major language content areas. Then, these content areas were analyzed and presented in a summary table, showing the frequencies of the content areas both in figures and percentages. After this, the extent of relationship between frequencies from the analyses of the textbooks and frequencies from the analyses of the examination questions has been determined using the Sidney Siegel's formula.

The results of the study showed that the degree of the extent of relationship between the textbook coverage and the examination content areas was found to be low since the correlation coefficient between the two observations reads 0.21. The allocation of percentages to content areas was also characterized by a disproportionate distribution. Besides, the match between the content of the textbooks and the exam was also distorted because of the inclusion of some tasks which did not appear in the textbooks. On the basis of the results, it was concluded that the exam lacked content validity.

CHAPTER ONE

INTRODUCTION

1.1. Background

It is a fact that language testing plays an important role in the entire process of language teaching and learning. Some of its uses, according to Hughes (1989), read as follows:

- *to measure language proficiency regardless of any language courses that candidates may have followed*
- *to discover how far we have achieved the objectives of a course of study*
- *to diagnose students' strengths and weaknesses, to identify what they know and what they do not know*
- *to assist placement of students by identifying the stage or part of a teaching program most appropriate to their ability (p.7).*

All of the above points summarized by Hughes refer to the benefits that can be gained by employing different types of language tests. Confirming this, Weir (1993) clearly states the fact that testing aids teaching-learning in a way that helps teachers encourage their students and assists teachers to monitor their students' progress in learning, etc. He further elaborates that the benefit which we get from test is not only limited to the classroom teacher. It has also a great role in the improvement of curriculum, provided that offering the necessary information about the implementation of the instructional material, the development of curriculum objectives, etc. Generally, scholars like Heaton (1982), Bachman (1990), Alderson (1991), Teshome (1995) and many others confirm that it is true that a test has a great role in the teaching-learning process.

However, perhaps the most important question to be asked is-What should a test possess to provide such an aid to the entire teaching-learning process? Indeed, it is undeniable that a test should possess the right quality and standard in order to be able to accomplish the above seemingly formidable tasks. To this effect, Harrison (19830) asserts that it is unthinkable to prepare

and to use a foreign language test without considering its different important qualities like validity, reliability, practicality and others. He further explains:

... these abstract nouns may seem rather daunting, but Justas it is impossible to play chess without knowing how a knight moves across the board, so it is pointless to try to write tests without a basic understanding of the principles behind them.

Similarly, many scholars, (for example, Lado 1961; Madsen 1983; Heaton 1988; Baker 1989; Bachman 1990) support the fact that test preparation should give a considerable attention to these qualities. Therefore, it could be deduced that it is a necessity rather than a choice to consider these qualities in test preparation.

It is, of course, inevitable that all of the four principal qualities (validity, reliability, comprehensiveness and practicality) contribute to the good quality of a given test. Nevertheless, the main focus of the present study is only limited to validity, especially content validity which is one of the other types of validity, such as face validity, criterion-related validity and predictive validity.

According to Hughes (1989), a test is said to have content validity when its content comprises a reasonable representative sample of language skills, structures, etc. with which it is supposed to cover. In the assessment of what has been learned, due concern is given to content validity as it is the most important aspect in language test development. In line with this, studies conducted by Oller (1979), Harrison (1983), Hughes (1989), Bachman (1990), Weir (1990), for example, assert that content validity is one of the important features of a test.

Moreover, Grieve (1964: 23) (in Alemu 1983) has this to say when elaborating the need of content validity in test preparation:

The important question to be asked about any examination is: How well does it measure what it is supposed to measure? This question can be effectively answered only in terms of two basic criteria: validity and variability... In a language testing, validity depends on the linguistic content of the test... The first essential of such validation is a detailed study of the test material. Such a study will establish whether a test contains a representative sample of the language problems likely to be encountered by the group of students taking the test. It is in this way the content validity of language test is established (p.7)

The point made by Grieve is that a test must be representative of the range of knowledge (language areas in our case) as outlined in the course. All what has been stated tells us that a test must cover an adequate and representative section of those areas it is desired to test. That is, if a test is designed to measure the content of a particular course of study, we should expect the test to be based upon a careful analysis of the outline of the course, and we should further expect the items to represent adequately each portion of the analysis of outline.

Realizing this fact, prior to the research discussed herein, studies had been conducted on testing areas in local contexts. Some of these are Asmare (2008); Aragaw (2004); Kifle (1995); Teshome (1995); Nuru (1992); Tibebe (1992); Alemu (1983). A report by Tibebe Alemayehu on the predictive power of ESLCE English in 1987/88 compared to two locally designed integrative language tests stated that the predictive validity of ESLCE in 1987/88 was found to be weak (cited in Asmare (2008). Moreover, the findings of almost all of the studies show that the examinations lack content validity at different levels.

1.2 Statement of the Problem

As explained in background section, a test should possess different principal qualities of a good test in order to play its role on teaching and learning process. It was also discussed that one of these principal qualities is content validity; that is to say that whether or not the test's content constitutes a representative sample of the language skills, structures, etc. with which it is meant to be concerned. In this section an attempt was made to highlight the consequences of lack of content validity. In other words, this section tried to show the danger otherwise.

In line with this, Hughes (1989) asserts that if a test lacks content validity, such a test would have a harmful backwash effect on the entire process of teaching and learning. He further elaborates that the consequences with the imbalance is many fold. For example, it is obvious that students would pay attention to only the content areas addressed in the test during the classroom instruction. It is also true that the lion share of the preparation for the test will be restricted to those areas which frequently appear in the test. Besides, it could be generalized that teaching and learning are likely to concentrate on what can be predicted. However, the reverse is true if the test's content constitutes a representative sample of the language skills, structures, etc. without making any bias. And it is obvious that such a test will result in beneficial backwash effect on the entire process of teaching and learning (Ibid).

Similarly, the consequence of lack of content validity in a given test is better summarized by Cumming and Berwick (1996) (in Argaw 2004). It reads as:

If a test lacks content validity: teachers would use methodology that would help them to prepare their students for the test and, finally, teachers would write tests which would mirror the contents of past examination.

Research in local contexts has determined that the problem of lack of content validity still prevails both in the national examinations and high school English language tests (Dejenie 1990; Atkins et al. 1996; Argaw 2004; Alemu 1983; Teshome 1995 and Kifle 1995). Of all previous studies mentioned above, the

study which was directly related to the present study is Alemu (1983) which dealt with an assessment of content validity of grades six and eight English National examination. However, the researcher felt that the issue still deserves further investigation because of two main reasons. First, the previous study was conducted only with references to the old curriculum and the former English textbooks which were in use until June 1982. Second, since then the grades six and eight examinations were prepared at national level rather than regional level. Hence, it is a fact that the content validity of the current Primary Education Completion English Examination prepared at regional level has not undergone observation to determine whether or not there is content validity between the examination and the new textbooks.

Moreover, in their study about the promotion and repetition rate of Primary Education Completion Examination of eight consecutive years (1995 E.C – 2002 E.C) in SNNPRS, researchers found that the repetition rate is increasing dramatically regardless of the lowering of the passing mark from 38 to 36. Besides, it has been remarked that the repetition rate and the declining of the candidates' result became worse in the 2002 E.C examination than any other years (Gebrehiwot et al., 2010). Why was that so? Why were the examination questions become so difficult for the candidates? So far, there has not been a systematic follow-up study to answer this question.

With this concern in mind, the present study aimed to determine whether or not there is a mismatch between the contents of the 2002 E.C Primary Education Completion English Examination in relation to the contents of the grades 7 and 8 English textbooks and thereby determining the degree of the extent of Textbook- Examination relationship, and examine the allocation of percentages to content areas so as to investigate the exam's content validity.

In doing so, the study tried to answer the following questions:

1. Are the contents of the two textbooks represented adequately and proportionally in the 2002 E.C English Examination?

If 'Yes', what is the extent of relationship?

2. How are the content areas of the textbooks allocated in terms of percentages?

1.3 Objectives of the Study

1.3.1 General Objective

The general objective of this study was to have some insights on whether or not the content validity is achieved between the 2002 E.C Primary Education Completion English Examination and Grades 7 and 8 English textbooks.

1.3.2 Specific Objectives

The specific objectives of the study were:

1. To determine whether the 2002 E.C English examination comprises reasonable representative samples of the language content areas of the two textbooks
2. To find out the degree of the extent of Textbook-Examination relationship
3. To examine the allocation of percentages to the content areas of the two textbooks

1.4 Significance of the Study

As was indicated in statement of the problem, the main contribution of the study is to try to answer partly the questions raised. By analyzing the content validity of the 2002 E.C English examination vis-à-vis the grades 7 and 8 textbooks, the present study was hoped to contribute some information to examination constructors, syllabus designers and textbook writers to modify, supplement or change the contents of the textbooks, syllabus and examination. Moreover, the results of the study would help test constructors to see whether or not the content areas to be tested are fairly represented in the test's content by examining the spread of the frequencies (in terms of percentages) among the content areas. Furthermore, it would be used as a source material for further related studies.

1.5 Delimitation of the Study

From the different validity concepts like face validity, content validity and criterion-related validity, the present study was confined to content validity. As to the examination years, the study was limited to the 2002 E.C academic year. Besides, the content analysis was made only on the Grades 7 and 8 English textbooks rather than including the content analysis of the syllabus of these Grades and the teachers' guides. The study was delimited in this way due to time and financial constraints.

1.6 Definition of Terms and Acronyms

1. Item analysis- is a study to determine how efficiently the individual items (questions) function.
2. Frequency- refers to the number of items a content area is dealt within the textbooks or examination contents.
3. Exercise – is an activity designed to teach content area through instructional words such as look and read, write, read and answer, change, complete, fill in, etc.
4. Content analysis- is a scheme in the form of a table marked vertically and horizontally to determine frequencies of content areas.
5. Backwash- refers to the effect of the testing on teaching and learning (Hughes, 1989: 1).
6. Harmful Backwash- the negative effect of the test on the teaching and learning.
7. Beneficial backwash- the positive effect of the test on the teaching and learning process.
8. Primary Education Completion Examination- refers to the regional examination given at the end of grade eight.
9. SNNPRS- stands for South Nations, Nationalities and Peoples Regional State.
10. SNNPRSEB- stands for South Nations, Nationalities and Peoples Regional State Education Bureau.

CHAPTER TWO

REVIEW OF RELATED LITERATURE

As stated in the first chapter, this study was planned to gain some insight into whether or ^{not} content validity was achieved between the 2002 E.C. Primary Education Completion English Examination and Grades 7 and 8 English textbooks. What follows is some of the reviewed literature which is relevant to the present study; it was hoped to develop the story line of this study.

2.1 The Distinction between Concepts-Test, Assessment and Examination

It seems that test, assessment and examination are difficult concepts to define and distinguish from each other. Indeed, there are people who have problems in defining these terms and making distinction among them as they are used in educational measurement and evaluation. Hence, it seems that understanding the difference between these terms is fundamental to avoid the kind of misleading overlap in their definition.

Thorndike (1995:8) defines, "Test is a tool or instrument used to gather information about the quality trait being measured." He further illustrate that test is a kind of holistic term which encompasses any assessment procedure used to systematically measure a sample of behavior provided and could be administered in a setting such as a classroom, laboratory etc. with standardized (uniform) procedures for administration and marking. It could be prepared either by the classroom teacher or any public (external) body and called classroom achievement test and standardized test respectively (Ibid).

Hence, it should be clear that the focus of this study is limited to the second type of test since the study deals with Primary Education Completion English Examination which is relatively one of the standardized tests produced by external bodies other than the classroom teacher.

Assessment takes slightly a different meaning than test according to some people. Griffin and Nix in Mulu and Daniel (2005:6), for example, define, "Assessment is regarded as the process of describing student progress, such as whether teaching objectives have been met. Whether implicit or explicit, assessment is most usefully connected to some goal or objective for which the assessment is designed." These scholars also note that assessment is like reporting for it deals with collecting, analyzing, recording and communicating information about the outcomes of learning and teaching.

Moreover, Bloom (1956) defines assessment as follows:

Assessment refers to the procedures used in appraising students learning. ... is concerned with the process involved in describing and reporting on the outcomes of learning and teaching. Assessment can be both qualitative and quantitative and may occur at any time during or at the end of a unit of instruction. It can cover a broad of learning outcomes- cognitive, affective, social- from the simplest to the most complex.

In short, from the above definitions of assessment it could be generalized that assessment is a process or a procedure of obtaining information about the progress of students and thereby evaluating the outcome of the teaching and learning process. It is also obvious that the process referred to are connected with collecting and explaining information that is related to learning and teaching outcomes. Yet, the question that seems to be asked here is: 'How/By what means do we collect this information?' The answer to this question is also obvious. It is through testing and other methods of assessment such as observation, rating scale and so on. Therefore, relatively speaking, it has become clear that assessment is wider, and can comprise testing. The next point is about examination.

In fact, it has been tradition to use the terms test and examination interchangeably. Oxford Dictionary of English (2005: 602) defines, "Examination is a formal test of person's knowledge or proficiency in a subject

or skill.” For Baker (1989), examinations are tests produced by public bodies. What is common to these two definitions of examination is the issue of being formal. Thus, it could be possible to say that what makes a given test an examination depends on the question whether that test is formal or not. That is, whether the tests are produced in the informal classroom situation by the teacher targeting his/her students or are they produced in the formal and standardized way by public bodies targeting a wider enough audience.

In summary, in view of what has been said about the terms test, assessment and examination, it seems that there exists a slight difference among them. Assessment is a process or procedure of obtaining information about progress of students and teaching and learning outcomes through various tools; one of the various tools used is test. It has also been said that examination is simply a formal test. Having given background to these concepts, now let us see some of the test purposes and test types.

2.2 Test Purposes and Test Types

The categories of tests and the names given to them vary considerably according to the preferences of different writers. The source of the variation is the purpose of the given test. That is, we choose a given test type if and only if we believe that it can provide us with the information we need from it.

For instance, if our purpose is to assign students to classes at different levels, then, we choose a placement test; whereas if our intention is to measure the progress that students are making, we need a progress test and so on. The issue of categorizing tests based on their purposes (the information they offer) is better summarized by Hughes (1989). It reads as:

We use tests to obtain information. The information that we hope to obtain will of course vary from situation to situation. It is possible, nevertheless, to categorize tests according to small number of kinds of information being sought. This categorization will prove useful both in deciding whether an existing test is suitable for a particular purpose and in writing appropriate new tests where these are necessary (p.9).

Though the intention here is not to give a complete list and explanation of different kinds of tests, let us see some of these tests and their purposes. The six types of tests that will be described in this section are: diagnostic, achievement, placement, progress, proficiency and aptitude tests. Of these types of tests all but achievement test is the primary concern of this study since it encompasses tests like Primary Education Completion English Examination.

2.2.1 Achievement Test

Achievement test (also called an attainment or summative test) is a type of test which covers what has been learned across a year, or a whole course(s) unlike to that of other types of tests such as diagnostic test. It can indicate whether or not students have reached a certain standard as compared to other students in the same level. The standard here could refer to international (e.g., Cambridge in EFL); or national (e.g., school leaving certificates); or it could also refer to different schools which certify students for following some courses (Harrison, 1983).

Similarly, for Hughes (1989) there is direct relation between achievement tests and language courses. They are designed primarily to evaluate the extent of the success of individual students, group of students, or the courses themselves in line with the objectives set before.

He adds that:

Final achievement tests are those administered at the end of a course of study. They may be written and administered by ministries of education, official examining boards, or by members of the testing institutions. Clearly the content of these tests must be related to the courses with which they are concerned (pp. 10-11, Ibid).

Besides, according to Allen and Davies (1977), an achievement test is a test which evaluates the knowledge gained from a given known program of instruction syllabus.

The important point which is common to all the above scholars' definitions of an achievement test is that it is (1) more formal than tests like diagnostic and progressive; (2) wider in its scope; that is, it covers a much wider range of material than others and relates to long-term rather than short-term objectives. This brings up problems of sampling, since what has been learned in a year or in a number of years (for example) cannot all be assessed in one day, yet the test must reflect the content of the whole course. Decisions therefore have to be made about what should be included in the test, and whether assessing one thing can be assumed to include another.

In the same token, since Primary Education Completion English Examination is the one which shares these characteristics of an achievement tests, it should be designed in a way that it properly covers and samples the course syllabus, the textbooks and the other materials used.

Therefore, with this concern in mind, the present study is intended to evaluate whether or not the 2002 E.C Primary Education Completion English Examination proportionally samples the contents of Grades 7 and 8 textbooks in keeping the content validity.

2.2.2 Diagnostic Tests

According to Hughes (1989), diagnostic tests are tests through which we try to get information about our students' problem area so as to ascertain further teaching.

Harisson (1983: 6) states, "A diagnostic test (sometimes called a formative or progress test) checks on students' progress in learning particular elements of the course. He adds that this type of test mainly rely on providing information about the progress which helps to offer an immediate remedial work. Besides,

Dejenie (1994:6) states the characteristic features of diagnostic tests as follows:

- *Identifies areas of difficulty, will be remedied later,*
- *Often a feature of progress test,*
- *Rather limited, e.g. to grammatical items,*
- *More relevant for whole group than individuals*

2.2.3 Progress Test

Progress tests, as their name suggests, are intended to measure the progress that students are making. In other words, these are short, regular classroom tests which enable teachers to find out about students progress how well they have mastered the language skills and items teacher taught. Indeed, they are the most important kinds of tests for teachers.

However, like achievement tests, progress tests are based on syllabus and objectives of courses. This is why they are also called progress achievement tests (Hughes, 1989).

2.2.4 Placement Tests

Harrison (1983: 4) explains, "A placement test is designed to sort new students into teaching groups, so that they can start a course at approximately the same level as the other students in the class. It is concerned with the student's present standing, and so relates to general ability rather than specific points of learning." Hughes (1989) states that placement tests, as their name suggests, play great role in yielding information which will help to place learners to the level which deserves their ability in a given teaching programme.

2.2.5 Proficiency Tests

Dejenie Leta (1994: 10) summarizes and defines proficiency tests in this outlined manner:

- *Tests predicted language proficiency needed to perform a certain task or follow a certain educational programme (e.g. IELTS) or employment,*
- *Measure skills rather than attainments,*
- *Not designed in relation to a particular previous teaching programme,*
- *Looks FORWARD.*

In the same fashion, Hughes (1989: 9) explains, "Proficiency tests that are designed to measure people's ability in a language regardless of any training they may have had in that language. The content of a proficiency test, therefore,

is not based on the content or objectives of language courses which people taking the test may have followed.”

Harrison (1983) also shares the above definitions of proficiency tests. He explains that the aim of these tests are to evaluate an individual based on the specification of what he/she is expected to perform in the future rather than based any particular course content delivered so far. In other words, they target to assess that whether or not the individual can cope with the future real life situation as the result of his/her past learning.

2.2.6 Aptitude Tests

These are tests meant to predict (guess) candidate’s success or failure to learn a foreign language; that is, they help to obtain information about how well a candidate will do well in learning a foreign language. With regard to this, Lado (1961 :370) describes, “Aptitude tests can be used to foreign language to study in the country where the language is spoken.”

Therefore, in the view of this future orientation, aptitude tests deal with the probable future performance (future potential) of candidates, testing personality attitudes the psychic aspect of the learner) more than language learning facts. What is more, based on the information in this way, aptitude testing might be used to sort students into fast and slow sections according to their proficiency and ability to learn.

2.3 Relation of Language Testing to Language Teaching

The linkage between language testing and language teaching is undeniable though different scholars perceive their relationship differently. For instance, Davies (1968: 5) believes that ‘the good test is an obedient servant since it follows and apes teaching.’ For Hughes (1989: 2), however, testing cannot be considered only to allow teaching affect bad teaching positively.

Although there seems to be slight difference in the opinions of the above two scholars, it could be inferred that they share one thing in common; that is, the fact that one depends on the other and thereby often influencing each other.

Similarly, Heaton (1988: 5) explains, "Both teaching and testing are so closely interrelated that it is virtually impossible to work in their either field without being constantly concerned with the other." He further elaborates that tests could be designed basically as a tool to enhance the learning to encourage the learner or as a way of measuring the learner's knowledge of the language. The test is geared/ strictly connected to/ the teaching that has taken place in the first case, but the reverse is true in the second case (ibid).

Strengthening the above point, Atkins et al. (1996) state that where there is good teaching there is good testing. In other words, it is to say that the two, good teaching and good testing must always go together.

Moreover, regarding the positive influence of testing on the entire process of teaching and learning activity, Venkatswaran (1995) in Asmare (2008) has this to say when elaborating the information that students gain from tests. According to him a test helps students to develop positive attitudes towards their class, master the language and foster learning by their diagnostic nature.

The benefit that students gain from a test is also clearly stated by Madsen (1983) as follows:

properly made English tests can help create positive attitudes towards instruction by giving students a sense of accomplishment a feeling that the teachers evaluation of them matches to what he has taught them ... tests also help students learn the language by requiring them to study hard, emphasizing course objectives and showing them where they need to improve.

According to him, teachers are also benefited from testing in a way that it helps them to have information about the effect of their teaching-learning process. In other words, it helps them to check whether or not their students have

achieved the entire lesson they delivered and thereby evaluating the effectiveness of their teaching. The following are some of the questions that teachers often ask in their frequent evaluation of the effectiveness of their teaching and learning process:

- *Are my lessons on the right level?*
- *Am I aiming my instruction to low or to high?*
- *Am I teaching some skills effectively?*
- *Which areas need reviewing?*
- *Should I spend more or less time on the materials of mine?*
- *Etc, (Ibid).*

The above descriptions mainly rely on the benefits that the classroom (teacher made) tests yield for both teachers and students. In addition to this, tests, especially external (public) examinations such as those set by National Organization for Examinations (NOEs) office (e.g. Ethiopian General School Examination Certificate (EGSEC) of grade ten and primary education completion examination etc), give information not only for the classroom teacher and student, but also they could influence the entire teaching and learning process and the syllabus, providing necessary information to the syllabus designers and concerned bodies. In this regard, Heaton (1988: 6) asserts that although there exist kinds of limitations, public examinations often exert a paramount influence on the syllabus and the teaching strategies.

As the result of such a strong relationship between teaching and testing, it has been said that both classroom tests and external (public) examinations have been helping students, teachers, syllabus designers and concerned bodies. Harrison (1983: 1) claims, "Far from being divorced from each other, testing and teaching are closely interrelated. 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 summary, all what has been said above tries to show that teaching and testing are interwoven and interdependent that it is difficult to treat them differently. And an attempt is also made to highlight the benefit can be gain

from both classroom tests (teacher made tests) and external examinations in the teaching and learning process.

However, Hughes (1989: 1) argues "Too often language tests have a harmful effect on teaching and learning; and too often they fail to measure accurately whatever it is they are intended to measure." Hence, it is a fact that the positive effect (benefit) can only be gained from a test when the test possesses the fundamental qualities required in good language testing of all kinds.

2.4 Qualities of a Good Test

In the previous section we have been dealing with the relation of language testing to language teaching. We have also seen that as the result of such a close relationship between the two, they have been influencing each other. The influence (backwash effect) could be either positive or negative. It is, of course, undeniable that nobody wants to see the harmful backwash effect of tests upon the teaching/learning process. So, this seeks the need of important qualities of tests in preparing any language test. There are four principal qualities which are often considered when developing and judging formal foreign language tests. These are validity, reliability, practicality and comprehensiveness.

2.4.1. Validity

Validity is a concept in testing that deals with how much the test(s) satisfactorily test what it is supposed to test (Atkins et al., 1996). Two questions must be considered when determining validity of a foreign language test, according to them, are:

- *What aspect of the language is it designed to measure?*
- *How well does it, in fact, measure the global skills or the discrete elements of grammar and vocabulary?*

There are different validity concepts. Some of these are content validity, criterion-related validity, face validity; each of these tries to examine validity of a test differently.

2.4.1.1 Content Validity

Different scholars and educators define content validity in different ways. But all definitions, if critically viewed, reflect the same essence. Perhaps the easiest place to start in the discussion of its definitions is the following statement by Heaton:

This kind of validity depends on a careful analysis of the language being tested and of the particular course objectives. The tests should be constructed as to contain a representative sample of the course, the relationship between the test items and the course objectives always being apparent (Heaton 1988: 160).

In a similar vein, Hughes (1989) states; “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.” (P. 22). In addition, Harrison (1983: 343) notes, “Content analysis of a language item involves inspection of the item to determine if it contains a language problem that is valid and representative of the problems that students have in mastering the foreign language.”

The above definitions clearly show that content validity is the validity around the representativeness of test. In other words, a test has content validity if the samples of the items are proportionate to the content blue print of the material taught.

Having said this much about its definitions, let us proceed to its role in language learning and the procedures we need to follow to achieve a proper content validity in foreign language testing.

- **Why Content Validity?**

As with other types of validity, content validity too plays an important role in test validation. Indeed, it is quite relevant here as it is a means to ascertain how thoroughly a language test samples the instructional objectives or the universe

of criterion behavior. Hughes (1989) clearly articulated its practical importance to language learning and teaching as follows:

... The greater a test's content validity, the more likely it is to be an accurate measure of what it is supposed to measure. A test in which more areas identified in the specification are underrepresented or not represented at all is unlikely to be accurate (p.22).

He further elaborates that unless the test's content adequately samples and represents the language areas covered so far, there is no doubt that it results in harmful backwash in the entire process of teaching and learning. For instance, it is obvious that areas which are not tested are likely to become areas ignored in teaching and learning.

The point made by Hughes shows us that content validity plays two important roles in test validation. First, it guarantees the accuracy of the test provided that the major areas identified in the specification have accurate representation in the test. Second, it helps to avoid a harmful backwash effect of testing upon the entire process of teaching and learning. As it is mentioned above, it is true that areas which are not tested are likely to become areas ignored in the teaching and learning. In other words, both the teachers and the students often gear their teaching and learning around the language content areas which are often tested. However, with careful analysis of the content, we can have a balanced representation without making bias among the language content areas and thereby avoiding such a harmful effect.

The importance of content validity has also been succinctly stated by Kohonom in Asmare (2008) as follows:

It is known that anticipation of testing procedures has a washback effect on learning; learners prepare for examinations and organize knowledge in memory in the light of how they are to be tested. ... Evaluation thus affects both quality and quantity of learning. Therefore, it needs to examine in terms of both the learning process and the outcomes of the learning (p.20).

Strengthening this fact, Weir (1993) states that it is the fact that language contents which are frequently addressed in the tests are studied and practiced by learners in a better way than others which do not appear in tests. This is because students often gear around the language contents which will be the content of the test during the instruction.

All of the above quotations explicitly state the consequence of negative effect (harmful backwash) of testing upon teaching and learning process. As it could be inferred from what has been said so far, the reason for this is not hard to find. This happened because of lack of proper content validity. That is, considering all contents of the syllabus/the material taught proportionally during testing facilitates language learning and thereby helping to have an acceptable decision about the learners' performance and the program or syllabus.

Content validation can thus be seen as such a crucial issue. Once this has been clear, however, we can go further and ask the question: 'How can content validity be achieved?' In other words, what are the things or procedures should we consider so as to have a proper content validation in any foreign language test?

Heaton (1988) provides the following considerations/procedures to be made in this regard. These considerations read as:

When embarking on the construction of a test, the test writer should first draw up a table of test specifications, describing in a very clear and precise terms the particular language skills and areas to be included in the test. ... In this way, the test should achieve content validity and reflect the component skills and areas which the test writer wishes to include in the assessment (pp. 160-161).

Madsen (1983), on his part, states that content validity can be achieved/obtained by determining how well the contents of the test samples the class of situations or the subject matter about which the conclusions are to be drawn.

In addition, in the process of establishing content validity, Hughes (1989) concludes that most of the time there is a tendency among test constructors to test language contents which are easy to test rather than testing language contents which are needed to be tested. This problem can only be solved by employing test specifications and thereby assuring whether or not there is a reasonable representation of the language contents.

As can be seen from quotations above, these scholars have tried to show some efforts to be made in establishing content validity. One of which is to have careful analysis of the language tested and the particular course objectives. The other is to have full test specifications and thereby ensuring whether or not the test content properly reflects these specifications. Generally speaking, content validation should therefore be the main concern of language test writers since it is one of key promoters of language learning.

2.4.1.2 Criterion Related Validity

Hughes (1989: 23) states, "... is to see how far results on the test agree with those provided by some independent and highly dependable assessment of the candidate's ability. This independent assessment is thus the criterion measure against which the test is validated." According to him, there are two kinds of criterion related validity: namely, predictive validity and concurrent validity.

Predictive validity deals with the extent to which how well the test scores predict some future behavior/future performance of the candidate (Ibid). In other words, this definition makes clear that this type of test validity helps us to judge whether or not a candidate can cope with the later learning based on the results (achievements) of the first test. An example would be how well the results of primary education completion examination could predict the student's ability to cope with subjects/courses in junior secondary schools (grades 9 and 10 in our cases).

Concurrent validity, on the other hand, is obtained by correlating scores of a given test with its parallel version of another test (Alderson et al., 1995). What should be clear here is the two tests are not different tests. Rather they are two tests produced from the same content and administered at about the same time to the same candidates.

Similarly, Hughes (1989) also notes that concurrent validity is indicative to the validity of the new test only if there is high correlation (high level of agreement) between the two sets of scores of the two tests. This point is better summarized by him in the following paragraph:

References to 'a high level of agreement' and 'little agreement' raise the question of how the level of agreement measured. There are in fact standard procedures for comparing sets of scores in this way, which generate what is called a 'validity coefficient', a mathematical measure of similarity. Perfect agreement between two sets of scores will result in a validity coefficient of 1. Total lack of agreement will give a coefficient of Zero (Ibid)

As it could be inferred from the above definitions, predictive validity and concurrent validity are similar in one aspect and different in the other. Regarding their similarity, both of them are obtained by the mathematical formulae for the computation of validity coefficient.

However, the main difference between the two resides in the time gap and the difficulty in the interpretation of the correlations. That is, since the duration of time to administer the two tests is almost the same in concurrent validity, it has got shorter time gap than predictive validity. Moreover, unlike correlations in concurrent validity, correlations in predictive validity are difficult to interpret. This is because the students' ability and performance is affected by different factors which result from the length of the time gap between the two tests.

2.4.1.3 Face Validity

Face validity is another branch of test validity which is hardly scientific concept. Many scholars (Lado 1961; Heaton 1988; Hughes 1989 ;Underhill 1991) note that this type of validity deals with the general appearance of a test like neatness of the test paper, font size, clarity of instructions, spacing, etc. All of which can be evaluated by any educated person with simple inspection. Let us see other qualities of a good test.

2.4.2 Reliability

Reliability is one of the qualities of a good test. It refers to the production of a test with regard to the consistency in its scores (results) on different occasions, while the conditions of the tests remain the same. Lado (1961) and Harrison (1983) state that for a test to be reliable it should yield the same (or as nearly as possible the same) score of that of the first score obtained by the student if we give the test gain.

Similarly, Atkins et al. (1996: 168) also explain the concept of reliability, "If a group of students takes the test on two occasions, their results roughly the same each time." These definitions imply that for a test to be reliable its scores should be consistent they are conducted to the same group in different period of time.

However, the reliability (consistency in scores) can be affected by different factors. Some of these are variation in testing conditions, differences in administrative instructions, inadequacy in sampling, personal factors such as motivation, illness, etc. (Heaton 1988: 161-162).

The description above given by Heaton clearly reveals that reliability of a test can be affected by a number of factors. Hence, the most important question that is likely to be asked is: 'How can these factors are avoided so as to achieve reliability in testing?'

Regarding making tests more reliable, Hughes (1989) outlines a set of ways. Some of these are:

- *Take enough samples of behavior (i.e. other thing being equal, the more items that you have on a test, the more reliable that the test will be.*
- *Do not allow candidates too much freedom.*
- *Write unambiguous items.*
- *Employ multiple, independent scoring etc.*

Before this session is winded up, it seems of more importance if something is said about the relationship between reliability and validity since their relationship has often been seen as rather complex. Indeed, different scholars perceive their relationship in different ways. But all expressions, if critically viewed, reflect the same essence. Harrison (1983), for example, explains, "On the one hand, a test can be reliable without being valid. ... A test cannot be valid, however, without being reliable." (P. 12).

What could be inferred from Harrison's seemingly controversial expressions is that these concepts are really complementary to each other. So, it seems that judging their relationship depends on a given context.

Bachman (1990) states that reliability is the necessary condition for validity. He further justifies that the primary concern in estimating the reliability is to examine the variance in the test scores they themselves. On the other hand, the primary concern in validity is to consider other sources of variance. That is, words, according to him, reliability deals with variance in test scores, whereas validity deals with things which cause this variance in test scores. Yet he admits that it is difficult to have a clear cut between them. Weir (1993) also illustrates that though there is no such a thing as validity in the development and use of any test, it is likely to be invalid simply because of absence of reliability.

2.4.3 Practicality

Being one of the precautions we take to improve the quality of our tests as instrument of measurement, practicality deals with issues like economy, scorability and administrability (Harrison 1983: 12-13).

Each of these is illustrated as follows:

- *Economy (i.e. to reduce cost): refers to being economical in cost (financial capacity, materials and the hidden cost of time spent).*
- *Scorability (i.e. easy of scoring): refers to being easy in preparation, marking, etc.*
- *Administrability: this criterion includes issues like. The number of people will be involved (the teacher time is a hidden expense), special arrangements to be done, the time arrangements for efficient filling of test materials, etc. (Ibid).*

Moreover, Dejenie (1994) and Atkins et al. (1996) also confirm that the test's practicality is highly important, especially in large scale testing. That is, when it is to be taken by a large number of students.

2.4.4 Comprehensiveness

Atkins et al (1996) have this to say about comprehensiveness:

A test may be said to be comprehensive when it contains an appropriate proportion of items from all aspects of the language or material to be tested. The number of items needs to be in proportion to the importance of the language area being tested, or in proportion to the problems of the students (p. 168).

At the same time, they point out that there should be great care in doing so. This is because of some consequences that result from making a test too long and too short. Some of the consequences which happen when a test becomes too long, for example, are problems of administration, economy, fatigue in the part of examinee and destruction of the test's utility. On the other hand, the consequences that result from short test are like problems of comprehensiveness and doubtful reliability. Hence, these points remind us that all of these cautions must be considered in the attempt to achieve comprehensiveness in a test.

In summary, it is learnt from the literature review that there exists a very slight distinction between terms like test, assessment and examination. It is also learned that there is strong relationship between language testing and language teaching. They are inseparable entities and often influence each other. Besides, from the different types and purposes of tests, it has been made clear that, especially the contents of an achievement examination like primary education completion examination should proportionally sample the contents of textbooks and the syllabus objectives. What is more, it is stated that a test should achieve content validity and reflect the component skills and areas which the test is supposed to measure. In other words, a test should measure what is important to measure rather than what is easy to measure.

The next chapter runs us through the methodology and procedures employed to achieve the objectives which were set in advance. It also gives brief descriptions of the two textbooks and the 2002 E.C. English Examination.

CHAPTER THREE

METHODOLOGY AND PROCEDURE

As has been indicated in the previous chapters, the main purposes of this study were to determine whether the 2002 E.C Primary Education Completion English Examination comprises a representative samples of the contents of grades 7 and 8 English textbooks and to find out the degree of the extent of relationship between contents of the two materials (the two textbooks and the 2002 E.C English exam). Besides, it was also the intention of the present study to examine the allocation of percentages to the language content areas intended to be tested. In order to achieve these objectives, therefore, it is important to give information on the general description of the above two textbooks and the 2002 E.C English exam, data sources and data collecting procedures.

3.1 A General Description of Grades 7 and 8 English

Textbooks and the 2002 E .C English Exam

As was mentioned in the 'background section', the present study focuses on the grades 7 and 8 English textbooks which were published in 2001 E.C. These textbooks have 16 and 15 units respectively. Each unit of the two textbooks has been divided into different lessons. The major language content areas in these textbooks are 'reading comprehension', 'vocabulary', 'language pattern' (the name given to grammar in the syllabus of grades 7 and 8), 'writing', 'listening' and 'speaking' as it was clearly stated in their corresponding syllabus.

The 2002 E.C English exam is the first exam which was prepared from the above textbooks since the textbooks were published in 2001 E.C. The contents of the exam are 'reading comprehension', 'references', 'vocabulary', 'dialogue completion', 'grammar', 'rearranging jumbled sentences', 'word arrangement', 'sentences comprehension', and 'sentence combination'.

3.2 Data Sources

The main data sources for the present study were the two textbooks and 2002 E.C exam which were described above. Besides, an attempt was made to gather relevant information from two test constructors of the 2002 E.C English exam and the director of Teaching _ Learning Assessment /Core Process in South Nations, Nationalities and Peoples Regional State Education Bureau (SNNPRSEB).

All of them were made to participate in the study as source of data on the basis of comprehensive sampling technique. Only the director is a full time worker in the aforementioned office, whereas the rest two are college English *instructors who were invited to offer their professional assistance*. All of them have their first degree. The one who works as a full time in the office has 9 years of working experience. The college instructors have 30 and 6 years of working experience in teaching English.

3.3 Research Design and Data Collecting Procedures

Descriptive study was used for it explains the present situation of the problem, and attempts to measure the problem exists without questioning why it exists. The major data gathering tool employed in this study was 'content analysis'. Content analysis (also called text analysis) is a suitable technique to deal with studying documents. It deals with the transcripts which the researcher must read and analyse (Sarantakos, 2005). Hence, it was assumed to help get a brief summary of the contents of the Grades 7 and 8 English textbooks, and the 2002 E.C. English examination. This tool was substantiated by unstructured interview (informal conversation) which was held with the three individuals mentioned above. The unstructured interview was chosen by the assumption that subjects would give detail and reliable information during the informal interview.

In general, the content validity analyses of the English textbooks and the 2002 E.C. English examination as well as their degree of relationship have been determined by employing different steps:

First of all, the contents of the textbooks and the exam were classified into major language areas, and analyzed and presented in a summary table showing the frequencies of the language content areas both in figures and percentages. Content area here refers to any language content in the textbooks or exam. For instance, dialogue completion is considered as one content area. Frequencies, on the other hand, refer to the number of times with which the given language content area is dealt within the textbook or the examination. That is, '10 practice exercises' on dialogue completion in a textbook or exam' means that the dialogue completion as language content area has a frequency of 10 in that textbook. If 2 questions were asked on dialogue completion in an exam, it would be said that dialogue completion as a content area had a frequency of 2 in the examination content.

After this, the extent of relationship between frequencies from the analyses of the textbooks and frequencies from the analyses of the examination questions was determined using percentage and the Sidney Siegel's formula.

According to Harper (1977: 190) (in Alemu 1983: 41) the breakdown of the Siegel's formula begins with X^2 (pronounced Ki-as in "kite"-squared). To find X^2 , three steps were followed. First, sets of observations in terms of frequencies were classified as major language content areas in a contingency table (i.e., table of frequencies from which the value of 'C' is computed) as row totals, column totals and grand totals. Second, "expected value" was computed for each "observed value", using the following formula:

$$\text{Expected value} = \frac{\text{Row total} \times \text{column total}}{\text{Grand total}}$$

Third, the expected values were subtracted from the observed values; the difference was squared, and divided by the expected values. The quotients were

added to give the value of X^2 . The contingency coefficient 'C' was then computed using the formula.

$$C = \text{square root of } \frac{X^2}{N + X^2} \text{ where } N \text{ refers to the grand total.}$$

To summarize in the mean time, it was, thus, hoped that the end-result of all these analyses would be helpful to get information about the main purposes of this study which were set before.

CHAPTER FOUR

DATA ANALYSIS, RESULTS AND DISCUSSIONS

In the previous chapter, an attempt was made to give a general description of grades 7 and 8 English Textbooks and the 2002 E.C English exam. In addition, data sources and data collecting procedures were also stated, assuming that they would help achieve the objectives which were set in advance. In this chapter, the contents of these textbooks, the exam and the interview questions were analyzed; the results were discussed and interpreted.

4.1 Textbook Analysis

As was described under chapter three, the major language content areas of the two textbooks are 'reading comprehension', 'vocabulary', 'language pattern' (the name given to grammar in the syllabus of grades 7 and 8), 'writing', 'listening' and 'speaking'. This classification was done in the same way as the content areas' classification in the syllabus.

The number of times that the items of practice (practice questions) in each content area of the textbooks was put in frequencies. For example, vocabulary extension practice items in the textbooks have 348 items (see Table 1) below. Therefore, it means that the vocabulary extension as a content area has a frequency of 348 in the textbooks.

However, for productive skills (speaking and writing) it was found to be difficult to put them in terms of frequencies. This is because the practice items of these skills are too general to be expressed quantitatively (in frequencies). Some of the examples of the speaking tasks are story telling, talking about the picture, debating, etc. Writing an essay, note taking, etc. are some of the examples of writing activities. Hence, counting the practice items of these skills quantitatively and comparing with the items in the exam is so difficult. For this

reason an attempt was made to see them in terms of periods allotted to them (see Table 2) below. For example, speaking was given a total of 71 periods.

'Dialogue completion' practice items and items on 'rearranging jumbled sentences' were treated separately though they seem to be aspects of speaking and writing respectively.

Consequently, the content areas drawn up as major language areas of the textbooks for sake of comparison with their corresponding items of practice in the 2002 E.C English exam are 'reading comprehension' 'vocabulary' 'grammar' 'listening' 'dialogue completion' and 'rearranging jumbled sentences.'

What follow are the analyses of the language contents of the textbooks in two different ways in two summary tables. Table 4.1 below shows the analysis of the six major language content areas in terms of periods allotted to them. However, Table4.2 shows the language content areas which were selected for the purpose of comparison from the textbooks and their frequencies of items of practice.

Table 4.1: The Six Major Language Content Areas of the Grades 7 and 8 English Textbooks and the Number of Periods Allotted to them

No	Content area	Total time	Percentage
1	Reading comprehension	46	14.33
2	Vocabulary	43	13.40
3	Grammar	44	13.71
4	Listening	38	11.84
5	Speaking	71	22.12
6	Writing	79	24.61
	Total	321	100

When considering validity of an English test, we should test a representative sample of skills and language areas from the instructional material we have taught. To achieve a fair representation, we need to write down all practice items both in the skills and language areas and thereby giving percentages to each of them on the basis of the time we spent teaching them /our feeling about their importance (Atkins et al., 1996). Hence, it is believed that the number of periods given to language contents can tell us the emphasis given to that particular content area.

As seen from the Table 4.1 above, from the total of 321 periods, writing was given 79 (24.61%). The second large amount of time was given to speaking which comprises 71 (22.12%). The much emphasis attached to the productive skills could play a practical role in improving the frequent problem among students, provided that the same attention is given to these skills in the classroom teaching and testing as well as in the national examinations. This is because practice for examination represents practice in these skills (Hughes, 1989). It could also be deduced from Table 4.1 above that the development of students' communicative ability both in speaking and in writing was given prominence over the development of students' ability in other language content areas in the two textbooks.

The difference in emphasis among the language content areas becomes more vivid if we compare and contrast the amount of time given to speaking and grammar, for example. That is, as seen from table 1, speaking was given 71 (22.12%) from the total of 321 periods, whereas grammar was given 44 (13.71%). It could be said that the former is nearly twice the later. Reading comprehension, vocabulary and listening have shared 46 (14.33%), 43 (13.40%) and 38 (11.84%) periods respectively

Table 4.2: Content Areas of the Grades 7 and 8 English Textbooks and their Frequency of items of Practice

No	Content area	Frequency of item	Percentage
1	Reading comprehension	371	26.11
2	Vocabulary	348	24.49
3	Grammar	427	30.05
4	Listening	251	17.66
5	Dialogue completion	10	0.70
6	Rearranging jumbled sentences to make coherent paragraph	14	0.99
	Total	1421	100

Table 4.2 above indicates the language content areas and their frequency of items of practice which were drawn up for sake of comparison with the language content areas in the 2002 E.C exam. As seen from the table, there are a total of 427(30.05%) grammar items in the two textbooks. Reading comprehension accounts for 371(26.66%). Vocabulary and listening constitute 348(24.49%) and 251(17.66%) respectively. Dialogue completion and rearranging jumbled sentences comprise 10(0.70%) and 14(0.99%) respectively. There exists difference regarding the number of items of practice among the content areas. For example, between grammar and reading comprehension, there is a difference of 56 items, but between vocabulary and listening there is a difference of 97 items.

In general, it is learnt from the analysis of the of six major language content areas of the two textbooks according to the time allotted to them that the productive skills (speaking and writing) are highly emphasized skills. With regard to expressing language contents in terms of frequency of items of practice, grammar constituted the highest frequency of items. Reading comprehension accounted the second

highest share. Listening constituted the third highest number of frequency of items of practice following Vocabulary items. Dialogue completion and rearranging jumbled sentences seem to have fewer shares.

4.2 Analyses of the 2002 E.C Primary Education Completion English Examination

Likewise, as was done in the analysis of contents in the two textbooks, the first step in the analysis of contents in the examination questions is also classifying the contents of examination questions into major language content areas. The second step is to determine the number (frequency) of practice items in each content area (Underhill 1991; Alderson et al. 1995).

Hence, the next two tables (Table 4.3 and Table 4. 4) show the classifications of the content areas of the 2002 E.C English exam. Table 4.3 depicts the detailed content areas as they appear in the exam, whereas Table 4.4 shows the reclassified summary of content areas which were used for the purpose of comparison.

Table 4.3: Content Areas of the exam and their Frequencies of Items

No	Content area	Frequency of Items	Percentage
1	Reading comprehension	6	10
2	References	4	6.67
3	Vocabulary	5	8.33
4	Dialogue completion	5	8.33
5	Grammar	24	40
6	Rearranging jumbled sentences	4	6.67
7	Word arrangement	3	5
8	Sentence comprehension	5	8.33
9	Sentence combination	4	6.67
	Total	60	100

As can be seen from the Table 4.3, the contents of the 2002 E.C English exam contained a total of 60 items, which were constructed with the assumption of assessing students' ability in different language content areas.

For example, according to the information gathered from test constructors and the director of Teaching-Learning Ass/Core Process in SNNPRSEB, the 'dialogue completion' questions in Table 4.3 were designed to assess students' speaking ability, whereas 'rearranging jumbled sentences to make coherent paragraph', 'word arrangement' and 'sentence combination' were designed to assess students' writing skills. However, the assumption behind the inclusion of these tasks has not been recognized by test experts.

Contrary to this, test experts like Hughes (1989) and Heaton (1975, 1988) claim that there is great uncertainty to judge one's ability of speaking and writing by measuring through multiple-choice speaking and writing tests. They further note that it is this uncertainty which lead for the need of direct testing than an indirect one. That is, it is advisable to test skills, especially productive skills (speaking and writing) directly. The concept 'direct testing' here refers to testing speaking or writing ability through speaking and writing employing techniques like interview, translation, etc rather than employing multiple-choice speaking /writing test which evaluates candidates ability indirectly (Ibid).

Similarly, Weir (1990) also argues that such indirect testing has no guarantee and trust whether or not a candidate who is successful in the indirect test (i.e. multiple-choice writing or speaking test) will be successful in the practical situations. It is only the direct testing of these skills which will result in a better content, face and wash back validity.

As mentioned in the above opening paragraphs, the 9 content areas in Table 4.3 were reclassified in Table 4.4. Here, it should be noted that the rationale for the reclassification was for the purpose of comparison and convenience rather than for any theoretical considerations.

Therefore, 'references' were categorized under 'reading comprehension'. But, questions on 'word arrangement', 'sentence combination' and 'sentence comprehension' were categorized under grammar and usage section. One possible reason to categorize these contents under the grammar section is the fact that answering them needs the knowledge of grammar in one way or the other. For instance, rearranging words to make a meaningful sentence needs ability of ordering subject, verb and object (complement) of a given sentence properly. Answering most of the sentence combination questions in this exam needs the knowledge of usage of 'relative pronouns' as linkers of sentences, and the sentence comprehension questions were supposed to test the meaning of grammatical components, such as "too + adjective + to", "unless" to mean "If... not", degree of comparison (positive degree- "as + adjective + as"), "asking permission" and "Neither... nor".

Accordingly, the content areas drawn up as major contents for the purpose of comparison with the selected contents of the textbooks (i.e., with the contents in Table 4.2) were reading comprehension, vocabulary, grammar, dialogue completion and rearranging jumbled sentences. The reclassified content areas and their frequency of items are summarized both in figures and percentages in the following table.

Table 4.4: Summary of the major content areas of the exam and their frequency of items

No	Content area	Frequency of Items	Percentage
1	Reading comprehension	10	16.67
2	Vocabulary	5	8.33
3	Grammar	36	60
4	Dialogue completion	5	8.33
5	Rearranging jumbled sentences to make coherent paragraph	4	6.67
	Total	60	100

As the distribution of data in Table 4.4 shows, grammar was given prominence in the exam as it was the dominant one in the textbooks. From the total of 60 items, it comprised 36 (60%) which is surprisingly the highest of the total percentage given.

Similarly, relatively speaking, reading comprehension 10 (16.67%) is also emphasized in the exam next to grammar as it was the second dominant in the textbooks. Vocabulary and dialogue completion have been given equal share. They constituted 5 (8.33%) each. However, in the textbooks there is great disparity among them. Jumbled sentences to make coherent paragraph accounted for 4 (6.67%), but it constituted 14(0.99%) in the two textbooks.

With regard to allocation of percentages to these contents, for example, reading comprehension, vocabulary and grammar in the textbooks, we can see that there was not such a big difference. However, in the exam, there exists a huge disparity between these language contents. In the textbooks, reading comprehension, vocabulary and grammar had (26.11%), (24.49%) and (30.05%) respectively, whereas in the exam they had (16.67%), (8.3%) and (60%) respectively, which can raise the question of proportional representation of content areas of the instructional material taught.

So far, two sets of data in terms of frequencies have been presented using figures and percentage. Table 4.2 indicates data from the analysis of the textbooks, but table 4 shows data from the 2002 E.C primary Education Completion English Examination. What follow are putting these two data sets which were selected to be compared to each other in one table (see Table 4.5), and determining the degree of extent of relationship between the two sets of data (i.e. contents of textbooks and the examination) by applying Sidney Siegel's formula. The formula is:

$$C = \sqrt{\frac{X^2}{N+X^2}}$$

Where, C = the value of relationship, **0.21** (see Table 4.6 below)

N = grand total, **1230** (see Table 4. 5 below)

X²= refers to the result found after employing three steps,
59.5076 (see 4.6 below)

To find X², three steps were followed:

Step 1 (Table 5), contingency table, sets of observations in terms of frequencies are classified under five components, namely: 'reading comprehension', 'vocabulary', 'Grammar', 'Dialogue completion', and 'rearranging jumbled sentences'. The row totals, as in 381 for reading comprehension, the column totals, as in 1170 for textbooks coverage, and the grand total 1230, are arranged.

Step 2, "expected value" is computed for each observed value, using the following formula:

$\text{Expected value} = \frac{\text{Row total} \times \text{column total}}{\text{Grand total}}$
--

For example, the expected value for the observed value 14 textbook frequency is calculated to be 17.1220.

In step 3, the expected values are subtracted from the observed values the difference is squared, and divided by the expected values. The quotients are added to give the value of x²

The contingency coefficient 'C' is then computed using the formula.

$$C = \text{square root of } \frac{X^2}{N+X^2} \text{ where N represents the grand total}$$

Table 4.5: The Comparison between total Frequencies of content areas of grades 7 and 8 English Textbooks and the 2002 E.C Primary Education Completion English Examination

No	Content area	Frequencies from textbooks	Frequencies from exam	Total frequencies
1	Reading comprehension	371	10	381
2	Vocabulary	348	5	353
3	Grammar	427	36	463
4	Dialogue completion	10	5	15
5	Rearranging jumbled sentences to make coherent paragraph	14	4	18
	Total	1170	60	1230

Table 4.5 above comprises two sets of data from the textbooks and the exam. These data were drawn up from the two materials (the textbooks and the exam) for the purpose of making comparison and thereby determining the degree of extent of relationship between the contents of the two materials. Table 4.6 below shows the attempt made to obtain their relationship

Table 4.6: Extent of Relationship between the Contents of Grades 7 and 8 English Textbooks and the 2002 E.C Primary Education Completion English Examination

	Observed value	Expected value	X²
Frequencies from textbooks	371	$\frac{381 \times 1170}{1230}$ = 362.4146	$\frac{(371-362.4146)^2}{362.4146}$ =0.2034
	348	$\frac{353 \times 117}{1230}$ =335.7805	$\frac{(348-335.7805)^2}{335.7805}$ =0.4447
	427	$\frac{463 \times 1170}{1230}$ =440.4146	$\frac{(427-440.4146)^2}{440.4146}$ =0.4086
	10	$\frac{15 \times 1170}{1230}$ =14.2683	$\frac{(10-14.2683)^2}{14.2683}$ =1.2768
	14	$\frac{18 \times 1170}{1230}$ =17.1220	$\frac{(14-17.1220)^2}{17.1220}$ =0.5693
	Frequencies from exam	10	$\frac{381 \times 60}{1230}$ =18.5854
5		$\frac{353 \times 60}{1230}$ =17.2195	$\frac{(5-17.2195)^2}{17.2195}$ =8.6713
36		$\frac{463 \times 60}{1230}$ =22.5854	$\frac{(36-22.5854)^2}{22.5854}$ =7.9676
5		$\frac{15 \times 60}{1230}$ =0.7317	$\frac{(5-0.7317)^2}{0.7317}$ =24.8987
4		$\frac{18 \times 60}{1230}$ =0.8780	$\frac{(4-0.8780)^2}{0.8780}$ =11.1012
Total			59.5076

$$C = \sqrt{\frac{X^2}{N+X^2}}$$

$$C = \sqrt{\frac{59.5076}{1230+59.5076}}$$

$$C = \sqrt{0.04614754}$$

$$C = 0.21$$

As it was said so far, the data in Table 4.6 is the conclusion of the whole analysis of both the textbooks and the exam. It shows that the degree of the extent of relationship between the contents of the two materials found to be 0.21. Now, before discussing the results of the whole analyses done so far, let us see the analysis of the interview questions.

4.3 Analysis of Interview Questions

This section consists of the summary of the three interviewees' (two test constructors and the director of the Teaching _Learning Ass/Core Process) reactions to the four interview questions provided to them.

Question 1: Which grades English textbooks are the content of the examination supposed to cover?

The responses to this question indicated that the contents of the examination were supposed to cover the language contents of grades 7 and 8. This is why the analysis of the present study focused only to these textbooks.

Question 2: What are the major language content areas which addressed in the examination?

According to their responses to this question, the major language content areas addressed in the exam were 'reading comprehension', 'vocabulary', 'grammar', 'speaking' and 'writing'.

Question 3: How do you think that the productive skills (speaking and writing) can be assessed in the examination?

The interviewees responded that there were activities included in the exam with the assumption of assessing students' ability of speaking and writing indirectly. For example, 'dialogue completion' part was designed with the

intention of assessing students' communicative ability. However, activities such as 'word arrangement' to make meaningful sentences, 'sentence combination' and 'rearranging jumbled sentences' to make a coherent paragraph were designed to assess students' ability of writing.

Question 4: Do you think that the contents of the two textbooks adequately and proportionally represented in the exam? If 'Yes', to what extent?

In response to this question, all of the interviewees showed some sort of reservation about the match between the contents of the two textbooks and the exam. As they further explained, the reason for their doubt is because there has not been any training given to them. The only thing they did was following the trend that they had got from their work experience.

The next section presents all the discussions and interpretations made regarding the sets of data indicated so far.

4.4 Discussions

Using the statistical test, the contingency coefficient 'C' by Siegel, the extent of relationship between textbooks and examination contents has been determined. As indicated in Table 4.6, the result is that the value of 'C' is 0.21. The value of 'C', according to Siegel and Spearman, should read 1 (one) where there is a perfect relationship between two observations. But if the correlation value reads less than 0.5, there is weak correlation between the two observations. Similarly, Sarantakos (2005) elaborates that a correlation can be interpreted into five different ways. For example, if the correlation coefficient has a value under 0.20 is considered to be very low. If the coefficient has a value between 0.21 and 0.41 is said to be low. If the coefficient has a value between 0.41 and 0.70 is said to be moderate. On the other hand, if the coefficient has the value between 0.71 and 0.91 and over 0.91 is said to be high and very high respectively.

Therefore, the computed value of 'C' (0.21) in this study is said to be low. This means that the examination coverage does not adequately represent the textbooks coverage. This issue was also reflected in the interview question 4 since the test constructors they themselves doubt about the adequacy in sampling, mentioning that they do not have any expertise gained by special training. As a result, they simply follow the trend that they have gained from their work experience. Thus, it could be concluded that under the present set up of the English examination a true assessment of the extent of a candidate's knowledge of the language remains under question until a test is made to represent more of the textbook coverage.

As indicated in Table 4.4, which shows the analysis of contents of the exam both in figures and percentages, there was a clear disparity in the allocation of percentages to the content areas tested. Consider, for instance, the gap between grammar (60%) and reading comprehension (16.67%). The former is more than three times the latter. However, in the textbooks' coverage there was not such a significant difference. They were given nearly equal coverage, 26.11% and 30.05% respectively (see Table 4.4). Besides, 'vocabulary' and 'dialogue completion' were given equal coverage, 8.33% each which is unfair since vocabulary constituted a total frequency of 348 (24.49%) in the textbooks, whereas dialogue completion accounted for only 10 (0.70%). Vocabulary was given surprisingly little attention in the test content.

Thus, weighing the performance of candidates according to disproportionate importance attached to content areas will only be unfair to candidates because experience suggests that all candidates may not be equally good at a given content area.

There were also some activities included in the exam which did not appear in the textbooks. For example, 'references', 'word arrangement' and 'sentence

comprehension' (see Table 4.3) were activities which students were not exposed to so far in the textbooks.

With reference to testing productive skills (speaking and writing), it has been determined that an attempt was made to test these skills only indirectly. That is, they do not involve pen and paper work and only through multiple choice speaking and writing test. Therefore, this improper representation of items seems to negatively affect the achievement of beneficial backwash effect, as well as fulfillment of learning objectives and above all, lower the content validity of the examination. In connection to this, Bachman (1990) notes that the principle of keeping a balance in assigning weights to the areas to be tested is must if we need the information which result from performance of students on achievement examination as an indication of the degree to which the expected objectives of the course are being attained.

CHAPTER FIVE

SUMMARY OF THE FINDINGS, CONCLUSIONS AND RECOMMENDATIONS

5.1 Summary of the Findings

In this section of the study an attempt was made to summarize some of its major findings. The first major finding is related to determining the extent of relationship between contents of the grades 7 and 8 English textbooks and the 2002 E.C. Primary Education Completion English Examination of SNNPRS, which was the first basic research question of the present study. With this regard, as seen from Table 4.6, comparison between contents of the textbooks and exam was made after thorough analysis of each content area so as to determine the extent of relationship between the two observations. The result shows that it is weak as it was interpreted in the discussion section; that is 0.21. In other words, it indicates that the degree of relationship between the textbook coverage and examination content is low.

The second most important finding was related to examining the allocation of percentages to each language content areas. The analysis of the content areas of the textbooks and the exam indicated that the allocation of percentages to content areas was characterized by disproportionate distribution (see Table 4.4). For instance, in the exam, from the total of 60 items, 'grammar' constituted 36 (60%). The rest language content areas like 'reading comprehension', 'vocabulary', 'dialogue completion' and 'rearranging jumbled sentences' together constituted 24 (40%).

There were also some other findings. For example, there were activities which did not appear in the textbooks that were included in the exam (see Table 4.3). These were 'references', 'word arrangement' and 'sentence comprehension'. These activities constituted about 12 (20%) of the total coverage given to the

exam. Besides, because the exam employed solely multiple-choice format, 'listening skills' were totally ignored, and an attempt was made to assess the productive skills (speaking and Writing) indirectly in the exam. Nevertheless, assessing candidates' speaking or writing ability through indirect testing (e.g. multiple-choice) test format can not be regarded as sufficient (Heaton, 1975, 1988; Hughes, 1989; Weir, 1990).

5.2. Conclusions

As it has been stated elsewhere in this thesis, the primary aim of this study was to assess whether or not content validity had been achieved between the grades 7 and 8 English textbooks and the 2002 E.C Primary Education Completion English Examination. Hence, on the basis of research objectives and major findings discussed so far, the following conclusions were drawn:

First, the degree of the extent of relationship between the textbooks coverage and examination content areas was found to be low, since the correlation coefficient between the two observations reads 0.21.

Second, the allocation of percentages to content areas was characterized by disproportionate distribution for the examination failed to adequately and proportionally represent the different content areas predominantly covered in the textbooks. Moreover, the match between the contents of the exam and the textbooks was also distorted because of the inclusion of some activities which did not appear in the textbooks.

In summary, due to these deficiencies, it was concluded that the 2002 E.C. Primary Education Completion English Examination of SNNPRS was poor at content validity.

5.3. Recommendations

Based on the summary and conclusions drawn from the findings, the following recommendations were forwarded:

- The inclusion of language content areas which did not appear in the textbooks should be avoided from the examination.
- It is advisable to make a content analysis scheme so as to determine whether or not balance is maintained in the distribution of percentages among content areas to be tested, and to check that content areas covered in the textbooks are given due test coverage.
- It is also advisable to attach test specifications with the textbooks so that the test contractors are able to construct tests which are valid contentwise.
- Teachers should give ample time to the direct testing of productive skills in the class since there is no room to test them in the national exam directly.
- A sort of training should be given to the test constructors regarding language testing.
- Since this study focused only on the content validity of the 2002 E.C., it seems difficult to generalize the result of this study to other years' examination papers. Thus, further research is required on issues like content validity; item analysis and backwash effect of each year's of the Primary Education Completion English Examination in SNNPRS.

BIBLIOGRAPHY

- _____. (1993). *Understanding and Developing Language Tests*. New York: Prentice Hall
- _____. (1988). *Writing English Language Tests, A Practical Guide for Teachers of English as a Second or Foreign Language*. London: Longman Group UK, Ltd.
- Alderson, J., Claphan, C., and Wall, D. (1995). *Language Test Construction and Evaluation*. Cambridge: Cambridge University Press.
- Alemu Tsegaye (1983). An Assessment of Grade Six and Eight English National Examinations (Unpublished MA Thesis). Addis Ababa University.
- Allen, J. and Davies, A. (1977). *Testing and Experimental Methods*. London: Oxford University Press.
- Argaw Sileshi (2004). The Content Validity of the EGSEC English Examination. (Unpublished MA Thesis) Addis Ababa University.
- Asmare Mehret (2008). An Assessment of the Content Validity of English Language Tests: The Case of Awassa Health Science College. (Unpublished MA Thesis). Addis Ababa University.
- Atkins, J., Hailom Banteyirga and Nuru Mohammed (1996). *Skills Development Methodology: Part 2*. Addis Ababa: Addis Ababa University Press.
- Bachman, L. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.

- Baker, D. (1989). (1989). ***Language Testing: A Critical Survey and Practical Guide***. Edward Arnold.
- Best, J.W. and Kahn, J.V. (1989). ***Research in Education (6th Ed.)***. New Jersey: Prentice-Hall, Inc.
- Brown, H. (1994). ***Principles of Language Learning and Teaching***. San Francisco State University: Prentice Hall Regents.
- Davies, A. (1968). ***Language Testing Symposium. A Psycholinguistic Approach***. London: Oxford University Press.
- Dejenie Leta (1990). Achievement, Wash back and Proficiency in School Leaving Examination: A Case of Innovation in an Ethiopian Setting. (Unpublished PhD, Thesis). Lancaster: Lancaster University.
- Gebrehiwot Salfore, Markos Mekuria, Seni Dutamo, and Asegedech Shawl (2010). The Survey Study on the Causes for the Decline of Students' Results in Grade Eight National Examination in SNNPRS (Unpublished), SNNPRSEB, Hawassa.
- Harrison, D. (1983). ***Language Testing Handbook***. New York: Longman Group Limited.
- Heaton, J. (1975). ***Writing English Language Tests***. London: Longman
- Hughes, A. (1989). ***Testing for Language Teachers***. Cambridge: Cambridge University Press.
- Kifle Kebede (1995). An Assessment Content Validity of High School English Language Tests in Relation to Textbooks Concerned. (Unpublished MA Thesis). Addis Ababa University.

- Lado, R. (1961). ***Language Testing***. London: Longman.
- Madsen, H. (1983). ***Techniques in Testing***. Hong Kong: Oxford University Press/London: Longman.
- Mulu Nega and Daniel Tefera (2005). Introduction to Measurement and Evaluation (Psyc 211): Modules 1-3. Addis Ababa University.
- Nuru Mohammed (1992). Level of Questions: A Description of Textbook and Examination Questions in Higher Secondary Schools. (Unpublished MA, Thesis). Addis Ababa University.
- Oller, J. (1979). ***Language Tests at School***. London: Longman Group Ltd.
- Oxford Dictionary of English (2005). ***The World's Most Trusted Dictionaries (2nd ed.)***. Oxford: Oxford University Press.
- Sarantakos, S. (2005). Social Research (3rd ed.). Online internet. Available: <http://www.palgrave.com/sociology/sarantakos>
- Teshome Demisse (1995). The Construction and Validation of Tests in English for Tertiary Education. (Unpublished PhD, Thesis). Addis Ababa University.
- Thorndike, R. (1997). ***Measurement and Evaluation in Psychology and Education (6th ed.)***. New Jersey: Printice-Hall, Inc.
- Tibebe Alemayehu (1992). ***The Predictive Validity of Ethiopian School Leaving Certificate English Examination and the Integrative Tests***. Comparative Study (EJE, Vol, XII, No 1).
- Underhill, N. (1991). ***Testing Spoken Language***. Cambridge: Cambridge University Press.
- Weir, C. (1990). ***Communicative Language Teaching***. New York: Prentice Hall.

APPENDIX A: INTERVIEW QUESTIONS

ADISS ABABA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

DEPARTMENT OF ENGLISH

Interview Questions for the test Constructors and the Director of Teaching _

Learning Ass/Core Process in South Nations, Nationalities and Peoples

Regional State Education Bureau (SNNPRSEB)

This interview is aimed at collecting information for the study entitled "Assessing Content Validity of Primary Education Completion English Examination of SNNPRS: 2002 E.C in focus". So, you are kindly requested to respond each item as genuinely and carefully as possible. Your cooperation is highly appreciated.

Thank you!

Part I: General information

1. Name _____
2. Sex _____
3. Educational Level _____
4. Number of work experience in that position _____

Part II: Regarding the 2002 E.C. Primary Education Completion English

Examination

1. Which grades English textbooks are the content of the examination supposed to cover?
2. What are the major language content areas which are addressed in the examination?
3. How do you think that the productive skills (speaking and writing) can be assessed in the examination?
4. Do you think that the contents of the two textbooks adequately and proportionally represented in the exam? If 'Yes', to what extent? Any additional comments

2

2

APPENDIX: B THE COPY OF
PRIMARY EDUCATION COMPLETION
ENGLISH EXAMINATION
SENE 2002/ JUNE 2010

NUMBER OF ITEMS: 60

TIME ALLOTTED: 60 MINUTES

Reading Comprehension

Direction: Read the passage below carefully to answer the questions next to it.

1. Nearly all animals have natural enemies, usually larger animals **who** hunt them for food. In most cases, nature develops a balance so that the different **species** can live side by side **without exterminating each other**. But human beings have changed all this. They have rushed into nearly every corner of the world. **Looking for more land** to use, and hunting down other species. Now many kinds of animals are in danger of extinction.
2. Some animals are hunted for food. A **case in point** is that of the Walia Ibex, the beautiful Ethiopian goat of which now only a few hundred specimens remain alive. Many animals are killed for **their** skins. The leopard is in a great danger of extinction by fur traders, because leopard skin coats are fashionable. Some animals are hunted for other products. Whales, for instance, have become rare because they are slaughtered for the oil **they** contain, which is used to make many products, including ice-cream and lip-stick. Some species are hunted for **pleasure**. Indian tigers nearly extinct because of the large number which have been shot by men for sport. In many such cases, men have gone on carelessly destroying the animals without considering the effects of their actions, or the danger of pushing a species to extinction.
3. Another great danger to wildlife is human beings' needs for land. All over the world more and more land is being taken over to grow food, and this means that the natural homes of wild animals are being destroyed. Gorillas, for instance, are

2

2

seriously threatened because the forests where they live are being cut down to make way for fields. Another example is the Indian buffalo, which has also become rare, because the grassy plains where it once lived have been take over by human beings' domestic animals. Among other animals that are in danger of this kind are the India rhino, the Mexican Grizzly bear, and many varieties of deer, lion and monkey, as well as many smaller animals.

1. Which one represents the main idea of paragraph 1?

- . Larger animals hunt their enemies
- . Different species are still living side-by-side
- . Hunting and expansion of farmland are the two dangers for the existence of wildlife.
- . Human beings improved the natural homes of wildlife

2. Paragraph 2 deals with _____.

- | | |
|---------------------|-----------------------------|
| A. Hunting animals | C. Pleasure |
| B. Making ice-cream | D. destroying natural homes |

3. Paragraph 2 indicates that _____ is hunted for food.

- | | |
|---------------|-------------|
| A. Whales | C. Leopard |
| B. Walia Ibex | D. Gorillas |

2

2

4. One of the statements below is not true about leopards.

- D. They are hunted for their skins
- D. They are hunted by fur traders
- D. their products are used to make fashionable coats
- D. Their products are good sources for oil

5. As per paragraph 2, _____ is hunted for pleasure.

- A. monkey
- B. Indian rhino
- C. Indian tiger
- D. Mexican bear

6. Gorillas are seriously threatened because _____.

- D. they are afraid of human beings' domestic animals
- D. the forests where they live are being cut down
- D. they migrate from place to place
- D. they live in vast national parks

7. Their (paragraph 2 line 3) refers to _____.

- A. beautiful Ethiopian goat
- B. fur traders
- C. Walia Ibex
- D. many animals

8. Who (paragraph 1 line 1) refers to _____.

- A. human beings
- B. natural enemies
- C. animals
- D. hunters

9. They (paragraph 2 line 6) refers to _____.

- A. whales
- B. some animals
- C. fur traders
- D. None of the above

2

2

10. It (paragraph 3 line 6) stands for

- D. grassy plains
- D. human beings' domestic animals
- D. Indian buffalo
- D. Indian rhino

Direction: For items 11-15, chose the best meaning for each of the words or phrases as they are used in the passage.

11. ...with out exterminating (Paragraph 1 line 3) means _____.

- A. defeating
- B. harming
- C. visiting
- D. helping

12. Looking for oral land ... (Paragraph 1 Line 4) means _____.

- A. examining
- B. taking care of
- C. despising
- D. searching for

13. Species (Paragraph 1 Line 2) means _____.

- D. group having common characteristics
- D. relatives
- D. money in the form of coins
- D. unique animals

2

2

14. A case in point ... (Paragraph 2 Line 1) means _____.
- | | |
|------------------------|------------------------------|
| A. forming this | C. container of such animals |
| B. an instance of this | D. food to build a body |
15. Pleasure (Paragraph 2 Line 7) means _____.
- | | |
|---------------|------------|
| A. Recreation | C. Humor |
| B. Politeness | D. Defense |

Direction: For items 16-20, choose the best alternative that completes the dialogue between Mary and Zinash.

16. Mary: Hi, Zinash. I'm Mary from Hawassa.

Zinash: _____ to meet you. I'm Zinash.

- | | |
|------------------|-----------------|
| A. I'm very well | C. Good morning |
| B. Nice | D. Fine thanks |

17. Mary: _____ ?

Zinash: I'm from Boditi.

- D. Did we meet before
- D. What language do you speak
- D. Where are you from
- D. Where from are you

2

2

18. Mary: Have you ever been to Hawassa before?

Zinash: No, _____.

A. I'm not

C. I didn't

B. I have never been

D. I have been once

19. Zinash: I'm a visitor here.

Mary: _____

A. So am I

C. Neither am I

B. So do I

D. I am not either

20. Mary: _____?

Zinash: Yes, I have.

A. Do you drive a car

B. Are you good at driving

C. Have you got a car

D. Are you looking for accommodations

Direction: For items 21-44, choose the best alternative to complete the given sentences.

21. I am not _____ a Biology teacher _____ a Unit-leader.

E. only/but also

C. either/or

A. neither/nor

D. so/that

2

2

22. Meseret can't run fast. _____.

A. So do I

C. So am I

B. Neither do I

D. Either do I

23. One of the girls _____ I met _____ long hair.

A. who/has

C. who/have

B. whose/has

D. whose/have

24. The woman _____ daughter _____ is the manager's secretary.

A. whose/was killed

C. whom/was killed

B. who/was killed

D. which/was killed

25. Excuse me _____ late.

A. for being late

C. for be late

B. to be late

D. to being late

26. Each of the girls _____ brought books.

A. is

C. has

B. are

D. have

27. If you _____ type of letter at once, you _____ the fee.

A. will not/didn't get

C. didn't /don't get

B. will not/will not get

D. don't/will not get

2

2

28. Tufa's wife was a good deal older than _____ when they married.

- A. he
B. him
C. his
D. her

29. Teacher : He is late.

Father : What is he saying?

Monitor : _____

- A. He says that he is late.
B. He says why he is late.
C. He said that he is late.
D. He said why was he late.

30. She was written letters while I _____ books.

- A. reads
B. read
C. am reading
D. was reading

31. At Leramo is the man to _____ tat letter was addressed to.

- A. who
B. whose
C. whom
D. which

32. He _____ late any more.

- A. should be not
B. should is
C. should not be
D. should not is

33. Marta sat down and _____ to write.

- A. begins
B. will begin
C. began
C. beginning

2

2

34. Ato Ayele has a house in Addis Ababa. He _____ be rich.

B. must

C. will

B. needn't

D. shouldn't

35. The man _____ three bottles of beer since he arrived.

B. drinks

C. drank

B. has drunk

D. was drinking

36. We won't need any light in this room. Please _____ the bulb.

B. switch off

C. put on

B. switch on

D. put of

37. The thief admitted _____ the money.

B. steal

C. stealing

B. stole

D. stolen

38. Visitor : How old is your school?

Teacher : It _____ in 1920.

B. was built

C. built

B. is built

D. has been built

39. Since they were not in the school compound, I could _____ find them in the classroom _____ in the library.

B. either/or

C. neither/nor

B. neither/or

D. either/nor

2

2

40. Muktar did not go with us to the lake as he was repairing his bicycle. I wish he _____ free.

B. were

C. will be

B. had been

D. is

41. My family and I _____ to Awash National Park next week.

B. will traveled

C. will be traveling

B. traveling

D. will traveling

42. Detamo was an outstanding student in the class; _____, he did not succeed in passing Primary School Leaving Certificates Examination.

B. because

C. however

B. so

D. besides

43. Our football team played _____ than its guests.

A. better

C. more better

B. best

D. as good

44. Muktar : Which question is the _____ in this examination?

Jemila : I think it is Number 20.

A. easier

C. most easy

B. more easy

D. easiest

2

2

Direction: Questions 45-48, are based on the sentences below. The sentences are not in their correct order. Decide and choose the letter which contains the correct order.

1. Eventually, with the discovery of their role in the spread of malaria and yellow fever, it has become a major focus of scientific study.
2. Many years ago, mosquitoes were regarded as insects with a nasty bite but of no more significance to cause degasses.
3. As a result, considerable attempts have been made to eradicate mosquitoes from the face of the earth.
4. Because of this, they were given very little attention by scientists and no attempt was made to control them.

45. Which sentence should being the paragraph?

A. 1

C. 3

B. 2

D. 4

46. Which sentence should come next?

A. 4

C. 2

B. 1

D. 4

47. Which sentence should be the third?

A. 3

C. 4

B. 1

D. 2

2

2

48. Which sentence should be the last?

A. 3

C. 4

B. 1

D. 2

Direction: Items 49-51, are about word arrangement. Choose the correct sentence that shows the best arrangement in each.

49. From/name/is/my/ Bruke Bekel Hawassa

A. I am Bruke Bekele from Hawassa.

B. My name is Bruke Bekele from Hawassa.

C. Name of my is Bruke Bekele from Hawassa.

D. Hwassa name is from my Bruk Bekele.

50. A student/am/at/primary school/Tabor/ I

A. I am a student at Tabor Primary School.

B. I am at Tabor Primary School a student.

C. Am I a student at Tobor Primary School.

D. Primary School at Tobor I am a student.

51. Are/fighting/we/Ethiopians/poverty/young/against.

A. We are fighting Ethiopians against young poverty.

B. We are Ethiopians fighting against poverty young.

C. We are young Ethiopians fighting against poverty.

D. We are fighting against poverty young Ethiopians.

Direction: For question 52-56, read the given sentences. Then choose the alternative that is the same or nearly the same in meaning as the given sentence or part of it.

52. Do you mind my opening the window?
- A. Does it bother you if I open the window?
 - B. Is it ok if you open the window?
 - C. Do you remember opening the window?
 - D. Have you forgotten to open the window?
53. He is not too young to marry her.
- A. He is so young that he can't marry her.
 - B. He is old enough to marry her.
 - C. He is too old to marry her.
 - D. She is so young that she can't marry him.
54. Unless you work quickly, you won't finish this exam.
- A. You won't finish this exam if you work quickly.
 - B. You will finish this exam even if you don't work quickly.
 - C. You are working so quick to finish this exam.
 - D. You won't finish this exam if you don't work quickly.
55. He is not as tall as her brother.
- A. He is taller than her brother.
 - B. Her brother is taller than her.
 - C. He is shorter than her brother.
 - D. His brother is shorter than her.

2

2

56. Neither the boy nor the girl will do it.

- A. Both of them will do it.
- D. One of them will do it.
- D. Both of them will not do it.
- D. Either of them will do it.

Direction: Item 57-60, are about sentence jointing. For each question choose the best combination from the given alternatives.

57. Ali can't swim. His sister can't swim.

- D. Ali can't swim and so can't his sister.
- D. Ali can't swim; neither does his sister.
- D. Ali can swim but his sister can't.
- D. Ali can't swim and his sister can either.

58. I went to visit Hana. She was not at home.

- D. I went to visit Hana and she was not at home.
- D. I was not only went to visit Hanna but I couldn't find her at home.
- D. I went to visit Hana so she wasn't at home.
- D. Although I went to visit Hana, she was not at home.

59. I met the man. You spoke to him.

- D. I met the man which you spoke to.
- D. I met the man that you spoke to.
- D. I met the man you spoke.
- E. I met the man whose you spoke to.

2**2**

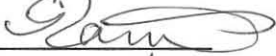
60. He gave Mary the keys. He had hidden them in the cupboard.

- A. He gave Mary the keys which he had hidden in the cupboard.
- B. He gave Mary the cupboard with the keys.
- C. Mary gave the keys to him which she had hidden in the cupboard.
- D. He gave the keys which Mary had hidden in the cupboard.

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 that all sources of materials used for this thesis have been duly acknowledged.

Name: Samuel Dossa

Signature: 

Date: 25/05/2011