

Addis Ababa University
School of Graduate Studies

EFL Teachers' Expressed Beliefs and Actual Practices in Testing
Reading Skills: The Case of Grade 10 in Three
High Schools of East Gojjam Zone



BY
MINWAGAW TEMESGEN

JUNE 2009
ADDIS ABABA

**EFL TEACHERS' EXPRESSED BELIEFS AND ACTUAL
PRACTICES IN TESTING READING SKILLS: THE CASE OF
GRADE 10 IN THREE HIGH SCHOOLS OF
EAST GOJJAM ZONE**

Minwagaw Temesgen

**A Thesis Submitted to the School of Graduate Studies
Institute of Language Studies
Department of Foreign Languages and Literature**

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

JUNE 2009

ADDIS ABABA

Addis Ababa University
Institute of Language Studies
Department of Foreign Languages and Literature
(Graduate Programme)

EFL Teachers' Expressed Beliefs and Actual Practices in Testing
Reading Skills: The Case of Grade 10 in Three
High Schools of East Gojjam Zone

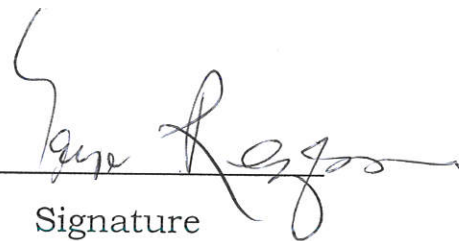
Minwagaw Temesgen



Approved by:


Taye Regassa (Ph. D)

Advisor


Signature

ASSEFA ZERU (PHD)

Examiner


Signature 19 June '09

DEDICATION

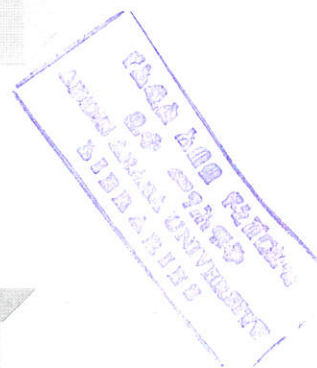
For the memory of:

My father-in-law, Kegn Geta Achamyehel Liyew

O father! What did God reveal to you
before you went to Wonkeshet Gebriel
Monastery?

And my uncle, Ato Fentahun Nigatu

Why did you prefer only birds of the sky
to attend your funeral ceremony, Aya
Fentie? Were we that much cruel to you?



Acknowledgements

First and foremost, glory be to God the Almighty for letting me begin in peace and end in peace. Praise also be to Hail Virgin Mary and all the saints, in whose intercession I have reached this moment.

Seeing me carrying my bag on my back, my classmate, Mebratu Mulatu, once asked me where I was going. “To the doctor to get my baby examined,” I jokingly replied. I was right. My ‘first-born’ (i.e. my thesis) was seriously sick of inconsistencies and inadequacies. It was through Dr. Taye’s meticulous treatment that it has ‘stood on its own feet’ healthily. Therefore, I would like to express my heart-felt gratitude to Dr. Taye Regassa, my advisor, for his critical suggestions and constructive comments, starting from shaping the title to the completion of the whole paper. The material he lent me also helped me a lot. I was lucky having him as an advisor to get comments and lessons in a friendly, non-threatening manner.

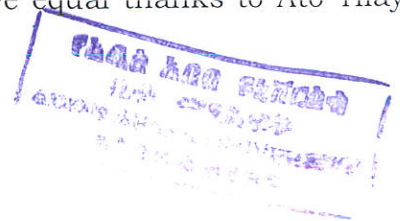
I also express my gratitude to Dr. Teshome Demisse, who was my instructor for the course *Language Testing* last year. He has surely opened my eyes to see what have been planted in the fields of language testing; I might not have seen as he thought me to, though. It was, however, because of his input that I dared to study some aspects of a reading skills test. He really made me remind what Ahrham Maslow once said: “My teachers were great in the world.” Yes! Mine are great, too!

My thanks also go to Kesis Sintayehu Abate, who has always been encouraging me to *learn both personally and professionally*. Here I am, Kesis. I will go on *learning* in the will of God.

My wife, Yirgedu Achamyeh, deserves great thanks for encouraging me along with carrying the load of rearing our daughter, Leamlak Minwagaw, ALONE. You have really showed me your courage and determination towards my success. Your patience and commitment has driven me up the ladder of my academic career.

My gratitude, further, goes to Ato Temesgen Nigatu and W/ro Tihun Lingerih, my ever-caring parents, for their moral support, advice and prayer for my overall journey in life.

Melkam Getnet and Mengesha Bogale, my friends, thank you for the great lessons you taught me in our stay in the campus together. I also owe equal thanks to Ato Tilaye



Alemneh, my friend, for his proof reading and encouragement. I wish you good luck to reach the same time next year, my dear.

Finally, I extend my indebted thanks to W/rit Eden Habte for typing my thesis efficiently and patiently. I know that my *handwriting* and this year's irritating electric power supply of the city tempted you. But you never showed me any change of welcoming expression. I learned so much from you. Thank you very much, indeed!

Table of Contents

Acknowledgements	i
List of Tables	v
Abstract	vi
CHAPTER ONE: INTRODUCTION	1
1.1 Background of the Study	1
1.2 Statement of the Problem	4
1.3 Objective of the Study	7
1.3.1 General Objective	7
1.3.2 Specific Objectives	8
1.4 Significance of the Study	8
1.5 Scope of the Study	9
1.6 Limitation of the Study	9
1.7 Operational Terms and Abbreviations	10
CHAPTER TWO: REVIEW OF RELATED LITERATURE	11
2.1 Language Teaching and Testing: An Interface	11
2.2 Teaching and Testing Reading Skills	15
2.3 Testing Reading Skills: Basic Considerations	17
2.3.1 The Nature of Reading: Unitary or Multi-divisible?	17
2.3.2 Approaches to Testing Reading Skills.....	18
2.3.2.1 Discrete-point Testing	18
2.3.2.2 Integrative Testing	19
2.3.2.3 Communicative Testing	20
2.3.3 Characteristics of a Good Reading Test	21
2.3.3.1 Test Validity	21
2.3.3.2 Test Reliability	23
2.3.3.3 Test Efficiency	23
2.3.4 Sub-Skills in Reading: Where Validity Resides	24
2.3.5 Test Methods	31

CHAPTER THREE: RESEARCH DESIGN AND METHODOLOGY 35

3.1 Research Setting and Subjects of the Study35
3.2 Data Collection Instruments36
3.3 Data Collection Procedures36
3.4 Data Analysis Technique37

CHAPTER FOUR: PRESENTATION, ANALYSIS AND

INTERPRETATION OF DATA 38

4.1 Analysis and Interpretation of Data Elicited through
Questionnaire: Parts One and Two38
4.1.1 Data from Questionnaire Part One.....38
4.1.2 Data from Questionnaire Part Two43
4.2 Analysis and Interpretation of Data Elicited through
Interview.....48

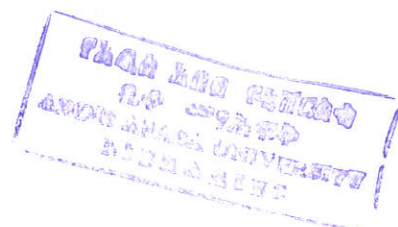
CHAPTER FIVE: SUMMARY, CONCLUSIONS AND

RECOMMENDATIONS 55

5.1 Summary55
5.2 Conclusions56
5.3 Recommendations57
BIBLIOGRAPHY59

APPENDICES:

Appendix 1: Questionnaire Part One64
Appendix 2: Questionnaire Part Two66
Appendix 3: Interview Questions68
Appendix 4: AASS Reading Skills Test69
Appendix 5: ASS Reading Skills Test71
Appendix 6: DMSS Reading Skills Test74



List of Tables

Table 1: Location of the Schools Taken as Sample	35
Table 2: Personal Information of EFL Teachers in the Sample	
Schools	38
Table 3: Sub-skills Believed to Have Been Taught by EFL Teachers...	40
Table 4: Number of Teachers who Agreed or Disagreed to Teaching	
the Sub-skills in Each School	42
Table 5: Test Items and Respective Sub-skills Tested as Judged by	
AASS Teachers	43
Table 6: Test Items and Respective Sub-skills as Judged by ASS	
Teachers	44
Table 7: Test Items and Respective Sub-skills as Judged by DMSS	
Teachers	47
Table 8: Test Methods Used in the Sample Schools	47

Abstract

The purpose of the study was to see if EFL teachers tested a representative sample of the reading sub-skills they believed have been taught in the classroom last semester. The study also tried to find out whether or not there was agreement among EFL teachers on assigning a reading sub-skill to a typical test item. Investigating what test method(s) was/were used was another focus of the research.

The study was conducted on the expressed beliefs and actual practices of high school EFL teachers in testing reading skills. 15 EFL teachers from three high schools of East Gojjam Zone, Amhara Region, were taken as subjects of the study. The reading skills test used in the analysis was that of 2001 E.C., grade 10 first semester.

*To collect data on teachers' beliefs and actual practices in testing reading skills, a two-part questionnaire which contained a list of sub-skills in reading was dispatched to all the 15 Grade 10 EFL teachers. **Part one** of the questionnaire was used to elicit data whether the teachers taught the reading sub-skills. **Part two** was used to check if the sub-skills taught were tested representatively in the achievement test of reading teachers constructed last semester. A semi-structured interview was also used to elicit data to supplement the questionnaire.*

*The data collected through these instruments were analyzed qualitatively, and the findings showed that high school EFL teachers did not test a representative sample of the sub-skills they claimed having taught in the classroom. It was also indicated that the reading sub-skills tested as a whole focused on lower-level ones. Among these lower-level sub-skills **referencing, vocabulary and explicitly expressed idea** were found to dominate the test batteries. EFL teachers also failed to agree in assigning the items which tested the listed sub-skills other than **vocabulary** and **reference**. Regarding the test methods teachers used, it was found out that only a single test method-multiple-choice-dominated the whole tests in each of the sample high schools.*

Finally, based on the findings of the study, conclusions were drawn and recommendations were made.

CHAPTER ONE: INTRODUCTION

1.1 Background of the Study

Gathering information regarding students' achievement is an integral part of instruction, and one means of gathering such information is through the use of tests. Testing should, therefore, be a normal and well-understood activity but that is often not the case. (Schreiner, 1979:v).

Teaching and testing are inextricably linked in the world of language teaching. Teachers teach and learners learn, but both *test-teachers* their method effectiveness; and learners, their learning. Thus, testing has an informative value for both groups. Using the test data, teachers can revise their methodology or area of emphasis. Learners as well get motivated or depressed depending on the tests (and scores) and their direct/indirect attachment to the courses taught or the real language use context. The impact of testing on teaching and/or learning in this regard is referred to as washback or backwash (Hughes, 1989; Brown, 1987; Weir, 2005).

Language instruction has three aspects-the *teaching*, the *learning* and the *testing*. The tests, in turn, have three demands, according to Davies (1991:1)-“the *language*, the *learning* and the *measurement*.” The *language* is *what is taught and tested*; the *learning*, the *washback*; and the *measurement*, the *reliability* over time and condition. Therefore, testing language means teaching for teachers and a way of learning for learners.

Weir (1990:15) clearly puts what teaching and testing mean in the communicative paradigm, taking the presence and absence of teachers' and peers' help in the teaching and in the testing respectively. He asserts that

The only necessary difference between teaching and testing ... relates to the amount of help that is available to the candidate from his teacher or his peers. The help that is normally available in the teaching situation... is removed in a test... [Therefore], the test may be viewed as an intermediate stage between the world of the classroom and the future target situation where the candidate will have to operate unaided (Emphasis mine).

The above expression goes with what Atkins et al. (1996:148) say “[testing is to teaching as teaching is to learning.]” To make testing serve a teaching purpose, the materials used (i.e. texts and test methods) need to resemble the real-world ones. This is what many scholars refer to as authenticity (Grellett, 1981; Nuttall, 1996; Cooper and Fairbairn, 1989; Weir, 1990, 2005).

With an English language teaching test, especially in a test of reading where the texts selected and questioning techniques used are aids for better understanding, making them replicate (simulate) the real-world reading is a vital matter. This helps learners develop the relevant enabling skills in reading.

According to Grellett (1981:7), in devising activities and test questions, it is very important to use authentic texts whenever possible. This is because reading tests based on real-world-like texts give students opportunities to tackle the problem of comprehension in any setting.

In addition to reading texts, test methods also affect the quality of a reading skills test. What is mostly practiced in schools and national examinations is a multiple-choice format. The format is criticized for being liable to guessing. Pyrczak (1975) refers this aspect of the multiple-choice format *text-independence*, for the answer can be supplied without reading the text (see also Duffelmeyer, 1980; Weir, 1990; Hughes, 1989).

It is also unnatural for learners to get reading texts with multiple-choice questions both in the academic reading and reading for pleasure. As a result, many teachers regard these types of questions associated with testing not teaching. They do not reflect an authentic native speaker response to a certain text (Cooper and Fairbairn, 1989). Thus, though the question format is not totally avoided, other test methods, i.e. free response, short answer, cloze, information transfer, etc. are recommended (Hughes, 1989; Weir, 1990, 2005).

Selecting the right test methods for a reading test has a twofold relevance: testing general comprehension and testing the sub-skills in reading. Regarding the sub-skills in reading, there has been a controversy for ages. Some scholars

argue that reading is a unitary competence which can be taught and tested as one overall skill. Others view reading as having multi-divisible nature. The latter argument seems prominent these days, though (Alderson and Lukmani, 1989; Weir et al., 1990).

The argument rather shifted to the issue of the number and level of the sub-skills in teaching and testing reading. Some say learning lower level skills (e.g. guessing unfamiliar words) is a pre-requisite for the acquisition of higher order ones (e.g. inference). For the others, learning higher order skills does guarantee the acquisition of the lower ones (Weir and Porter, 1994; Alderson, 1990a, b).

Whichever direction the argument goes, the language teachers should have the ability to test a representative sample of sub-skills listed by Munby (1978). Research, however, proved that language teachers (testers) differ in their views to treat the different sub-skills (and in the methods they employ) to test the sub-skills (see Alderson, 1990a; Alderson and Lukmani, 1989; Weir et al., 1990).

While most scholars agree to the issue of teaching and testing the different sub-skills, they differ in the number and order of the sub-skills. Munby (1978), for example, identified more than eighteen sub-skills, while Davies (1968), cited in Alderson and Lukmani (1989), grouped them into eight. The same skills for Gray (1970), cited in Alderson and Lukmani (1989), are summed up into three: reading the lines, reading between the lines and reading beyond the lines (see also Atkins et al., 1996; Taye, 1999; Dejene, 1994).

Whether eighteen or so, or eight or three, what should be taken into account is to teach and test the skills in as a representative sample as possible, making the exam “more communicative, more authentic..., and more realistic in terms of students’ needs and expectations” (Cooper and Fairbairn, 1989:309). A valid test of reading comprehension test, thus, should be a representative sample of the higher order and lower order sub-skills (Weir et al., 1990). How? Matthews

answers like this: “If texts are carefully chosen and tasks are sufficiently global, then the relevant enabling skills will be naturally sampled” (1990:515).

This representative sampling, in turn, will be achieved, as mentioned earlier, using a variety of test methods. “Good reading tests are likely to employ a number of different techniques... since in real-life reading, readers typically respond to texts in a variety of different ways.” (Alderson, 2000, cited in Weir, 2005:119).

Especially in testing the reading skills, many English language teachers were found to be incapable of constructing items themselves to test the sub-skills. And they were not able to match the different sub-skills to the already written items. As Alderson (1990a, b) reported, language teachers lack the competence to “test” the reading sub-skills appropriately. Though Weir et al. (1990) criticized Alderson’s methodology and his ‘ambiguous’ definitions of higher- and lower-level sub-skills, they seemed to admit that Alderson was right in his report on testing reading comprehension.

The case is not different in Ethiopia, where English is given as a foreign language school subject from grade one up to colleges and universities. Though reading passages are in every unit of the textbooks in junior and high schools, teachers do not seem to teach and test them as the literature dictates. If it were not the case, many research findings would not prove that students at the high schools and college levels are poor comprehenders (Zelege, 2007; Haile, 2007; Sileshi, 2007; Wogari, 1995). Of course the learners themselves and the learning/teaching materials could also contribute to the low comprehension ability of students.

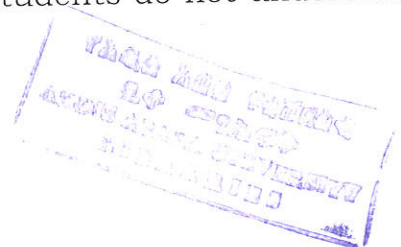
1.2 Statement of the Problem

Teachers who teach English as a foreign language in Ethiopian high schools are supposed to make learners use it for communication. Even though there is little opportunity for students to encounter native speakers outside the classroom, they are expected to apply the language for academic matters. In this regard,

therefore, teaching reading skills needs critical attention; the more competent students are at reading skills in English, the better their understanding of other school subjects at the high schools or colleges and universities will be. It goes with what (Teshome, 1995:4) says: "If individuals have a good command of language, they may be at an advantage to do well in their academic performance".

In spite of the efforts made to teach reading at the high schools, colleges and universities, most Ethiopian students have been found to be low comprehenders. This may be caused by a variety of reasons: the structure of the textbooks, the type of reading texts and questions, the teachers' area of emphasis while teaching, etc. (Nuru, 1992; Mekonnen, 2005; Getachew, 1996). The writer of this paper also argues that the language teachers' testing method and the belief they have about reading skills have an impact on the issue. To what extent do English language teachers test the students' competence of the sub-skills in reading?

Regarding the level of questions in textbooks, and ESLCE and classroom exams, Nuru (1992) proved that the Grade 11 and 12 textbooks along with comprehension questions and classroom exams focus on lower level skills, i.e. skills which demand only literal comprehension of a text. Another research by Mekonnen (2005) supports Nuru's idea in that grade 10 textbooks focus on lower-level cognitive questions. Getachew's (1996) finding is also similar. He further found out that reading skills are not at all taught; the teaching of content and vocabulary at the expense of reading are given much emphasis. Sileshi (2007) also ascertained that the present day preparatory students are poor at comprehension. Wondifraw (2005), in his study of comparing text difficulty of biology, history, geography, and English textbooks, indicated that English texts are relatively easier than others; the students do not understand the English texts for sure, though.



All the above studies, except Nuru's, focus on the teaching of reading to the high schools. None of the studies mentioned whether the texts provided in the textbooks or test papers were interesting and authentic so that students could read them with motivation to develop the sub-skills in reading.

In another research, Desalegn (2000) reported teachers should use a variety of test techniques, such as free response, so that females would also get opportunities as males. In his study, males outscored females in multiple-choice items and the vice versa in free response ones. Wogari (1995), studying the impact of the multiple-choice and the short answer formats on high school students' comprehension, indicated students scored lower in the comprehension parts of both test formats than their respective structure parts. This is another proof for the students' low comprehension ability.

The real causes of this low comprehension ability of students have not clearly been identified so far. Zeleke (2007) suggested limited proficiency in English, lack of practice of extensive reading, lack of motivation to read the passages, failure to adjust their reading with the text structure and the teachers' failure to encourage students to become efficient comprehenders as factors.

However, in none of the studies above and others (to the extent of my knowledge), no thorough study has been conducted on EFL teachers' expressed beliefs and actual practices in testing reading skills. Testing as a teaching tool seems to have been neglected. It is this neglect that made the present study focus on investigating the testing of reading skills at the high schools. The way teachers believe about their competence in reading test construction and their practice on that plays a great role on students' performance. If teachers, for example, write questions which demand learners only to "read the lines," this does not help learners. As Taye (1999:57) puts "real understanding requires the reader to be able to paraphrase the context of what he/she [the student] has read in his or her own words and sentences rather than produce a mere

repetition of the writer's words." This is so because real-world reading does not invite such copying; things are not all the same everywhere.

In order to make learners practice detailed comprehension, the classroom EFL teacher should use a variety of test techniques and should present authentic texts which are exploitable to test the expected reading sub-skills mentioned in Munby (1978). Weir (1990:12) recommends the importance of using different formats in a reading skills test. He says "[f]ormat effect relates to the possibility that test results may be contaminated by the test format employed, i.e. a different estimate of a skill such as reading might be obtained if a different format is employed." Therefore, sticking oneself to very limited methods is a disadvantage.

The texts selected and the test techniques used need critical attention because they are basic for the testing of reading sub-skills, which are the blood (and vein) of the test. Alderson and Lukmani (1989) and Alderson (1990a) found out that teachers at Lancaster, London, were unable to match the sub-skills in Munby (1978) and the test items constructed at Bombay, India. Can teachers match a sub-skill with a typical test item constructed by themselves in Ethiopian high schools? This was the main focus of the present paper.

To this end, the present study was designed to see EFL teachers' beliefs about what and how to test reading. The study tried to check if their tests were reflections of their beliefs.

1.3 Objective of the Study

1.3.1 General Objective

The main objective of the study, as mentioned above, was to investigate if high school EFL teachers' beliefs about testing reading skills were reflected on the final achievement tests they wrote.

1.3.2 Specific Objectives

The research, to meet the main objective, tried to:

- a. find out whether or not high school EFL teachers tested the sub-skills they claimed having taught.
- b. see whether or not teachers tested the different sub-skills of reading representatively.
- c. examine to what extent teachers agreed in assigning sub-skills to typical test items in an achievement test of reading, and
- d. investigate what type of test methods they used.

1.4 Significance of the Study

This study, in light of the findings drawn, may have the following significance for teachers, researchers and test materials developers.

For teachers, primarily, it may help them to reconsider testing as one part of teaching. It is not only through what they present in the classroom as reading lessons can teachers teach students, but the tests they construct (texts they select and methods they employ) can also help a lot. This, in turn, will have a positive washback on students' learning the reading skills.

For researchers, the finding will probably give them an insight on how to approach a reading skills test problem, taking the teachers, the texts and the test methods that elicit competence on sub-skills as focus areas for further research.

Finally, the test materials developers, especially test experts at the national level, may get the finding helpful. It will serve as a ringing bell to take great care in developing test items and in selecting test texts, for the items they construct and the sub-skills they focus influence school teachers so much.

1.5 Scope of the Study

The research was conducted in East Gojjam Zone of the Amhara Region. Out of 22 high schools in the Zone, only three (Abreha Woatsibeha, Amanuel and Debre Markos) high schools were taken. The study treated only the sub-skills included and test methods used in a reading skills test in the three sample schools. It did not investigate the type of texts selected for the reading test.

1.6 Limitation of the Study

Obviously, it is not possible to generalize about all the high schools in the zone based on three schools only. Therefore, the research will have weaknesses in treating only three schools. It would also be good if both the rural and urban (i.e. zonal) high schools were included. This would help the researcher see whether there comes a difference in testing belief and practice because of the setting which teachers have been working at. This limitation happened because in both the urban and rural schools, some reading exams were found to use very short passages (a paragraph) and few questions (about 5 only).

Besides, all or half of the schools in the zone would be taken if there were plenty of time, material and human resource to collect data. The financial constraint to do all these is another factor. The researcher hopes that succeeding researchers will have more to say about this problem in a wider context.

1.7 Operational Terms and Acronyms

Actual practices: The manifestations of teachers' beliefs reflected on classroom test papers

Expressed beliefs: Teachers' self-understanding of their abilities in test construction, *representing the content taught*

Reading skills test: A test which is designed to test the possible sub-skills of reading skills

AASS: Abreha Woatsibeha Secondary School

AAU: Addis Ababa University

ASS: Amanuel Secondary School

DMSS: Debre Markos Secondary School

EFL: English as a Foreign Language

ESLCE: Ethiopian Schools Leaving Certificate Examinations.

CHAPTER TWO: REVIEW OF RELATED LITERATURE

2.1 Language Teaching and Testing: An Interface

Whatever we do, wherever we are, and whoever we are with, we always experience something, and we react in some way. The reaction is reasonable and systematic. This is a proof for our character as rational beings. Eating at a restaurant with people, for example, we appreciate or dislike the taste of the food immediately we test it, in the mid of the eating or at the end. As a result, we may have a chance to change the dish as a whole, make improvements (adding salt, for example) or finish it with little dissatisfaction. We feel and react similarly about the restaurant and the people we are eating with. It seems from this common experience that Brown (1987:218) says the following.

Whether we realized it or not we test every day in virtually every cognitive effort we make. When we read a book, listen to the news on TV, or prepare a meal, we are testing hypotheses and making judgments.

Brown refers the experience of reacting to a situation as testing hypotheses. Likewise, in the field of education, especially in language teaching and learning, teachers teach some content from the syllabus to students in a classroom. It is like eating food with people in a restaurant, where the food is the teaching content; the restaurant, the classroom; and the people, the pupils. In so doing, the teacher makes hypotheses about the teaching in two respects. Firstly, he judges whether his invitation of that sort of “food” is suitable for the people. Secondly, the way the food is served and the effect it has on the invitees is evaluated. Both can be referred to as *content evaluation* and *methodology evaluation* respectively. The invited people also make hypotheses in a similar way.

In the process of language acquisition in general and classroom teaching in particular, making hypotheses and testing their worth is indispensable (Richards, 1974; Brown, 1987). Second language learning is perfected in the

process of continuous, approximative hypotheses formation of the target language system. This mental process of approximating the target language system is termed as *interlanguage*, a term first coined by Selinker (1967), in Brown (1987).

In the same way, in a classroom teaching, there are manifestations of interlanguage, among which the use of tests is one. As interlanguage occurs between two languages (the learner's mother-tongue and the target language), testing happens between what the learner has learned so far and what s/he is expected to acquire.

Weir's (1990:15) idea proves the above point clear. He says "...the test might be viewed as an *intermediate stage* between the world of the classroom and the future target situation where the candidate will have to operate unaided." (italics mine)

Just like interlanguage in which "second language learners are forming their own self-contained linguistic systems" (Brown, 1987:169), tests are seen as "a way of finding out or demonstrating how well they [learners] are progressing, or in order to seek some reassurance that they are keeping up with their peers" (Pearson, 1988:104). Morrow (1979) also has a similar vein. He says "...the role of the test is to show how far it [testing] has moved towards an approximation of a native speakers' system." (p. 145). Though interlanguage data are manifested in productive skills, receptive skills also undergo the process. This process is best expressed by Morrow (1979) as: "Tests of receptive skills will ... be concerned with revealing the extent to which the candidates processing abilities match those of a native speaker." The similarity between interlanguage and testing is clearly seen in Davies (1991:1) and Genesee and Upshur's (1996:141) explanation of the "demands" and "constituents" of a test respectively.

Davies characterizes language instruction as having three aspects: the *teaching*, the *learning* and the *testing*, while "tests have three demands: the *language*, the

learning and the measurement." (1991:1). Genesee and Upshur claim that "tests contain three things: the *content*, the *method* and the *measurement*" (1996:141). From these assertions, we can see that "learning", which is taken as the *backwash* by Davies, is common for both the "teaching" and the "testing". The only difference is its replacement by "method" in Genesee and Upshur's statement. This is because testing is a *method* through which learning takes place. Because of this Brown (1987:219) defines testing as "a *method* of measuring a person's ability or knowledge in a given area" (my italics). For Brown three things also characterize testing: a test is a method; it has a purpose of measuring; and it measures a person's ability or knowledge.

If a test has a content (Hughes, 1989), is a method (Brown, 1987) and has a purpose of measuring (Davies, 1991; Genesee and Upshur, 1996), we can say that testing has a similar status to teaching, where the only difference lies to the amount of help available in teaching, but not in testing (Weir, 1990). This is because tests discriminate among learners for the purpose of placement, diagnosis, proficiency or achievement (Alderson, 1996; Pearson, 1988; Brown, 1987). Tests also have consequence which is termed as washback or backwash (Brown, 1987; Atkins et al., 1996; Alderson et al., 1995; Hughes, 1989; Weir, 1990, 2005; Bachman, 1990).

Weir (1990) puts the relationship between teaching and testing in clear terms analogous to a dog and its tail. According to Weir, two decades ago, where discrete-point and integrative testing dominated the testing domain, it was the teaching that was believed to direct the testing, just like the dog wagging its tail in any direction the dog wants. However, the belief about the relationship between language teaching and testing has changed since the 1970's, when the idea of communicative language teaching emerged (Richards and Rodgers, 2001; Brown, 1987; Richards, 2001). "...The 1970s saw a change from teaching language as a system to teaching it as a means of communication" (Weir, 2005:9).

As a result, testing has got a different status; it is believed to direct the teaching. Following this, Hughes (1989:2) comments that “[w]e cannot expect testing to follow teaching... it should be supportive of good teaching and... *exert a corrective influence* on bad teaching” (italics mine). As a result of this, Weir (1990:27) tells us a possibility of saying “a dog-wagging tail” instead of “a tail-wagging dog.” He states that

...the tail may wag the dog in that a communicative approach to teaching is more likely to adopt when the test at the end of a course of instruction is itself communicative.

Generally, a test can be taken as a self-contained syllabus (having content, methodology and measurement) (Richards, 2001). It is another form of teaching content through a method to bring about learning.

Furthermore, testing is a good opportunity for teachers to adapt a language teaching material in a manageable fashion. If the syllabus is malformed, the teacher can adapt it through tests, at least. A reading material which uses inauthentic texts and tasks, for example, can be adapted using tests. This means that “... a good test will exert a beneficial influence upon a bad course” (Pearson, 1988:103). A dog-wagging tail, that is!

Whether the dog wags its tail or the tail wags the dog, what matters is having a “perfect” dog. As the tail alone is not called a dog, a dog without tail is a tailless dog. In the same way, to make teaching-learning process successful, teaching and testing should act in a similar way; they should both be for learning. “Teaching is to learning as testing is to teaching” (Atkins et al., 1996:148).

The relation between teaching and testing is also expressed by Underhill (1982:18), cited in Teshome (1995:3), saying “AS YE TEACH, SO SHALL YE TEST”. Teshome himself takes assessment (testing) as a “necessary twin to language teaching” (Ibid.). This goes with what Pearson (1988) recommends. “It will be suggested that one way to shed at least some light ...is to consider tests... as teaching-learning activities, and vice versa” (p. 103).



2.2 Teaching and Testing Reading Skills

The interface between teaching and testing works for the four macro skills (listening, speaking, reading and writing), grammar and vocabulary. These are areas from which we derive contents to teach and test. The whole of this paper deals with testing reading skills. However, we shall start with teaching the skill; it is from what we teach that we can write representative test items. Otherwise, it will be “a far cry” from the danger (Weir, 1990; Pearson, 1988). A valid achievement test of reading should be based on the taught content (sub-skills). “If such skills are teachable, then they are probably testable.” (Alderson, 1996:219) The skills are taught based on texts presented for a purpose.

In real life people read different things for different purposes. They read for information or for pleasure or for meaning in a text (Grellet, 1981; Nuttall, 1996). We read a lot of things, ranging from newspapers through textbooks to novels (Grellet, 1981; Williams, 1984; Nuttall, 1996). The purpose of our reading determines what we read and how. The purpose we read a newspaper for and the way we read it is quite different from reading academic texts. In the former, we read for information or for pleasure very quickly, but in the latter, we read for meaning slowly and critically.

Therefore, teaching reading skills needs a greater emphasis at the high schools, where academic texts are studied. As “language improvement is a natural by-product of reading” (Nuttall, 1996:30), reading ability is also recognized as centrally important because with its absence very little academic learning can take place (Teshome, 1995; Abdu, 1993; Williams, 1981).

However, in Ethiopia, where English is taught as a foreign language, teaching the skill and the way it is tested does not seem to be in a good condition. Although there is no local research conducted regarding the status of testing reading skills (to the extent of my knowledge), the teaching is proved to be ineffective. (See section 1.2 for details.)

Getachew (1996), in his study about the teaching of reading in government high schools, found out that reading as a skill is not taught; the teaching of content at the expense of reading skills is emphasized. This finding goes with what Nuttall comments saying “[g]iving a lesson [content] based on a text is not the same as giving a reading lesson” (1996:30). A reading lesson, according to Nuttall is used in a classroom to

enable students to enjoy (or at least feel conformable with) reading in the foreign language, and to read without help unfamiliar authentic texts, at appropriate speed, silently and with adequate understanding (p. 31).

This role can be achieved by choosing suitable texts to work on and by devising effective tasks and activities to elicit the underlying sub-skills of reading. This idea is best put by Grellet (1981:3) as she says, referring to Munby (1978), “[t]he above skills, (Munby, 1978), question types and question functions are constantly related since a given exercise uses a certain type of question, with a certain function, to develop a particular reading skill.”

As it has been mentioned earlier, EFL teachers failed to help learners be good comprehenders. Why do the teachers fail to help learners? Is it because they are unable themselves or is it, as Nuttall (1996:32) says, because the teachers believe that “reading cannot be taught, but learnt”?

How can reading be learnt if it is not taught? This is a misconception. Reading can both be taught and tested to be learnt. Reading defined as “the activation, application, and interaction of decoding, encoding, and comprehension processes that results in knowledge gain from something written or printed” (Lee, 1997:152), can be taught using activities which allow learners to interact with the message in the text using their background knowledge. “Comprehension involves both reader-based and text-based factors” (Ibid.). But teachers at the high schools seem to fail to teach or test reading as the literature dictates (Haile, 2007; Sherif, 2007; Sileshi, 2007; Wondifraw, 2005).

As has been mentioned earlier, the reading teacher can use test materials for teaching purposes. It is, therefore, reasonable to use communicative learning activities as a test, and the vice versa. It is only practical and finical constraints that can be seen to limit testing to a small sub-set of teaching-learning activities (Pearson, 1988:106).

2.3 Testing Reading Skills: Basic Considerations

2.3.1 The Nature of Reading: Unitary or Multi-divisible?

Reading specialists argued for and against the nature of reading. While some say that reading is a unitary competence, others argue against this issue. Alderson (1990) takes reading as a unitary competence. He derived this statement from a research he conducted in 1990 on reading teachers. He found out that teachers (he called *judges*) differed in assigning a typical skill to a test item. Again the subjects were asked to identify which skill a typical test item was testing, which again resulted in inconsistency. Alderson cited Lunzer et al. (1974) as his supporters saying “they [Lunzer et al.] conclude that reading, at least as carried out when taking reading tests, can be considered to be one single aptitude.” (p. 427).

Alderson’s conclusion is unreasonable as it is not possible to call a human being a complete being without taking the different body parts separately. The different parts make him human; we see the person with all the parts. Of course, the sum of the different parts does not equate the complete being. It is unity with diversity or diversity within unity. “The whole is greater than the sum of its parts” does not mean that the “whole” does not have separate parts.

In the same way, reading as a “unitary” competence constitutes different sub-skills which, when added, do not equate reading. As the way one reaches the correct answer needs a multiple interaction of skills, it is undeniable that reading is multifaceted. Reading comprehension requires the reader beyond knowing how to read words and sentences. Lee (1997:152) asks a basic

question: "Can people believe that they can teach others to read based solely on the fact that they know how to read?" Reading is beyond knowing individual letters, words and sentences. Otherwise, "differences exist between textual content and what a reader extracts from a text" (Ibid.); to make meaning of a certain text, readers use the textual information as a base.

Other researchers (Matthews, 1990; Weir et al., 1990; Weir and Porter, 1994) argue that reading has various component sub-skills on which course design, teaching, and test and materials development are based. Alderson in a research prior to the 1990's also seems to contradict with himself. In 1989, he talked about Gray's classification of questions reading the lines, reading between the lines and reading beyond the lines- which in turn tell us about the divisible nature of reading (Alderson and Lukmani, 1989; Weir and Porter, 1994). Reading, therefore, can be broken down into "...underlying skills components for the purposes of teaching and testing" (Weir and Porter, 1994:2). Following are beliefs about the testing of reading at different times.

2.3.2 Approaches to Testing Reading Skills

"Seeing where we have come from... helps us better understand where we are today" (Weir, 2005:5).

Language teaching and testing have manifested changes for ages depending on the theory of language, teaching and learning (Brown, 1987; Weir, 1990, 2005; Heaton, 1988). Below is a brief presentation of approaches to language testing in general and the testing of reading in particular. "Language testing reflects language teaching, which in turn reflects a particular model of competence." (Spolsky, 1975:70).

2.3.2.1 Discrete-point Testing

This is associated with the teaching of reading for the sake of language learning. Language learning in turn was viewed as learning the culture in which the language was used. Williams (1984) refers to this as "reading for language." In here the learner was seen as a passive recipient of information in the text. He

was “spoon-fed with selected language rather than given a chance to reflect upon it in his own time and try to work things out for himself” (Ibid:15).

The testing of reading, as a result, focused on phonology, morphology and syntax of the language independently, assuming that the sum of the different parts equate reading competence (Brown, 1987; Weir, 1990, 2005). However, reading is beyond understanding individual letters, words and sentences. It requires the learner to bring his background information to extract meaning from the sentences as a whole (Nuttall, 1996).

Let us take an example from Nuttall (1996:20-21) to demonstrate that understanding words or the entire sentence does not mean getting what the sentence means.

Sentence: *Examination results are misleading.*

Discretely, this sentence can be used to test the structure of a simple sentence (statement), having *examination results* as subject where *are* and *misleading* are *a verb* and *a complement* respectively. However, in reading it does not give sense because the proposition is liable to doubt, question or argument. Therefore, it needs a certain context in which it was said. This contextual use of language is the concern of integrative testing.

2.3.2.2 Integrative Testing

Integrative language testing “...denies the atomistic nature of language as a basis for language testing” (Morrow, 1979:148). While discrete-point testing tries to put elements of language apart, integrative testing puts them back together (Weir, 1990). The example given above does a certain function in this approach when used in context. Nuttall herself explains the communicative value of the sentence when used as a plea or complaint by a mother to the head teacher who is proposing to expel her son. As a result, the sentence does have this purpose if used interactively with another sentence as follows.

You should not expel my son just because he has failed. Examination results are misleading. (p. 20-21)

2.3.2.3 Communicative Testing

The communicative testing of reading goes with the top-down processing approach to reading in general. Unlike the discrete-point testing which focuses on bottom-up approach, communicative testing relies on understanding of the meaning through interaction of the schema in the readers' head and the information in the text. It requires active interaction of the reader and the writer through the text (Williams, 1984; Grellet, 1981; Nuttall, 1996).

Although the importance of knowing a language in order to read it has been stressed, this is a means to an end (Williams, 1984:16), or it is a necessary but not a sufficient condition for communication (Weir, 1990). What matters is not adding bricks on bricks but to have a complete service-giving building. Williams puts it as follows:

Rather than seeing reading as a matter of 'stamping in' selected pieces of language, the communicative approach sees language learning as a development in the learner's language ability, which occurs as the learner carries out relevant tasks. The stress is on 'growth' on the part of the learner, rather than 'building' on the part of the teacher (1984:17).

Generally, reading skills are assumed to be "identified, taught, tested and researched" (Alderson, 1990:428). Hence, teachers should be very careful of the divisible nature of reading, sub-skills in it and how to teach and test them to produce efficient readers. *The Journal of Reading* (1971) remarks the following:

Wordsworth said, "The child is father of the man." When that child is figuratively orphaned-intellectually, socially, morally- the man becomes no more mature than the child he was. If a child is an efficient reader but orphaned instructionally he will have little more efficiency as an adult than he had as a child (Editorial, p. 287).

2.3.3 Characteristics of a Good Reading Test

Something is called a sort of its kind if it has some peculiar characteristics pertinent to it. Likewise, tests in general and language tests in particular have certain distinguishing features.

A language test, therefore, is said to be appropriate if it fulfills the criteria of being valid, reliable and practical. Brown (1980:210) asserts that a test is “administrable within given constraints, be dependable, and actually measure what it intends to measure if the three axiomatic criteria are carefully met.”

Brown (1996) states the inter-relatedness of the three qualities of good language test as: “...a test should give the same results every time it measures [reliability], ...should measure exactly what it is supposed to measure... [validity], and should be practical to use.” (p. 185)

2.3.3.1 Test Validity

Validity refers to whether a test measures what it is intended to measure (not something else) (Hughes, 1989; Weir, 1990, 2005; Bachman, 1990). A reading skills test in this regard should test reading (not writing, for example). “A valid test of reading ability,” says Brown (1980:212), “is one that actually measures reading ability and not a 20/20 vision, previous knowledge in a subject, or some other variable of questionable relevance.”

Validity of a test can have different manifestations. It can be of content, face, construct, criterion or washback validities (Brown, 1987; Weir, 1990, 2005; Bachman, 1990; Hughes, 1989).

A. Face Validity

Face validity, according to Anastasi (1982), cited in Weir (1990:26), is “...not validity in the technical sense; it refers, not to what the test actually measures, but to what it appears superficially to measure.” This type of validity is considered only for the purpose of public relations (Davies, 1990; cited in

Teshome (1995); *ibid.*), because a test which lacks face validity "...may not be accepted by candidates, teachers, education authorities or employers" (Hughes, 1989:23). A reading skills test, for example, should have a reading passage followed by comprehension questions to assure the presence of face validity.

B. Content Validity

A test is said to have content validity if it presents a representative sample of the areas covered in the classroom teaching (Hughes, 1989; Brown, 1987; Weir, 1990, 2005). If a test lacks content validity, it not only under-represents the language areas to be tested but also it is likely to have harmful washback effect (Hughes, 1989). Content validity proves the comprehensiveness of a test.

Brown (1980:213) defines content validity as the ability of a test in "...showing how well the content the test samples the class of situations, the universe of subject-matter, about which conclusions are to be drawn." For Anastasi (1982), cited in Weir (1990:25), content validity is determined by "...whether it [the test] covers a representative sample of the behavior domain to be measured."

Thus, the content validity of a reading skills test is maintained if a representative sample of the sub-skills (listed under section 2.3.4) has been included. Both the higher- and lower-level sub-skills should also be tested. Generally, the test should "...depend on truly measuring reading skills and abilities, not the knowledge or lack of knowledge in specific areas" (Livingston, 1972:408).

The distinction among construct validity, criterion-related validity and content validity is not clear (Anastasi (1982); cited in Weir (1990:25); Weir, 2005). Hence, the three validities are supposed, in this paper, as dealt with content validity.

C. Washback Validity

Washback/backwash is defined as the effect a test result has on teaching and learning (Weir, 1990, 2005; Hughes, 1989; Brown, 1987). If a test is valid

content-wise and has criterion validity, the learners will have positive attitude towards taking such tests and learning the language. These sorts of test serve a teaching purpose. Washback validity, as for Weir (2005:210), is also referred to as “consequential validity.”

2.3.3.2 Test Reliability

Weir (2005) states the two main qualities of a test as *validity* and *practicality*, taking *reliability* as a *scoring validity*. A reliable test, for Brown (1980:211), is “consistent and dependable.” Consistency and dependability are attained when the same test is given to the same group of individuals at different times, and if the score is relatively similar (Weir, 1990; Brown, 1987; Hughes, 1989).

Reliability of a test is mainly concerned with scoring of a test; that is why Weir (2005:22) takes it as scoring validity. He also used the name because of “the growing consensus that it [reliability] is a value part of a test’s overall validity” (ibid.).

However, in the practice of language testing, there happen occasions which reduce test validity at the expense of its reliability. That means “a test can be [made] reliable without being valid” (Brown, 1996:213). Weir (1990:33) refers to this practice as a “reliability-validity tension.” And he comments that, though the two concepts are mutually exclusive, “...but if a choice has to be made, *validity after all is the more important.*” (Italics mine).

2.3.3.3 Test Efficiency

‘Test efficiency’ is also called ‘test practicality’ (Brown, 1987) and ‘usability’ (Brown, 1996). A test is said to be practical with regard to “financial limitations, time constraints, ease of administration, and scoring and interpretation” (Brown, 1987; Weir, 1990; Hughes, 1989).

Brown (1980:211) best expresses a test practicality problem as follows:

A test that is prohibitively expensive is impractical. ... a test that requires one-to-one proctoring is impractical for a group of 500 people and only a handful of examiners. A test that takes a few minutes for a student to take and several hours for the examiner to correct is impractical for a large number of testees and one examiner if results are expected within a short time.

Reading skills tests, which contain too short or too long passages and too few or too many reading questions, therefore, are impractical. Alderson, et al. (1995:15-16) state that a reading skills test “should contain approximately 12 items for each reading passage”. They also assert that a reading skills test should use a variety of item types that can be objectively scored.

Finally, we shall sum up this section with the following remark, from Weir (1990:22).

However problematic, there is clearly an imperative need to try and develop test formats and evaluation criteria that provide the best overall balance among reliability, validity and efficiency in the assessment of communicative skills.

2.3.4 Sub-skills in Reading: Where Validity Resides

“And the LORD God formed man of the dust of the ground, and breathed into his nostrils the breath of life; and man became a living soul” (Gen. 2:7).

A test cannot be called a good, appropriate reading test merely because it presents a passage and requires students to answer some questions based on that passage. It goes beyond that. What sort of passage is the passage? Is it a sample of which students could encounter in real-life reading? What test methods does it use to elicit the desired behavior? What sub-skills are included in the test? etc are burning issues that should be raised.

The content of a reading test is the spirit of the test by which it can “move”. A language teacher, as a “creator” of the communicative reading test, should breathe in a life-giving spirit to the test: validity. The validity which is a real one

in which the test is based, i.e. content validity, should be based on representative sample of the sub-skills in reading.

According to Weir (2005:1), communicative testing (of reading) demands both trait and method. The trait, for Weir, is “concerned with the underlying constructs/abilities we wish to measure in students, the what of language testing.” This is what we call validity of a test. Validity (content validity) in a reading test is taken to mean “reading skills defined either by the test itself or the reading authorities” (Livingston, 1972:405). Therefore, our main focus in a reading skills test is on the sub-skills taught and tested so that they can have a beneficial washback on the students’ future reading (Alderson, 1996).

Reading as a multi-factorial competence is witnessed by Munby’s book in which about 19 sub-skills have been listed (Munby, 1978). These are:

- Recognizing the script of the language
- Deducing the meaning and use of unfamiliar lexical items
- Understanding explicitly stated information
- Understanding information when not explicitly stated
- Understanding conceptual meaning
- Understanding the communicative value (function) of sentences and utterances
- Understanding relations within the sentence
- Understanding relations between the parts of a text through grammatical cohesion devices
- Interpreting a text by going outside it
- Recognizing indicators in discourse
- Identifying the main point or important information in a piece of discourse
- Distinguishing the main idea from supporting details
- Extracting salient points to summarize (the text, an idea, etc).

- Selective extraction of relevant points from a text
- Basic reference skills
- Skimming
- Scanning to locate specifically required information
- Transcoding information to diagrammatic display

Alderson and Lukmani, of course, show how differentiating the different sub-skills is problematic.

It is likely that the manner in which individuals arrive at any answer to a question will vary. One person may have difficulty with a particular word and need to infer connections across sentences. Another may understand the word and, therefore, not need to infer. Thus a product, that is, a right answer may be arrived at in a variety of different ways using different processes, strategies or skills. (1989:264).

Because Munby's list of the sub-skills is unclear about the relationships among the enabling skills, Weir's (1983), cited in Alderson (1990a), are taken appropriate, because they are said to be ordered in a list-from lower to higher down the list. These skills are:

- a) Reference skills
- b) Deducing the meaning (and use) of unfamiliar lexical items through understanding word formation and/or contextual clues
- c) Understanding contextual clues within the sentence
- d) Understanding relations between parts of a text through grammatical cohesion devices
- e) Understanding relations between parts of a text by recognizing indicators in discourse
- f) Understanding the communicative function of sentences, with and without indicators
- g) Understanding conceptual meaning
- h) Understanding explicitly stated ideas.

The above skills are taken as **lower-order** because the information about which students are required to answer is clearly stated in the passage. On the other

hand, the following enabling skills are taken as **higher-order** since the information is implicitly stated in the text. Even some are beyond the text. As a result, they demand the reader to read critically.

- i) Understanding ideas in a text not explicitly stated
- j) Separating essential from non-essential in text distinguishing the main idea from supporting details, etc.
- k) Transfer of information from one form to another
- l) Skimming (i.e. surveying to obtain the gist).
- m) Critical evaluation of the text.

Davies (1968), in Alderson and Lukmani (1989:255), reduced the number of the enabling skills into eight, in which sub-skills 1-3 are lower-level and those from 4-8 are higher-level. These are:

- 1) Recalling word meanings
- 2) Drawing inferences about the meaning of a word from context
- 3) Finding answers to questions answered explicitly or in paraphrase
- 4) Weaving together ideas in the content
- 5) Drawing inferences from the content
- 6) Recognizing a writer's purpose, attitude, tone and mood
- 7) Identifying a writer's technique
- 8) Following the structure of a passage

These same skills are classified consisting of the following educational objectives, according to Bloom (1956), cited in Perrot (1982) and Petty (2004): *knowledge, comprehension, application, analysis, synthesis and evaluation*. 'Comprehension' in turn is sub-divided into three: *translation, interpretation, and extrapolation* (Alderson and Lukmani, 1989; Atkins, et al., 1996; Nuru, 1992). 'Knowledge' and 'translation' of 'comprehension' are taken as lower-order skills, while the others are higher-order. Lower-order skills, according to Winne (1979), cited in Nuru (1992:13), are those "which supposedly call on the student merely to recall verbatim or [sic] in his own words material previously read or taught by teachers," and the higher-order ones require the student to

“mentally manipulate bits of information previously learned to create an answer or to support with logically reasoned evidence.”

Other researchers view the enabling sub-skills underlay in levels of comprehension. They sum the sub-skills into three levels: *reading the lines*, *reading between the lines* and *reading beyond the lines* (Gray, in Alderson, 1990); *textually explicit*, *textually implicit* and *scriptally implicit* (Pearson and Johnson (1978), in Taye (1999). These latter levels are equivalent to the former ones with their respective orders.

A question is textually explicit if the question and its answer are derivable directly from the text. If the answer to a question is derived “without any logical or grammatical cue that ties the question to the answer”, the relationship is said to be textually implicit (i.e. reading between the lines). An answer and its question are said to be scriptally related if a nontextual, but plausible answer is given going outside of the text (Taye, 1999: 57-58).

Another classification of the levels is given as: *literal comprehension*, *interpretation* and *critical reading* (Livingston, 1972); and *basic level*, *intermediate level* and *advanced level* (Cooper and Fairbairn, 1989). The difference is only in terminology; what is called *reading between the lines* in one is termed as *textually implicit* or *interpretation* or *intermediate level* in others. Following is an example to show how each level is manifested even in a single sentence, from Livingston (1972:404).

Sentence: Joe, a seven year-old boy, threw the base ball one hundred yards.

Literal comprehension: the reader, to comprehend literally, would gather the primary meaning of individual words and understands a male child of seven threw a spheroid three hundred feet.

Critical reading: if the reader questions the possibility of a seven-year old throwing a baseball one hundred yards, he is engaged in evaluating the truthfulness of the proposition.



Generally, “literal comprehension” refers to the skill of getting the primary, direct meaning of a word, phrase or sentence in context. “Interpretation”, on the other hand, includes skills necessary for getting deeper meanings which are not directly stated in a text (passage). “Critical reading” or evaluation is the one

which includes “literal comprehension” and “interpretation” as defined above, but it goes further than either of these in that the reader evaluates... the quality, the value, the accuracy and the truthfulness of what is read (Smith, 1963; in Livingston, 1972:404).

These enabling skills can be elicited using a variety of test techniques which in turn are based on a reading text. As a direct test of communicative competence, reading is tested through reading. Therefore, the text teachers select for this purpose has a vital role. Knowing or identifying the sub-skills has nothing to do with the cognitive development of students’ information processing if the text selected is not appropriate, and if the type of questions we use to elicit such behavior are defective. Following is an example of questions which are used to elicit different sub-skills of reading based on a text, from Nuttall (1996).

The Son to be Proud of

Last week, Rahman’s wife, Leila, had an accident. Rahman’s youngest child, Yusof, was at home when it happened. He was playing with his toy car. Rahman had given it to him the week before, for his third birthday.

Suddenly Yusof heard his mother calling ‘Help! Help!’ He ran to the kitchen. His mother had burned herself with some hot cooking oil. She was crying with pain and the pan was on fire.

Rahman had gone to his office. Both the other children had gone to school. Yusof was too small to help his mother, and she was too frightened to speak sensibly to him. But he ran to the neighbor’s house and asked her to come and help his mother. She soon put out the fire and took Yusof’s mother to the clinic.

When Rahman came home, Leila told him what had happened. He was very proud of his son. ‘When you are a man, you will be just like your father,’ he said. (p. 233)

Based on the above passage, Nuttall constructed different questions to elicit the different sub-skills of reading.

1. Questions of literal comprehension

- a) Where did Leila have an accident? (*In the kitchen*)
- b) What was Yusof doing when the accident happened? (*Playing with his toy car*)
- c) Why didn't Yusof help his mother? (*Because he was too small*)

The answers to these questions are directly stated in the passage. So, the testee is expected only to follow the lines to 'pick' the answers 'up'.

2. Questions involving reinterpretation or reorganization

- a) How old was Yusof? (Reinterpreting *third birthday, the week before*) – *Three years and three months old.*
- b) How many children has Rahman? (Reorganize: 1 (*Yusof*) + 2 (*both the other children*)) - *He has three children.*
- c) Was Yusof playing in the kitchen? (Reinterpret *ran to the kitchen*) – *He was playing outside.*

3. Questions of inference

- a) Which people were in Rahman's house when the accident happened? - *His neighbour.*
- b) Why was Rahman proud of his son? *Because he saved his mother (Rahman's wife).*

These questions are more difficult (intellectually rather than linguistically) than the above two, because the reader has to understand the text well to work out its implications (p. 188).

2.3.5 Test Methods

The methods we use to measure language ability influence performance on language tests. Performance on language tests... varies as a function both of an individual's language ability and of the characteristics of the test method (Bachman, 1990:113).

Test techniques are “means of eliciting behavior from candidates which will tell us their language abilities” (Hughes, 1989:59). For Hughes, language teachers should use methods which

- elicit behavior that is reliable and valid indicator of the ability in which they are interested.
- elicit behavior which can be reliably scored.
- are as economical of time and effort as possible.
- have a beneficial washback effect.

These methods/techniques include multiple-choice, cloze, C-test, guided short answer, information transfer, true-false, matching, etc. (Hughes, 1989; Weir, 1990, 2005; Heaton, 1988). It is recommended in any reading test to use more than one test technique so that students can have opportunities to practice real-world tasks. “Good reading tests are likely to employ a number of different techniques... since in real reading readers typically respond to texts in a variety of different ways.” (Alderson, 2000; cited in Weir, 2005:119).

Dejene (1994), Desalegn (2000) and Wogari (1995) state similar idea about the different test technique application in reading tests. Desalegn (2000) comments that sticking only to multiple-choice tests disadvantages females. Thus, free response should also be used to help females, Desalegn recommends. In Wogari's (1995) study, multiple-choice questions have been found to be easier to test comprehension than short answer formats. He recommends that the use of a variety of techniques is useful.

Alderson (1996:223-225) listed methods of testing reading skills, dividing them into two: *discrete point versus integrative* methods, and *real-life* methods. The first category includes cloze and gap-filling, multiple-choice and short answer.

For him, short answer formats come closer to the real world. The second category constitutes information-transfer test technique. C-test, cloze-elide, free-recall and summary tests are taken as *alternative* test methods, Alderson asserts.

Reading test specialists (Pyrzczak, 1975; Tuinman, 1971) take objective test formats (e.g. multiple-choice, cloze, C-test, etc) as reading-independent, for the answer can be given from some clues either in the stem or the choices, without processing the text, because the stem or the choices can serve as clues to the answer. Thus, they advise teachers to use reading-dependent questions in which "it is impossible to obtain a high score on such a measure without having read a prerequisite passage" (Tuinman, 1971:289). Reading-independence occurs, according to Pyrczak (1975:308), as a result of:

- a) general knowledge of the point in question.
- b) the interrelatedness of the questions in a given passage.
- c) keyed choices that are longer or more precisely stated than other choices in a multiple-choice question.

So called reading-dependent questions, in most cases, are found to be lower-level both in textbooks and in teacher-made tests (Nuru, 1992). The level test items focus is, of course, not necessarily restricted to the objectivity of the test format. An objective test format, if written well, can test both literal comprehension and interpretation. The example below illustrates this, from Nuttall (1996).

Text

Standing with his back to the window was a man. His hands were round the throat of a woman who faced him, and he was slowly, remorselessly, strangling her.

Question

The murder victim was young.

A. True B. False C. It doesn't say

Here the testee cannot get the correct answer (i.e. **C**) unless he/she interprets the text to say that *woman* refers either middle-aged or old person, not young. The tester should care about where to focus in the text and the way constructing the question plays a significant role.

Cooper and Fairbairn (1989:307) also present a good example of questions to elicit the skill of *reading the lines* and *reading between the lines*.

Text

BOOKING INFORMATION

Course Registration Fee: 25 pounds - not refundable

Accommodation booking deposit: 35 pounds-credited to your accommodation account.

Questions

1. *Basic level (reading the lines) question*

How much is the registration fee?

2. *Intermediate (reading between the lines) question*

Apart from the registration fee, how much more do you need to send if you would like the school to arrange your accommodation?

The answer to the first question is very simple; it can simply be picked up from the text, i.e. 25 pounds. However, the second one demands the testee to read between the lines for interpretation. Because the answer is textually implicit, the test taker should add 35 to 25 to get the correct answer, i.e. 60 pounds. The textual cues “how much more” and “apart from” help the candidate to interpret the text.

Generally, though the direct test techniques (e.g. short answer and information-transfer) are more advantageous than the indirect ones, because they are time-consuming to score and as a result become unreliable, objective test formats can be used to assess the cognitive processing of candidates if made text-dependent and critical.

The main purpose of this part of the paper, as a whole, is to see what the nature of reading is, the sub-skills in reading and the different ways of testing

the enabling skills. Therefore, literature dictates that the text, the sub-skills and test methods need critical consideration in writing reading skills test. To conclude, Davies' (1988:13) advice should be traced here. He remarks that "...what properly distinguishes communicative language testing is not the methods and techniques of testing but the problems, the tasks, the skill manipulations, that what the testee must do with and in the language." To do all these, the teacher's expertise in test construction is the "pillar" for the "hall" of reading skills test.

CHAPTER THREE: RESEARCH DESIGN AND METHODOLOGY

3.1 Research Setting and Subjects of the Study

The present study was conducted in three high schools (Abreha Woatsibeha, Amanuel and Debre Markos) of East Gojjam Zone, Amhara Region. Out of 22 secondary schools in the zone, the researcher selected only three secondary schools. One rationale for selection was the research convenience to collect data. The other was that the passages used to test reading skills in Grade 10 in these schools during last semester (of 2001 E.C. first semester) were found to be longer than a few others in the same setting; there were more questions with better coverage of aspects of reading comprehension skills in the sample schools taken.

Table 1: Location of the Schools Taken as Sample

No	Name of the School	Zone	Woreda
1	Abreha Woatsibeha Secondary School	East Gojjam	Enebisie Sar Midir
2	Amanuel Secondary School	East Gojjam	Machakel
3	Debre Markos Secondary School	East Gojjam	Debre Markos Town

Grade 10 achievement test was taken for analysis because the grade level is the end of general secondary schooling and the transition for preparatory levels. Grades 9 and 11 were excluded not because they were inappropriate, but because both of them are entrance stages for general secondary and preparatory levels respectively. Grade 12 could have been taken as an alternative if there were not time, material and financial constraint. The number of teachers who were teaching in Grade 10 was also more than those in grade 12. Thus, taking grade 12 would have required the researcher to take more schools which in turn would demand many achievement test analysis.

The study mainly focused on EFL teachers expressed beliefs and actual practices in testing reading skills. Hence, the subjects of the study were all EFL teachers who were teaching English in grade 10 of the selected schools. The

researcher checked from the school principals that all the teachers supplemented their teaching through plasma TV last semester. The total number of teachers taken was 15 (4 in Abreha Woatsibeha, 5 in Amanuel and 6 in Debre Markos), of these 4 were females and 11 were males.

Piloting was made on EFL teachers in Motta and Bahire Giorgis high schools, which the researcher excluded because of the short passages and few questions used. The piloting helped to reject an item which was ambiguous from the questionnaire.

3.2 Data Collection Instruments

As for the method of data collection, questionnaire was dispatched to all the 15 teachers to get their competency beliefs in teaching and testing reading skills. The questionnaire had two parts. Both parts were about the 13 sub-skills that are believed to be taught and tested (Weir, 1983; cited in Alderson, 1990a). *Part one* was about teaching the sub-skills, while *part two* was about testing them. Semi-structured interview was also used to supplement the data in the questionnaire. The data obtained through questionnaire and interview, then, were seen in relation to the final achievement test of reading skills in each school.

Piloting was made on EFL teachers in Motta and Bahire Giorgis high schools, which the researcher excluded because of the short passages and few questions used. The piloting helped to reject an item which was ambiguous from the questionnaire. As the piloting did not deal with the details of part two questionnaire (because the questions in the schools were very few – i.e. up to 5 only), the reliability and other related issues were not calculated for the questionnaire.

3.3 Data Collection Procedures

After preparing the questionnaire (i.e. a list of the sub-skills in reading), adapted from Alderson (1990a), EFL teachers were instructed in part one to put

(✓) for the sub-skills they believed were taught last semester, and to put (X) in front of those not taught. For part two of the questionnaire, the same sub-skills in part one were listed parallel to the constructed test items. The sub-skills and test items were used together to check which sub-skill(s) was/were tested by which test item(s). At the back of *part two*, the question paper teachers constructed last semester was attached so that teachers could do the labeling accordingly.

Dispatching the first part of the questionnaire, the researcher conducted a brief discussion on the “whatness” of each of the sub-skills with each teacher. There was a two-day gap between the administration of parts *one* and *two*. Regarding interview, a teacher from each school was selected; the ones with relatively greater teaching experience were chosen.

3.4 Data Analysis Technique

Having collected the data, the researcher applied qualitative data analysis. As the research mainly followed qualitative study design, descriptive analysis was used. The description dealt with the sub-skills teachers believed have been taught and appeared in the achievement test of reading they constructed last semester. In addition, which sub-skill was tested by which test item was seen with respect to the judgments teachers made and on the empirical evidence of the test papers under examination in each school.

Regarding the test methods teachers used to elicit the different enabling sub-skills of reading, the actual test was analyzed empirically to reach a judgment about which test method(s) dominated in the test battery of the schools under investigation.

The above section (i.e. Design and Methodology) presented where the setting of the research was, who the subjects were, what instruments were used to collect data, and how the data were analyzed. It also indicated the way of test analysis for finding out the test methods used in the reading skills tests constructed in the sample schools.

CHAPTER FOUR: PRESENTATION, ANALYSIS AND INTERPRETATION OF DATA

This section presents the data collected through questionnaire and interview. Data from both parts of the questionnaire, and the interview are put in a respective order for presentation, analysis and interpretation. The following, thus, is the personal information of the sample EFL teachers.

4.1 Analysis and Interpretation of Data Elicited through Questionnaire: Parts One and Two

4.1.1 Data from Questionnaire Part One

Table 2: Personal Information of EFL Teachers in the Sample Schools

No	Name of the School	Teachers	Sex	Edu. Background	Teaching Experience	Field of Study	
						Major	Minor
1	Abreha Woatsibeha Secondary School (AASS)	T1	M	Diploma	12 years	English	Amharic
		T2	M	Degree	2 years	English	Amharic
		T3	F	Degree	6 years	English	Amharic
		T4	M	Degree	1 year	English	Amharic
2	Amanuel Secondary School (ASS)	T1	F	Degree	16 years	English	Amharic
		T2	M	Diploma	9 years	English	Amharic
		T3	M	Diploma	5 years	English	Amharic
		T4	M	Degree	6 years	English	Amharic
		T5	M	Degree	8 years	English	Amharic
3	Debre Markos Secondary School (DMSS)	T1	M	Degree	12 years	English	Amharic
		T2	F	Degree	17 years	English	Amharic
		T3	M	Degree	13 years	English	Amharic
		T4	M	Degree	11 years	English	Amharic
		T5	F	Degree	8 years	English	Amharic
		T6	M	Degree	18 years	English	Amharic

* T1, T2, T3, ... T6 = Teacher 1, Teacher 2, Teacher 3, ... Teacher 6

Although the above data do not have any direct relation with the objectives of the study, they are presented only for giving information about the sex, educational status, teaching experience and field of study of the subjects. Especially, the teaching experience was taken as a criterion for selecting

interviewees for the study. This was so because experience is supposed to be important in understanding the problems, threats and opportunities of language testing in general and the testing of reading skills in particular.

As the above table depicts, most of the high school teachers have first degree which is a proof for them taking courses related to teaching and testing reading skills. Those who are diploma holders are attending in-service training during summer as the researcher got through informal talks.

These teachers were given a two-part questionnaire to label. The first part contained reading sub-skills which are believed to be taught and tested. The second part was about which sub-skill was tested by which typical test item. The sub-skills used in this study are listed below.

1. Referencing (anaphoric, cataphoric)
2. Deducing the meaning of unfamiliar words through understanding word formation or contextual clues (morphology or context)
3. Understanding relations within the sentence (understanding syntax)
4. Understanding relations between parts of a text through grammatical cohesion devices (e.g. pro-forms)
5. Understanding relations between parts of text by recognizing indicators in a discourse (i.e. next, that is to say, namely, as we said, for example, etc.)
6. Understanding the communicative functions of sentences, with and without indicators (i.e. related to pragmatic or interaction-dependent function in which the writer expects readers to feel, think or act as a result of having read)
7. Understanding conceptual meaning (propositional meaning), e.g. cause, result, purpose
8. Understanding explicitly stated ideas
9. Understanding ideas in a text not explicitly stated (implications and inferences)

10. Separating essential ideas from non-essential ones in a text: distinguishing the main idea from supporting details
11. Transfer of information from one form to another
12. Skimming (i.e. surveying to obtain the gist)
13. Critical evaluation of the text (writer's attitude, interest, point of view)
(Weir, 1983; cited in Alderson, 1990a:429; Nuttall, 1996)

From the above sub-skills of reading those from 1-8 are said to be **lower-level** because the answer to questions requiring these skills are directly taken from the passage; they are a sort of recall-type. The rest, i.e. 9-13, are **higher-level** as they demand higher cognitive thinking, i.e. analysis, synthesis or evaluation of the discourse. These help learners to form their own interlanguage system. (See also section 2.3.4 in the Literature Review.)

The table below shows to what extent EFL teachers in the selected schools taught the sub-skills last semester. Although the proportion differs, as the researcher checked in the Grade 10 textbook, almost all the sub-skills are present. To what extent do teachers agree in teaching these sub-skills? Here is what the EFL teachers believed having taught.

Table 3: Sub-skills Believed to Have Been Taught by EFL Teachers

Sub-skills Taught	English Teachers in Each School														
	Abreha Woatsibeha				Amanuel					Debre Markos					
	T ₁	T ₂	T ₃	T ₄	T ₁	T ₂	T ₃	T ₄	T ₅	T ₁	T ₂	T ₃	T ₄	T ₅	T ₆
1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
2	✓	✓	✓	✓	✓	✓	✓	✓	x	✓	✓	✓	✓	✓	x
3	x	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	x	✓	✓
4	x	✓	x	✓	✓	✓	✓	✓	✓	x	✓	x	x	x	✓
5	x	✓	✓	✓	✓	✓	✓	x	✓	✓	✓	x	x	✓	✓
6	x	✓	✓	x	✓	✓	✓	x	x	x	✓	x	✓	✓	x
7	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
8	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	x	✓
9	x	x	x	✓	✓	x	✓	x	x	x	✓	x	✓	✓	x
10	✓	✓	✓	✓	x	✓	✓	✓	✓	✓	✓	x	✓	✓	✓
11	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	x	✓	x
12	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
13	x	x	x	✓	x	✓	✓	✓	x	x	x	x	✓	✓	✓

As depicted in Table 3 above, all EFL teachers in each school agreed to having taught sub-skills 1, 7 and 12. With sub-skill 8, only a teacher in Debre Markos

disagreed to having taught it. Two teachers in Debre Markos with sub-skill 11, a teacher in Abreha Woatsibeha and one in Debre Markos with sub-skill 3, and a teacher in Amanuel and one in Debre Markos with sub-skill 2 disagreed to having taught the sub-skills. Two teachers in Abreha Woatsibeha and four in Debre Markos believed sub-skill 4 not taught last semester. The same number of teachers (i.e. two) in Abreha Woatsibeha and one more (i.e. three) teachers in Debre Markos believed sub-skill 6 not taught. Sub-skills 9 and 13 were the ones on which most teachers disagreed; three teachers in Abreha Woatsibeha, three in Amanuel and three in Debre Markos with sub-skill 9, and three teachers in Abreha Woatsibeha, two in Amanuel and three in Debre Markos with sub-skill 13 disagreed to teaching it last semester.

For each school independently, the sub-skills believed to have been taught are described here. Among 4 teachers in Abreha Woatsibeha, all teachers agreed to teaching sub-skills 1, 2, 7, 8, 10 and 12. With sub-skills 3 and 5 only teacher one (T₁) believed them not taught. Sub-skills 4 and 6 were believed not having been taught by two teachers. Three teachers for each believed sub-skills 9 and 13 not taught last semester.

In Amanuel from where 5 EFL teachers were taken, sub-skills 1, 3, 4, 7, 8, 11 and 12 were believed having been taught by all teachers. Sub-skills 2 and 5 were believed not taught only by a teacher each. While sub-skills 6 and 13 were thought not taught by two teachers each, sub-skills 6 and 13 were believed not taught by only two teachers each. Only sub-skill 9 received the least belief of being taught.

Teachers in Debre Markos (i.e. 6 first degree holders) expressed their belief of teaching the sub-skills of reading. All the teachers in the school agreed to teaching sub-skills 1, 7 and 12. With sub-skills 2, 3, 8 and 10 a teacher for each disagreed to teaching them. Two teachers disagreed on sub-skills 5 and 11. The sub-skill on which most teachers (i.e. 4 teachers) disagreed was sub-

skill 4, while those on which 3 teachers for each disagreed were sub-skills 6, 9 and 13. The above description can be summarized as follows.

Table 4: Number of Teachers who Agreed or Disagreed to Teaching the Sub-Skills in Each School

Sub- skills Taught	No of Agreed or Disagreed School Teachers					
	Abreha Woatsibeha		Amanuel		Debre Markos	
	Agreed	Disagreed	Agreed	Disagreed	Agreed	Disagreed
1	4(100%)	0(0%)	5(100%)	0(0%)	6(100%)	0(0%)
2	4(100%)	0(0%)	4(80%)	1(20%)	5(83%)	1(17%)
3	3 (75%)	1(25%)	5(100%)	0(0%)	5(83%)	1(17%)
4	2(50%)	2(50%)	5(100%)	0(0%)	2(33%)	4(67%)
5	3(75%)	1(25%)	4(80%)	1(20%)	4(67%)	2(33%)
6	2(50%)	2(50%)	3(60%)	2(40%)	3(50%)	3(50%)
7	4(100%)	0(0%)	5(100%)	0(0%)	6(100%)	0(0%)
8	4(100%)	0(0%)	5(100%)	0(0%)	5(83%)	1(17%)
9	1(25%)	3(75%)	2(40%)	3(60%)	3(50%)	3(50%)
10	4(100%)	0(0%)	4(80%)	1(20%)	5(83%)	1(17%)
11	4(100%)	0(0%)	5(100%)	0(0%)	4(67%)	2(33%)
12	4(100%)	0(0%)	5(100%)	0(0%)	6(100%)	0(0%)
13	1(25%)	3(75%)	3(60%)	2(40%)	3(50%)	3(50%)

As depicted in Table 4, the EFL teachers in each school put their belief (agreed most) in teaching both the higher-and lower-level sub-skills of reading. Accordingly, from the total of 13 sub-skills taught in the sample schools, only sub-skills 9 and 13 were believed to have received the least, i.e. 25% and 25%; 40% and 60%*; and 50% and 50% respectively in each school. With the rest of the sub-skills, more than 50% of the teachers agreed to teaching them. If they believed teaching the sub-skills, did they test these skills representatively in the achievement test they constructed last semester?

* 60% could not be taken as the least, but it is presented here only for 'binary' description of the two sub-skills. 50% was taken as such because it is the cut-point for 'high' and 'low'.

The following section is about how teachers judged the sub-skills the items were supposed to be testing. For empirical analysis, the three levels of analysis were mainly used from Taye (1999:55-59). According to Taye (1999), based on Pearson and Johnson (1978), a question-answer relationship is said to be textually explicit (reading the lines) if the answer is directly found in the reading passage. If the answer needs a sort of reorganization of words or sentences based on clues or indicators in the discourse, the relationship is called textually implicit (reading between the lines). The last category, i.e. scriptally implicit, is maintained if the answer is given by relating what the reader previously knew to the information in the passage, or answering a question going outside of the text. As indicated in the literature review, the difference between the latter categories lies on where the answer is found. In reading beyond the lines, the answer is not given directly or indirectly in the reading passage; it is through relation with the outside world or experience that the answer can be given. (See the Literature Review for details.)

4.1.2 Data from Questionnaire Part Two

Table 5: Test Items and Respective Sub-skills Tested as Judged by AASS Teachers

Items	Teachers and their Judgments of the Sub-skills Tested				The Sub-skill the Items Empirically Tested
	T ₁	T ₂	T ₃	T ₄	
1	7	12	7	4, 6	(8) textually explicit, reading the lines
2	8	10	8	12	(12) textually implicit, reading between the lines
3	8	8	8	-	(8) recall, textually explicit
4	8	5	7	-	(8) recall, textually explicit
5	12	8	12	6	(8) textually implicit, reading the lines
6	2	2	2	2	(2) recall, textually explicit
7	2	2	2	2	(2) recall, textually explicit
8	2	2	2	2	(2) recall, textually explicit
9	1	1	1	2	(1) recall, textually explicit
10	1	1	1	1	(1) recall, textually explicit

* For the actual test items, see Appendix 4

* The empirical judgments (in the tables below and above) were made based on Perrot (1982), Petty (2004), Nuttall (1996), and Taye (1999).

Table 5 depicts that questions of reference (see items 9 and 10) and vocabulary (i.e. items 6, 7 and 8) were the ones on which all the EFL teachers agreed. With the questions that asked some sort of ‘analysis’ to get the correct answer (like item 2), none of the teachers agreed which sub-skill it was supposed to be testing. Total disagreement or little agreement was seen with items 1 and 4. Teachers also seemed to agree wrongly (see T₁ and T₃ for item 5). Of course, 2 teachers agreed to judging the test item (i.e. item 5) testing skimming. But the actual question did not require students to do so. It was only with item 3 that three teachers agreed in assigning the item to testing explicitly expressed idea empirically. The two teachers (for item 5) agreed wrongly again in assuming the test item testing a higher cognitive level sub-skill.

Most of the test items as judged by the subject teachers (rightly or wrongly) tested the lower-level reading sub-skills. Contrary to the belief teachers made on teaching both levels, they were found to be testing only the lower-levels. This is a proof for the test’s lack of validity (i.e. content validity). This in turn may have a negative backwash effect on students’ future reading. The data also shows that the test was not comprehensive.

Table 6: Test Items and Respective Sub-skills as Judged by ASS Teachers

Items	Teachers and their Judgment of the Sub-skills Tested					The Sub-skills the Items Empirically Tested
	T ₁	T ₂	T ₃	T ₄	T ₅	
1	6	8	8	3,13	8	(8) recall, literal comprehension
2	3	9	8	1,8	8	(9) analysis, textually implicit
3	3	7	8	4,7,9	9	(9) analysis, reading between the lines
4	8	8	8	-	8	(9) analysis, reading between the lines
5	1	1	1	1	1	(1) recall, reading the lines
6	1	1	1	1	1	(1) recall, reading the lines
7	7	5	5	-	12	(9) analysis, textually implicit
8	2	-	-	-	2	(2) recall, reading the lines
9	2	2	2	2	2	(2) recall, reading the lines
10	2	2	2	2	2	(2) recall, reading the lines
11	2	2	2	2	2	(2) recall, reading the lines

* For the actual test, see Appendix 5

As Table 6 reveals, three teachers agreed in assigning item 1 to sub-skill 8, but with item 4 four teachers agreed wrongly to what the test was empirically testing. Like the AASS EFL teachers, the ones in Amanuel agreed to test items which were supposed to test referencing (items 5 and 6) and contextual vocabulary (i.e. items 8, 9, 10 and 11). Similarly, the items constructed tested more of the lower-level than the higher-level ones. This again made a mismatch between what was believed having been taught and what was tested. Even those items which teachers agreed to test a certain sub-skill did not test what they supposed them to be testing. Item 4, for example, empirically tested learners what was not put directly in the passage. However, most teachers judged the item as if it were testing explicitly stated idea. Surprisingly, a teacher judged one item (i.e. item 2) testing two extremely different sub-skills (see T₄ in the above table). This might have been because of taking “reference” the same as/similar to “explicitly stated idea”, lack of knowledge. Reference is what a word or phrase refers to, while explicitly stated idea is an expression said in the same way as the idea in the passage. The same confusion happened with item 3 as judged by T₄. The teacher became uncertain about whether the item tested cohesion, propositional meaning or implicitly stated idea, while the item tested what was expressed implicitly, when seen empirically.

Generally, EFL teachers in Amanuel, like the ones in Abreha Woatsibeha failed to test the sub-skills they claimed having taught representatively. In the same way these teachers disagreed in labeling items which were supposed to be testing higher cognitive ability. (See items 2 and 3 in Table 6.)

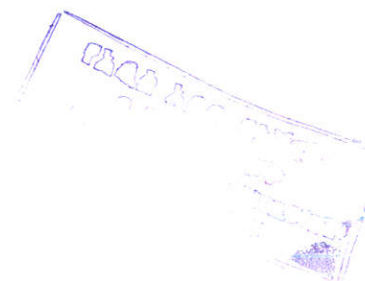
Table 7: Test Items and Respective Sub-skills as Judged by DMSS Teachers

Items	Teachers and their Judgment of the Sub-skills Tested						The Sub-skills the Items Empirically Tested
	T ₁	T ₂	T ₃	T ₄	T ₅	T ₆	
1	7	5	3	7	7	8	(8) recall, textually explicit
2	12	12	9	13	7	12	(13) scriptally implicit, reading beyond the lines
3	1	1	1	1	1	1	(1) recall, reading the lines
4	6	7	5	8	-	3	(8) recall, textually explicit
5	-	7	10	8	-	8	(8) recall, textually explicit
6	2	2	2	2	2	2	(2) recall, reading the lines
7	3	8	11	11	3	11	(8) recall, textually explicit
8	2	2	2	2	2	2	(2) recall, reading the lines
9	2	2	2	2	2	2	(2) recall, reading the lines
10	2	2	2	2	2	2	(2) recall, reading the lines

* For the actual test, see Appendix 6

Just like the former schools, Table 7 presents that EFL Teachers in DMSS agreed on items which were supposed to test lower-level sub-skills. The referencing and contextual vocabulary meaning were the agreed upon ones in this school. The higher-level cognitive sub-skills which demand learners to analyze or evaluate the text were ignored. Teachers even testing lower-level sub-skills agreed little, except on referencing (i.e. item 3) and vocabulary (see items 6, 8, 9 and 10), to identify which item was testing which sub-skill. With items 1 and 7, the teachers agreed with each other wrongly: different from what the test empirically tested. The only exception from the former two schools was that a question which required learners to read beyond the lines was asked. (See item 2 in the above table.)

What the EFL teachers claimed they taught and what they actually tested did not match. The higher-level skills which they believed were taught last semester appeared almost no where in the test. This shows lack of validity of the test. In



addition, as the items did not make learners interact with the writer through the text, there seemed no possibility of interlanguage system formation, because, as Davies (1988) says, though the answers in a reading test are predictable, the way the questions are asked matter a lot in interlanguage formation.

The presentation of the questions is a possibility for learners to process information through active interaction in the text. The items in this case are taken as intermediate stages between what the learner has constructed and what is actually in the text, which the writer has intended to transmit. On eliciting this ability of extracting information, the test methods used play a great role. Following are the test methods which EFL teachers in the sample schools employed to test the reading ability of their students.

Table 8: Test Methods Used in the Sample Schools

Test Methods	Schools					
	Abreha Woatsibeha		Amanuel		Debre Markos	
	Yes	No	Yes	No	Yes	No
Multiple-choice	✓	-	✓	-	✓	-
Short-answer	-	✓	-	✓	-	✓
Information-transfer	-	✓	-	✓	-	✓

* For the actual tests, see Appendices 4-6

The above test methods are supposed to be employed in integrative and communicative testing of reading skills. Though there have been lots of short answer and information transfer activities in the Grade 10 English textbook, the achievement test constructed in the sample schools used only multiple-choice format. The reasons might have been its being time-saving in scoring. But information-transfer and short answer formats could have been used with ease of scoring to assure the reliability of the test. Although a multiple-choice item could test a higher-level sub-skill, what has empirically been seen in the items constructed in the sample schools was focusing on lower-level, i.e. getting the right answer directly from the passage.

A cognitively demanding-question can make learners process information to form their own system cognitively, i.e. interlanguage. It goes with what Davies says, "Language tests [of any sort] are operational definitions of the knowledge of language that learners require and native speakers [or writers] possess." (1988:13)

Except the test methods analysis which was directly taken from the actual test papers, regarding the sub-skills and any other practice in testing, reading skills, the most experienced teacher from each school was interviewed. As mentioned in the methodology part, this helped the researcher to supplement the data elicited through questionnaire.

4.2 Analysis and Interpretation of Data Elicited through Interview

In all the schools, a highly experienced teacher from each school was interviewed on his/her practice of testing reading skills. Although there were other responses which the interviewees expressed about their belief and practice in testing the reading skills, for convenience, only four basic questions were chosen to show the analysis and interpretation of the interview data.

A. Do you discuss with each other on the reading skills test prior to its construction?

In all the schools, the teachers said that they did not thoroughly discuss what to include in the test of reading skills. Neither did they discuss which sub-skills should be tested with which typical test items. In none of the schools was there a discussion about the test methods to test the reading sub-skills. It seemed from this ignorance that all the schools used only multiple-choice format. This could also be a reason for their failure to agree to which item was intended to test which sub-skill.

B. Do you think the courses you took in undergraduate programmes helped you in your practice of testing reading skills?

Regarding whether or not the courses they took in their undergraduate courses helped them construct good reading skills test items, they said the courses helped them. But their focus was found to be on reference and vocabulary. This is a mirror image of what was reflected on their actual achievement tests of reading skills. (See Appendices 4-6 for the actual test papers). What they wrote on the questionnaire (part one) and what they said in the interview did not go with what they actually did.

C. Are there factors which you think hinder you from constructing good reading skills tests?

All the school teachers interviewed agreed that the school principals forced them to limit the number of questions and the length of reading passages. This may be why a two-up-to-four paragraph length passages appeared in the tests. Even the tests which the researcher intended to include from Motta, Bahire Giorgis and Menkorer (all in East Gojjam) included only a paragraph and 5 questions. Among these questions half or so were reference and vocabulary. How can we say that reading skills is tested in such a way? How can a paragraph and 5 questions be a representative sample of the sub-skills taught in about 14 passages of the Grade 10 textbook?

D. What do you think should be done in the future to improve a reading skills test at a high school level?

Expressing their belief about the importance of a good reading skills test in improving students' reading ability, regarding 'What should be done in the future?', teachers recommended that they should teach the sub-skills thoroughly, convincing learners about the importance of reading skills. The school administrators should allow them to write longer passages and appropriate number of items to the extent of the schools' resources, the

teachers added. The school principals also agreed to this issue, when the researcher talked to them about the same problem.

Generally, as elicited in data from interview, the sub-skills of reading that are believed to be taught and tested were not given due emphasis at a high school level. Especially, the reluctance teachers and principals have in Grade 10 reading skills test was a great threat for the efficient implementation of the curriculum to bring the expected outcomes. If learners are not taught and tested the reading sub-skills representatively, it is believed that they will not be able to understand any subject in their academic area. (See the Literature Review). Below is a presentation of what the interviewees expressed regarding reading skills test.

i. An Interview Conducted in Abreha Woatsibeha Secondary School

Interviewer (M): Welcome to our interview session. First of all, I ask you to tell me your name and your teaching experience.

Interviewee: Thank you. My name is M₁ and I have been teaching English for about 12 years.

M: Have you ever been a department head?

M₁: Yes, since 1999 E.C. I have been a department head.

M: In what ways do you construct tests in your school? Do you construct them individually or in collaboration?

M₁: As a department head, I always advise colleagues to construct tests cooperatively, and this is how we usually do in our school.

M: Do you do the same for reading passage?

M₁: Yes, we do the same. We find out a passage which is understandable by learners from magazines or newspapers, and we adapt it in some form. We do not directly take the items from the original material.

M: Do you think the test items you constructed are representative samples of the sub-skills in reading?

M₁: Yes, I do. The skimming, the reference, scanning, etc. are both in the textbook and in the test items we construct. Thus, I can say, it represents the taught skills.

M: Do you think testing reading can help learners to develop their reading ability?

M₁: As there are many comprehension questions in the textbook, teaching these helps learners.

M: Teaching the skills, yes. How about testing? Does it help?

M₁: Absolutely.

M: Have you ever faced any problem in testing reading skills?

M₁: One problem is that students give much focus to grammar. They do not want to work on comprehension. Secondly, school administrators usually restrict us not to construct a longer passage having many questions. Of course, we are given more chance than other departments; it is not enough, though.

M: What should be done in the future to improve the practice of reading skills test?

M₁: We teachers have a problem, a knowledge problem. Therefore, first, we need to know the sub-skills, select and teach them to students. Secondly, we should construct test items purposefully to test the reading sub-skills representatively. Finally, inputs (like newspapers, magazines, etc) should be provided in the school so that we teachers can have alternatives to adapt passages for a reading skills test.

ii. An Interview Conducted in Amanuel Secondary School

M: Welcome to our interview.

Interviewee: Thank you.

M: I would like you to tell me about yourself in brief, your name and teaching experience?

Interviewee: My name is A and I have a 16 years teaching experience.

M: Have you ever been a department head?

A: No, Never.

M: Where did you attend your undergraduate program?

A: At Bahir Dar University.

M: How do the courses you took there helped you in testing at the high school?

A: I took many courses, such as *Communicative Skills* parts I and II, *Teaching Methodology*, *Reading Skills*, etc.

M: In your Methodology course did you learn something in relation to testing?

A: Yes, I did. About lexical words, skimming, references, multiple-choice, etc.

M: How do you construct reading skills test in your school? Do you discuss?

A: Yes, we discuss to include the sub-skills like skimming, referencing, lexical words and so on.

M: Is the reading skills test (passage) constructed by a teacher or many teachers?

A: Each of us has shares. And a teacher who is better at reading skills test construction takes the responsibility. After that we discuss the questions constructed. The rest of us construct grammar items.

M: Do you think testing reading can help learners improve their reading ability?

A: Yes, I do believe. Reading can be improved through reading different passages. Teaching what are in the textbook also helps a lot.

M: It is about teaching. Teaching helps, sure. How about testing?

A: I think so.

M: What should be done in the future in relation to testing reading skills?

A: Constructing longer passages and 'many' test questions will help.

M: Thank you very much.

A: It is my pleasure.

iii. Interview Conducted in Debre Markos Secondary School

M: Welcome to the interview willingly. Could you tell me in brief about your name and teaching experience?

Interviewee: Yes, indeed. My name is B. I have an experience of 18 years in teaching English.

M: Have you ever been a department head?

B: Yes, in 1998 E.C.

M: Do you construct test items together or individually?

B: It depends. First, individually; then, together. We share the questions among 6 of us. And a teacher takes the responsibility of constructing a reading skills test. After that we discuss on all the items together.

M: What do you do then?

B: We decide whether the test is simple or difficult. I once, for example, constructed a very simple reading test.

M: How do you say the test was very simple?

B: Because I know that my students were poor.

M: How do you see your students' participation in answering the questions in the textbook?

B: Except the clever, very few participate to answer.

M: What do you think is the reason?

B: It is their background. They have very poor foundation (base). At elementary level you can shape students, but here impossible.

M: Why is that so?

B: Because it is not the right time.

M: When you attended your undergraduate studies, how much did the courses you took helped you in test construction?

B: We learned, but we do not apply it. Even our teachers at the universities did not themselves do what they say should be done.

M: How can the reading sub-skills in the textbook be tested?

B: We do not teach all what is there, though possible. We focus only on skimming, reference and vocabulary.

M: Why is it impossible to test the sub-skills in reading?

B: Because teaching all the sub-skills is impossible.

M: Can a good reading test help learners improve their reading ability?

B: It helps.

M: What should be done to improve the testing of reading skills?

B: First of all teachers must teach all the sub-skills. The other, students should be advised not to be reluctant in doing comprehension exercises. Finally, school principals should not restrict language teachers on the length of the passages and the number of test items

M: Thank you for your cooperation

B: You're welcome.

This section of the paper as a whole tried to see which sub-skills teachers believed having taught last semester appeared in the achievement test of reading they constructed themselves. Again the section dealt with investigating with which test items and sub-skills teachers agreed in the test-sub-skill assignment. It also showed the investigation of the kind of test method used.

As the data from questionnaire, interview and test paper analysis of the section indicated, teachers failed to test a representative sample of the sub-skills they claimed having taught. This shows that their test lacked content validity, i.e. the expected enabling skills in reading were not tested. The only sub-skills appeared on the tests were found to be *reference*, *vocabulary* and *explicitly stated idea*-all lower-level. As a result, teachers agreed little in assigning the higher-level sub-skills (e.g. inferences) with typical test items, in the tests they constructed using only a single test method.

CHAPTER FIVE: SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

5.1 Summary

This study is about EFL teachers expressed beliefs and actual practices in testing reading skills. Out of 22 high schools in East Gojjam Zone, 3 high schools, in which 15 teachers teach English in Grade 10, were taken as samples. Data about testing reading skills were elicited through questionnaire and interview. The elicited data, then, were seen in relation to the actual tests constructed. The analysis of the data and their interpretation is summarized as follows.

As many researchers (Getachew, 1996; Haile, 2007) found out reading skills were not taught at high schools, this paper proved that they were not tested as well. When the subjects were asked to indicate which sub-skills they believed having taught, they labeled most of the sub-skills listed by Weir (1983), cited in Alderson (1990a). Their actual tests, however, were found to be testing only a few of the sub-skills listed. Unlike the teaching, teachers disagreed a lot in assigning a sub-skill with a typical test item.

The excuse they proposed when interviewed for the short passages and few numbers of sub-skills tested, they blamed school principals for restriction. In any way, EFL teachers constructed items which made learners only read the lines; they encouraged only literal comprehension. This aspect was also manifested in the use of only a single test method. Multiple-choice test format dominated all the EFL teachers practice in all the schools.

Generally, the following points can be made in relation to the objective of the study.

- EFL teachers did not test a representative sample of the sub-skills they claimed having taught last semester.

- As a whole the sub-skills were not tested in a reasonable proportion. Only *referencing*, *vocabulary* and *explicitly stated idea* dominated the test battery.
- EFL teachers failed to agree in assigning the items with tested sub-skills other than referencing and vocabulary.
- They totally used only one test method (i.e. multiple-choice). The other alternative formats like short answer, gap-filling or information transfer were given no value.

5.2 Conclusions

From the analysis and interpretation of the data, the following conclusions can be made.

- The EFL teachers seemed to lack enough training on the ‘whatness’ of the sub-skills of reading, how to teach and test them as well.
- Reading passages appeared on achievement tests only for their own sake. If that were not so, the passages would be followed by items which tested both higher- and lower-order sub-skills. However, the actual test was found to test only the *reference* and *vocabulary* in focus.
- The EFL teachers might have received little training about how to teach and test reading sub-skills. Especially, the testing knowledge teachers were found to have needs critical consideration, for it highly influences students future academic life.
- The content of the reading skills tests in each school seemed to have been “...determined by what is easy to test rather than what is important to test” (Brown, 1980:23); validity was sacrificed on the altar of practicality.
- Testing as a tool for teaching seemed to have been neglected. The sub-skills included in each test item and the test methods used were witnesses for this negligence.

- The reading skills tests constructed at the high schools were a sort of discrete-point, not integrative or communicative. The tests mainly encouraged the application of the bottom up approach of reading skills.

5.3 Recommendations

Based on the findings of the study, the researcher would like to recommend the following so that future reading test practices could be improved.

- In undergraduate courses, detailed training should be given regarding teaching and testing reading skills. This is so because reading skills is a key for other academic successes. Especially, the teaching value of tests should be given due consideration.
- At high schools in which teaching and testing take place, seminars, workshops, short-term trainings and discussions should be organized by the school, the departments or the teachers themselves. This may help teachers be aware of teaching and testing the reading skills, for it opens the door for experience sharing. Teachers in a department should also discuss the passages selected, the sub-skills included and the test methods used in a reading skills test thoroughly.
- Encouragement should be provided to EFL teachers in constructing test items based on longer passages employing a variety of test methods. Restriction on the part of the school principals should be minimized. To do so, discussions between EFL teachers and school principals should be held.
- A variety of reading materials should be displayed in libraries in order for students to practice reading and understanding. Providing magazines, newspapers, etc has a two-fold purpose. One, it exposes learners with authentic language use. The other, it can serve teachers as sources for reading passages.

- Alternative test methods which can be scored objectively should also be included in the testing of reading skills both at the national levels and in the classrooms.
- Finally, though students are more interested in grammar parts than in the reading passage (as the researcher experienced in his teaching at the high schools for the past six years), teachers should draw students' attention to passages in which the grammar is seen doing some work. They should convince their students that learning grammar in isolation is nothing unless it is used in context. The real context for grammar in use is a reading passage, teachers should emphasize.

BIBLIOGRAPHY

- Abdu, M. (1993). Reading Preferences of High School Students, Unpublished MA Thesis, AAU.
- Alderson, J. C. (1990a). Testing Reading Comprehension Skills (Part One), *Reading in a Foreign Language*, 6(2), 425-439.
- _____. (1990b). Testing Reading Comprehension Skills (Part Two), *Reading in a Foreign Language*, 7(1), 465-503.
- _____. (1996). The Testing of Reading: In Nuttall, C. (1996). (New Edition). *Teaching Reading Skills in a Foreign Language*. Hong Kong: Heinemann.
- Alderson, J.C. and Y. Lukmani. (1989). Cognition and Reading: Cognitive Levels as Embodied in Test Questions, *Reading in a Foreign Language*, 5(2), 253-270.
- Alderson, J.C., C. Clapham, and D. Wall. (1995). *Language Test Construction and Evaluation*. Cambridge: CUP.
- Atkins, J., Hailom, B. and Nuru, M. (1996). *Skills Development Methodology (Part 2)*. Addis Ababa: Addis Ababa University (AAU) Press.
- Bachman, L. (1990). *Fundamental Considerations in Language Testing*. Oxford: OUP.
- Brown, J.D. (1996). *Testing in Language Programs*. New Jersey: Prentice-Hall, Inc.
- Brown, H.D. (1980). *Principles of Language Learning and Teaching*. Englewood Cliffs: Practice-Hall, Inc.
- _____. (1987). *Principles of Language Learning and Teaching*. (2nd ed.). Englewood Cliffs: Prentice-Hall, Inc.



- Cooper, J. and I. Fairbairn. (1989). Ten Years of the Communicative Use of English Examinations-Reading, *Reading in a Foreign Language*, 5(2), 299-309.
- Davies, A. (1988). Communicative Language Testing: In Hughes, A. (1988). *Testing English for University Purposes: ELT Document 127*. Modern English Publishers.
- _____ (1991). Language Testing in the 1990's, *Language Testing in the 1990's*, 1(1).
- Dejene, L. (1994). Language Testing and Its Practical Application: The Testing of Reading in Focus (Teaching Handbook). Department of Foreign Languages and Literature. Institute of Language Studies, AAU.
- Desalegn, G.S. (2001). Evaluation of Multiple-Choice and Free Response Formats in English Language Tests: Are They Gender-Biased? Unpublished MA Thesis, AAU.
- Duffelmeyer, F.A. (1980). The Passage-Independence of Factual and Inferential Questions, *Journal of Reading*, 24(2), 131-140.
- Genesee, F. and J. A. Upshur. (1996). *Classroom-based Evaluation in Second Language Education*. Cambridge: CUP.
- Getachew, A. (1996). The Teaching of Reading in Government High Schools in Addis Ababa: A Descriptive Study, Unpublished MA Thesis, AAU.
- Grellett, F. (1981). *Developing Reading Skills: A Practical Guide to Reading Comprehension Exercises*. Cambridge: Cambridge University Press.
- Haile, S. (2007). The Application of Interacting Reader Variables with Text Variables in a Reading Lesson: The Case of Debre Markos College of Teacher Education First Year 10+3 Programme, Unpublished MA Thesis, AAU.
- Heaton, J.B. (1988). *Writing English Language Tests*. (New Edition). London: Longman.

- Hughes, A. (1989). *Testing for Language Teachers*. Cambridge: Cambridge University Press.
- Lee, J. F. (1997). Non-native Reading Research and Theory, In Bardovi-Harlig, K. and B. Hartford. (eds.). (1997). *Beyond Methods: Components of Second Language Education*. New York: McGraw-Hill Companies, Inc.
- Livingston, H. F. (1972). What the Reading Test Does not Test-Reading, *Journal of Reading*, 15(6), 402-410.
- Matthews, M. (1990). Skills Taxonomies and Problems for the Testing of Reading, *Reading in a Foreign Language*, 7(1), 511-517.
- Mekonnen, G. (2005). The Cognitive Levels of the Reading Questions and the Teachers' Classroom Supplement to the Questions: English for Grade 10, Unpublished MA Thesis, AAU.
- Munby, J. (1978). *Communicative Syllabus Design*. Cambridge: Cambridge University Press.
- Nuru, M. (1992). Levels of Questions: A Description of Textbook and Exam Questions in Higher Secondary Schools, Unpublished MA Thesis, Addis Ababa University.
- Nuttall, C. (1996). (New Edition). *Teaching Reading Skills in a Foreign Language*. Hong Kong: Heinemann.
- Pearson, I. (1988). Tests as Levers for Change, In Chamberlain, A and R.B Baumgardner (eds.). (1988). *ESP in the Classroom: Practice and Evaluation*. (ELT Document 128). Modern English Publishers.
- Perrott, E. (1982). *Effective Teaching: Practical Guide to Improving your Teaching*. Longman: Longman Group Limited.
- Petty, G. (2004). *Teaching Today: A Practical Guide*. (3rd ed.) London: Nelson Thornes (Publishers) Ltd.

- Pyrczak, F. (1975), Passage-Dependence of Reading Comprehension Questions: Examples, *Journal of Reading*, 18(4), 308-311.
- Richards J. C and T. Rodgers (2001). *Approaches and Methods in Language Teaching: Description and Analysis*. Cambridge: Cambridge University Press.
- Richards, J.C. (ed.). (1974). *Error Analysis: Perspectives on Second Language Acquisition*. Longman: Longman Group Ltd.
- _____ (2001). *Curriculum Development in Language Teaching*. Cambridge: CUP.
- Schreiner, R. (ed.). (1979). *Reading Tests and Teachers: A Practical Guide*. Minnesota: International Reading Association, Inc.
- Sherif, A. (2007). The Effect of Schema-Based Pre-reading Tasks on Reading Comprehension: Purpose-Focused Study, Unpublished MA Thesis, AAU.
- Sileshi, K. (2007). An Investigation of Preparatory Students' Reading Comprehension Abilities with Particular Reference to Yekatit 12 Preparatory School, Unpublished MA Thesis, AAU.
- Taye, R. (1999). The Effect of Academic Background on Reading Comprehension: The Case of First Year Students at the College of Social Sciences at AAU, Unpublished Doctoral Dissertation, AAU.
- Teshome, D. (1995). *The Construction and Validation of English Language Tests at Tertiary Level*, Unpublished Ph. D Dissertation. AAU.
- Trinitarian Bible Society. (2007). *The Holy Bible*. London: Tyndale House.
- Tuinman (1971). Asking Reading-Dependent Questions, *Journal of Reading*, 14(5), 289-293.
- Weir, C. J., A. Hughes and D. Porter. (1990). Reading Skills: Hierarchies, Implicational Relationships and Identifiably, *Reading in a Foreign Language*, 7(1), 505-510.

- Weir, C.J. (1990). *Communicative Language Testing*. London: Prentice-Hall International Ltd.
- _____. (2005) *Language Testing and Validation: An Evidence-based Approach*. Chippenham and Eastbourne: Antony Rowe Ltd.
- Weir, C.J. and D. Porter. (1994). The Multi-divisible or Unitary Nature of Reading: The Language Tester Between Scylla and Charybdis, *Reading in a Foreign Language*, 10(2), 1-19.
- Williams, E. (1984). *Reading in the Language Classroom*. Modern English Publishers.
- Wogari, N. (1995). The Impact of the Multiple-choice and the Short Answer Formats on Students' Performance: Three Government High Schools in Focus, Unpublished MA Thesis, AAU.
- Wondifraw, W. (2005). A Comparative Study of Text Difficulty and Learners' Comprehension Level at Secondary Schools, Unpublished MA Thesis, AAU.
- Zelege, A. (2007). An Exploration of Some of the Factors Affecting Preparatory II Students' Reading Comprehension, Unpublished MA Thesis, AAU.

Appendix 1: Questionnaire Part One

Addis Ababa University

School of Graduate Studies

Department of Foreign Languages and Literature

Questionnaire to be filled by EFL Teachers at _____ High School

PART ONE

Dear teacher,

This questionnaire is designed for a study at AAU to investigate the practice of EFL teachers in testing reading skills at your school. Your cooperation is very important for the completion of this study which may have significant influence on future testing practices. Thus, I kindly request you to fill this questionnaire carefully. Put a tick (✓) mark for your choice. It is no need to write your name.

Thank you in advance.

I. Personal Information

1.1 Sex: Male Female

1.2 Educational Background MA BA/BEd Diploma

1.3 Teaching Experience: (Please specify) _____ years

1.4 Field of Study: Major: _____ Minor: _____

II. Main Body

Following are the sub-skills in reading which are believed to be taught and tested in EFL classroom. Which skills do you think you taught in the classroom last semester? Put a tick (✓) mark parallel to the sub-skills you taught, and an (X) mark parallel to those not taught.

1. Reference skills (anaphoric, cataphoric)
2. Deducing the meaning of unfamiliar lexical items through understanding word formation or contextual clues
3. Understanding relations within the sentence

4. Understanding relations between parts of text through grammatical cohesion devices
5. Understanding relations between parts of a text by recognizing indicators in discourse
6. Understanding the communicative functions of sentences, with and without indicators
7. Understanding conceptual meaning, e.g. cause, result, purpose
8. Understanding explicitly stated ideas
9. Understanding ideas in a text not explicitly stated (implications and inferences)
10. Separating essential ideas from non-essential ones in a text: distinguishing the main idea from supporting details, etc.
11. Transfer of information from one medium to another
12. Skimming (i.e. surveying to obtain the gist)
13. Critical evaluation of the text (writer's attitude, interest, point of view)

(Adapted from: Weir, 1983; cited in Alderson, 1990a:429)

Appendix 2: Questionnaire Part Two

PART TWO

This questionnaire is an extension of **PART ONE**. It focuses on testing reading sub-skills. The questionnaire is to be answered based on the question paper attached to it. The *question paper* is what YOU constructed last semester to test your students' reading ability. Please, put a tick (✓) mark parallel to the sub-skills you think are tested in a typical test item.

No.	Sub-skills in reading	Test items constructed to test the sub-skills														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	Referencing (anaphoric, cataphoric)															
2	Deducing the meaning of unfamiliar lexical items through understanding word formation or contextual clues															
3	Understanding relations within the sentence															
4	Understanding relations between parts of a text through grammatical cohesion devices															
5	Understanding relations between parts of a text by recognizing indicators in discourse															
6	Understanding the communicative functions of sentences, with and without indicators															

7	Understanding conceptual meaning, e.g. cause, result, purpose																		
8	Understanding explicitly stated ideas																		
9	Understanding ideas in a text not explicitly stated (implications and inferences)																		
10	Separating essential ideas from non-essential ones in a text: distinguishing the main idea from supporting detail, etc.																		
11	Transfer of information from one form to another																		
12	Skimming (i.e. surveying to obtain the gist)																		
13	Critical evaluation of the text (writer's attitude, interest, point of view)																		

Appendix 3: Interview Questions

1. Do you discuss with each other on the reading skills test prior to its construction?
2. Do you think the courses you took in undergraduate programmes helped you in your practice of testing reading skills?
3. Are there factors which you think hinder you from constructing good reading skills tests?
4. What do you think should be done in the future to improve a reading skills test at a high school level?

Appendix 4: AASS Reading Skills Test

Direction 4: For Items 31-41 read the passage. Then, answer items based on the information in the passage

Pulp Friction

Every second, 1 hectare of the world's forest is destroyed. That is equivalent to two football fields. An area size of New York city is lost every day. In each year, that adds up to 35 million hectares, more than the land area of Poland. This alarming rate of destruction has serious consequences for the environment; scientists estimate, for example, that 137 species of plant, insect or animal become extinct every day due to logging. In British Columbia thirteen rainforest valleys have been clear cut and 142 species of salmon have already become extinct since 1990, and cheap products of all kinds of consumers; so the government is reluctant to restrict or control it.

Much of Canada's forestry production goes towards making pulp and paper. According to Canadian Pulp and Paper Association, Canada supplies 34% of the world's wood pulp and 49% of its news print paper. If these paper products could be produced in some other way, Canadian forests could be preserved. Recently, a possible alternative way of producing paper has been suggested by agriculturalists and environmentalists; a plant called hemp.

(GEQAEA, EGSECE)

1. How many species of salmon have become extinct in British Columbia?
A. 137 B. 13 C. 31 D. 142
2. What is the main idea of paragraph 1?
A. The extinction of animals B. The effects of logging
C. The government's effort to make money from logging
D. Canada's pulp and paper production E. All except 'D'

3. A possible alternative way of producing paper as suggested by agriculturalist and environmentalist is
- A. logging B. using hemp C. clearing forest D. all
4. How much of the world's rain forest is destroyed every second?
- A. 31 million hectares B. 1 hectare
- C. An area of two football fields D. B and C E. None
5. One of the following sentences is not true. Which one?
- A. America supplies 34% of the world's wood pulp.
- B. 137 species of plant, insect or animal become extinct due to logging.
- C. A forest area of two football field is destroyed every second.
- D. Canada supplies 49% of newspaper.

Direction 5: For items 36-38 guess the meanings of the underlined words form the context.

6. This alarming rate of destruction has serious consequences.
- A. Increasing B. Decreasing C. Exciting D. Threatening
7. 142 species of salmon have already been extinct since 1990,
- A. Reproduced B. Imposed C. Died out D. Sold
8. ...the government is reluctant to restrict or control it.
- A. Willing B. Committed C. Careful D. Unwilling

Reference Questions

9. ...so the government is reluctant to restrict or control it.
- What does 'it' refer?
- A. Hogging B. Government C. Consumer D. Product
10. ...34% of the world's wood pulp and 49% of its news print...
- A. Wood pulp B. Association C. Canada D. World

Appendix 5: ASS Reading Skills Test

III: Answer the following questions based on the passage.

IS THERE LIFE ON MARS?

It is now generally agreed that life of a kind exists on Mars. Through a telescope one can see the dark areas which have interested astronomers for hundreds of years. They look rather like vegetation and they behave as vegetation would be expected to do. There is no valid reason why there should not be vegetation. But we must be rather cautious. It is often said that we may find vegetation on Mars which is similar to our lichens and mosses. That may be though there is no evidence for it. All we can say is that Mars contains lowly vegetation we can almost certainly be right, but there is probably no animal life there.

There is a considerable atmosphere on Mars. It is thinner than the earth's atmosphere. It is even thinner than the air on top of Mount Everest. We certainly could not breathe it, nor could any earth type animal, on the other hand, there is enough pressure to support plant life. The Martian atmosphere is made up mainly of nitrogen, so far as we can tell. Though there is a certain amount of carbon dioxide gas, there is very little oxygen and not much water vapor when we finally succeed in getting to Mars. Therefore, we are going to find that the atmosphere there will be of little help to us.

The day temperatures on Mars are moderate. On a rarely hot summer's day on the Martian equator the temperature can go up to about 50° Fahrenheit (about 25° centigrade). The nights, however, are extremely cold because the atmosphere is too thin to keep in the sun's heat even on the Martian equator. We would find that a summer night is good deal colder than a winter night in Greenland or Antarctica.

Taken from: College English (vol. 1)

1. In the first paragraph, the author says we must be careful because
 - A. It is difficult to live on Mars as it is very cold at night.
 - B. We can see dark areas which have interested astronomers for hundreds of years.
 - C. It is often said that we may find vegetation on Mars.
 - D. There is no evidence that the vegetation on Mars because of our own lichens and mosses.
2. We may suppose that vegetation exists on Mars because
 - A. The atmosphere on Mars is thinner than the earth's atmosphere.
 - B. There is enough pressure to support plant life.
 - C. We could not breathe the atmosphere on Mars.
 - D. There is very little oxygen.
3. There is not likely to be any animal life on Mars because
 - A. There is very little oxygen in the Martian atmosphere.
 - B. When we got there we will find that the atmosphere on Mars will be of little help to us.
 - C. There is considerable atmosphere on Mars.
 - D. Its day time temperature came up to about 50°F.
4. The Martian atmosphere won't be of much help to visitors from earth because
 - A. There is enough pressure to support plant life.
 - B. The day temperatures on Mars are moderate.
 - C. Such visitors would not be able to breathe it.
 - D. It contains nitrogen.
5. They look rather like vegetation. ...They in line three refers _____
 - A. Astronomers
 - B. Dark areas
 - C. Life on Mars
 - D. Vegetations on Mars
6. There is not evidence for it. 'It' refers to _____
 - A. Vegetation
 - B. Mars
 - C. Evidence
 - D. Mosses

7. Which of the statement is false according to the passage?
- A. The vegetation on Mars will probably on lower stage of development than the vegetation on earth.
 - B. The Martian atmosphere is enough to support plant life.
 - C. There is a big difference between day and night temperatures on Mars.
 - D. Although there is only a thin atmosphere on Mars, some earth type animals would be able to breathe it.
8. But we must be cautious, "cautious" means:
- A. Serious B. Careful C. Certain D. Dangerous
9. There is considerable atmosphere on Mars, "considerable" means ____
- A. Well-based B. Extreme C. Very little D. Dangerous
10. Mars contains lowly vegetation. Lowly means:
- A. Dense B. Fall C. Simple D. Short
11. The Martian atmosphere is made up mainly of nitrogen; "made up" means _____
- A. Supported B. Consisted of C. Formed from D. Created by

Appendix 6: DMSS Reading Skills Test

I. Read the following passage carefully and answer the questions that follow.

Anyone of us is responsible for his/her actions, feelings and thoughts and their consequences. Others may seek or wish to be responsible, for a person may use various means to influence another's beliefs, feelings etc., but people still ought to be responsible for themselves. It may feel easier to divert responsibility to others, blame others for our failures and shortcomings, but the harsh and empowering reality is that we are responsible for ourselves. We get to decide what we do, how we feel whether we change or not and for these choices we are ultimately accountable, even if only to ourselves.

I am responsible to others by acting ethically, legally and morally, and by communicating my feelings, thoughts, needs, etc. appropriately. I am accountable to my children to teach them good principles so that they can survive and prosper in their lives. I am responsible to my work colleagues to perform my tasks effectively and in a manner that supports our joint activities. Being responsible to other requires trust and vulnerability on our part, difficult when we have fears and feel the need to be protected. The result is greater opportunity for more meaningful relationships, which can endure problems with greater resistance.

Responsibility is positive, powerful and forward-moving, and enables us to engage with the world effectively, accountably and can provide the liveliness needed to pursue our goals, dreams and desires. The opposite of responsibility is irresponsibility. Irresponsibility is draining. Through irresponsibility we rob ourselves and to some extent others of energy and power to make choices and exercise free will...

1. How is one accountable to his children? (Paragraph 2).
 - A. By performing his tasks effectively in a manner that supports his colleagues.
 - B. By teaching them good principles.
 - C. By acting illegally towards them. D. All of the above are correct.
2. What do you think this text is written for?
 - A. Students B. Teachers C. Society D. Doctors
3. What does the word "others" in paragraph 3 refer to?
 - A. Responsibilities B. Human beings C. Feelings D. Educators
4. To blame others for our failures and shortcomings is _____
 - A. Harsh and empowering reality.
 - B. Easier to divert responsibility to others.
 - C. Reality that shows we are responsible to ourselves.
 - D. A sign of responsible person.
5. Acting ethically, legally and morally is _____
 - A. Showing irresponsibility to others.
 - B. Showing responsibility to others.
 - C. Appropriate way of exercising responsibility.
 - D. B and C are possible answers.
6. "Shortcomings" in paragraph 1 (line 5) refers to _____
 - A. Strong sides B. Weaknesses C. Shortcuts D. Successes
7. Through irresponsibility _____
 - A. One robs him/herself and to some extent others.
 - B. We rob ourselves and the energy and power of others.
 - C. We should be role models. D. A and B are possible answers.
8. "Vulnerability" in paragraph 2 (line 16) means
 - A. Sensitivity B. Accountability C. Possibility D. None
9. "Meaningful" paragraph 2 (line 17) means
 - A. Sensible B. Careful C. Changeable D. Undefined
10. "Blame" Paragraph 1 (line 5) means
 - A. Criticize B. Accuse C. Complain D. All

Declaration

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

Minwagaw Temesgen

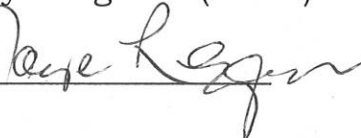


Date

12/10/2001 E.C

This thesis has to be submitted for examination with my approval as a university advisor.

Taye Regassa (Ph. D)



Date

19/9/03