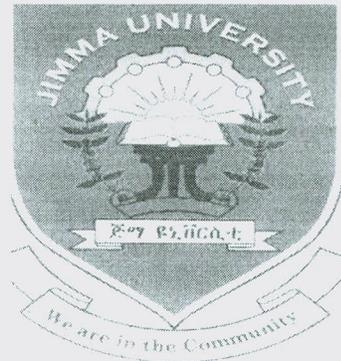


Acc-43661
C-1 JIMMA UNIVERSITY
COLLEGE OF SOCIAL SCIENCES AND LAW
DEPARTMENT OF ENGLISH
(GRADUATE PROGRAMME)

PN
4001
.A87
2013



0043661/007

An Exploration of Teachers' Beliefs and Practices of Alternative Assessments to Improve Learners' Speaking Skills:

Public High School EFL teachers in and around Jimma Town in Focus

By Assefa Gebre Miko

A Thesis Presented to the Department of English

(Graduate Program)

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

PN
4001
.A87
2013
Doc.

June, 2013

Acknowledgements

I am grateful to my advisor, Yohannes Tefera, for his unfailing support and continuing faith in me, and for his kindness to exert quiet and positive rather than voiced pressure on me to get this work finished. Yohannes encouraged me up to the last date of accomplishing my study.

I also would like to thank my co-advisor, Dr. Yemanebrehan for he commented me how to proceed the study at early of the proposal defense.

My classmates, my counterparts and my family (Mira and my brothers) have encouraged and cheered me up whenever my research has made me feel insecure. I say thank you all!

Abstract

The main purpose of this study was to examine what EFL teachers' beliefs and practices of alternative assessments look like in developing students' speaking skills in some government high schools in Jimma Zone. Besides, it studied the most challenging factors on the practices of the teachers as it is known that assessing speaking skills obviously underpass these challenges. To carry out the current research, the researcher designed descriptive survey study. Frequency, percentage, mean values and standard deviation were employed to obtain the main findings. The result showed that EFL teachers have strong beliefs about using alternative assessments but hardly implemented for they were using test-based (traditional) assessment in oral classes. It was also investigated that teachers' practices were highly challenged by large classes, teaching loads and learners' background respectively. Therefore, it was recommended that all concerned bodies and the teachers themselves have to strive to alleviate the investigated problems.

Table of contents

Contents	Pages
Acknowledgements.....	I
Abstract	II
Chapter-One: Introduction.....	1
1.1. Background to the Study.....	1
1.2. Statement of the Problem.....	3
1.3. Objectives of the Study.....	5
1.4. The Significance of the Study.....	6
1.5. Delimitation of the Study.....	6
1.6. Limitation of the Study.....	7
1.7. Acronyms and Definitions of Key Terms.....	7
Chapter-Two: Review of Literatures	11
2.1. Reasons for assessing Speaking Skills and Using Alternative Assessments.....	11
2.1.1. Reasons for Assessing Speaking Skills.....	11
2.1.2. Reasons for Using Alternative Assessment.....	17
2.1.2.1. Main Categories of Alternative Assessments.....	22
2.1.2.1.1. Observational Assessment.....	22
2.1.2.1.2. Self and Peer assessments.....	24
2.1.2.1.3. Portfolio Assessment.....	26
2.1.2.1.4. Direct Oral Testing.....	28
2.1.2.1.4.1. Techniques of coping with challenges in Oral testing	32
2.1.2.1.4.1.1. Holistic Ratings.....	34
2.1.2.1.4.1.2. Analytic Ratings.....	36
Chapter-Three: Design and Methodology.....	40
3.1. Research design.....	40
3.2. Samples and Sampling Techniques.....	40
3.3. Data Collection Instruments.....	41

3.3.1. Document Analysis.....	41
3.3.2. Questionnaires.....	41
3.4. Methods of data analysis.....	44
Chapter-Four: Data Presentation and Analysis.....	45
4.1. Teachers' Background Information.....	45
4.2. Analysis of Lead-in Questions.....	46
4.3. Analysis of Teachers Beliefs on Alternative Assessments.....	47
4.4. Analysis of Data on the Challenges Teachers Face.....	50
4.5. Analysis of Teachers' Practices of Alternative Assessments.....	52
4.6. Analysis of Data Gathered from Students.....	54
Chapter-Five: Summary, Findings and Recommendations.....	56
5.1. Summary.....	56
5.2. Findings.....	57
5.3. Conclusions.....	58
5.4. Recommendations.....	59
Bibliography.....	61
Appendices.....	68
Lists of tables	
Table 3.1.....	43
Table 4.1.....	45
Table 4.2.....	46
Table 4.3.....	47
Table 4.4.....	50
Table 4.5.....	52
Table 4.6.....	53

Introduction



1.1. Background of the Study

The emergence of new approaches, theories and methods in the field of language teaching and learning has affected the beliefs of many researchers on the assessment and testing language performance of many learners.

While large-scale standardized language tests are receiving strong attention in the literature and research, relatively little interest has been directed at what teachers actually do in the classroom to assess learners. This lack of interest in classroom assessments can be explained by the fact that large-scale tests have a major impact on the lives of test-takers and institutions, and are therefore often considered more significant for testing institutions and researchers to study (Law and Eckes, 2007:13). Authors like Genesee (1994) and Tsagari, C (2004) pinpointed the purpose of assessment as follows:

- to motivate pupils and teachers
- to inform the teaching and learning process
- to inform 'relevant others'
- to encourage cooperative styles of work
- to encourage responsibility and involvement, and
- to effect a healthy backwash upon learning and teaching.

Therefore assessment should not be simply an 'add-on' activity, seen as distinct from the rest of the program and carried out under formal conditions at the end of a large block of work or at the end of a course. Assessment should be regular and integrated into normal teaching and learning. Assessment has now taken on a high profile and is required to fulfill a wide range of purposes. Classroom assessment in foreign language (FL) teaching is not a separate element, it is an integral part of a complex jigsaw.

Luoma (2001 and 2004) has forwarded that assessment of spoken language abilities deserves further studies due to the nature of spoken language and the challenges in its assessments. The birth of communicative language teaching approach outshined the need to teach language and assess learners' progresses on these bases. In the former times, linguists believed that testing language ability on the basis of separate elements would have been the best way of measuring learners' performance. Today, a lot of literatures depict that oral assessment is a much more comprehensive term that encompasses largely observational assessment and recording information about the learners' progress. Language testing, on the other hand, is one of the essential part of assessment commonly used for summative purpose; for instance, teacher made classroom tests, terminal tests and high-stake tests. While these types of assessment are mainstay in the history of language assessment, educators and critics from various backgrounds have been raising a number of concerns about the usefulness of testing as a primary measure of student achievements (see Black, 1998; Heaton, 1990; Luoma, 2001 and 2004; Marshall, 2011; Law and Eckes, 2007; Bailey, 1999; Baker, 1989; Bailey, 1998; Simachew, 2012).

As testing for language ability through traditional ways has received its criticisms, many current authorities are devising assessment approaches. Multiple-choice, matching, true-false responses and mere gap fillings are becoming inconvenient measures of speaking abilities though they are used as a tool for collecting information. Furthermore, they lack some qualities in collecting information about learners' motivation, learning strategies and learning styles (Hamayan, 1995; Shohamy, 1998; Brown and Hudson, 1998a). Thus, there are beliefs that test-based assessment overemphasizes the grading function more than learning function in language classrooms. Black (1998) and Bloxham and Boyd (2007), for instance, illustrate that there would be high tendency of focusing on the normative evaluation rather than criterion based evaluation with frequent testing approaches which in turn create competition among learners rather than personal improvement. These tests are simply to show the ranks from the highest to the lowest without the consideration of the above elements.

Those learners who will fail one or more tests but still have some sort of proficiency will be demotivated and lose confidence to continue to learn the new language. On one side, learners who are talented in paper-and-pencil tests lack opportunities to investigate their oral language abilities in actual situations because there is low tendency to expose them to the communicative situations. Moreover, contents of speaking skills tests are usually undermined by school teachers in favor of the other skills such as grammar, reading comprehensions and vocabularies.

Yet, Dickins(2004 as cited in Shohamy,1998) also expresses the role of teacher is crucial in the use of assessment as part of learning. She notes that foreign language teachers are in reality agents of assessment, in that they routinely design and implement assessments and interpret student performance resulting from those assessments. The issue we have to raise here is whether EFL teachers in the selected schools of Jimma Zone believe in and make use of the available alternatives meaningfully, because some investigators forward that most teachers of EFL are attracted by traditional objective tests for these are easily quantifiable and marked. The availability of commercially produced objective tests, the lack of sufficient training on using rubrics and confidence in the subject matter etc. have driven these teachers towards this. As a result of this, as a lot of scholars speculated in their scripts, students who are gifted in a paper-and-pencil environment tend to become weak in real-life oral abilities. To end this, attention is being directed toward alternative assessment regardless of its challenges(Mcloughlin and Lewis,1994,Baron and Boschee,1995; Brown,1994 and 2004,Law and Eckes,2007; a Conference on Computer-Based Assessment (CBA) of Speaking skills held in Brussels in June 2010 by European Commission).

1.2. Statement of the Problem

Until recently the interconnection between teachers' beliefs and practices of alternative assessment, being the current approach to language assessment, has not been explored yet in local studies and particularly in public Secondary Schools of Jimma town and the surrounding schools.

Most of the times assessment of oral skills have been carried out to the knowledge of the researcher as in **testing discrete elements** of the skills such as grammar, vocabulary and

reading comprehensions. In some schools this raises to 50% within teaching periods and 50% in the final examination as it was informed to the current researcher.

It is also believed many teachers, students and people have passed under the umbrella of such assessment even though scholars in the field of SLA research have made great strides towards alternative assessments. Testing agencies such as IELTS, ACTFL, UCLES, TOEFL, OPI and FSI etc.(see these acronyms in section 1.7)have implemented and implicated for such assessments. By the same token, domestic teachers engaged in training EFL teachers to carry out alternative assessments to maximize the oral abilities of learners in Ethiopian schools apart from traditional ways of assessment. Unfortunately there comes a lot of complaints about the oral proficiency of students at various levels (see Animaw, 2012; Simachew, 2012; Zerihun, 2009;Mebea,2008;Kifle,1995;Teshome,1995;Wagari,1995;Tadesse,1990;Tesfaye,1982)etc.so that the researcher suspected about the practices of alternative assessments within the existing challenges to enhance the speaking skills of secondary school students.

The researcher has also been teaching in Jimma College of Teachers Educator(JTCE)-South West Ethiopia. He has got chances to teach English language to those students who join this college from the zone. Unfortunately oral proficiency of these students is not at reasonable development as to the expected level. This also initiated the researcher to investigate teachers' practices of alternative assessments required for oral language development for learners are expected to pass from grade level to the next level after effective assessment practices have been underway as well as with the reasonable skills development.

EFL teachers in the schools understudy may also have or lack of strong beliefs on the practices of alternative assessments to enhance their learners' oral skills. Teachers' beliefs are built up gradually overtime and consist of both subjective and objective dimensions which may range from simple to complex (Richards and Lockhart, 1996 and Marshall, 2008;Chang, 2005). Partly because beliefs may grow from various sources such as knowledge of the existing theories, approaches, methods, personality factors, syllabus materials and even the established curriculum in general. Teachers may hold a strong view about the uses alternative assessments but unable to implement due to workloads, backgrounds of learners, their own proficiency level, nature and levels of tasks in the textbooks, lack of communication among teachers and

logistic materials etc. Thus the teachers' beliefs may affect their assessment practices or enhance it.

1.3. Objectives of the Study

0043661/1001

This study has the following general objectives:

- To investigate how much EFL teachers' beliefs reflect their practices of alternative assessments in developing learners' speaking skills.
- To explore the most serious challenges teachers face in the practices of alternative assessments in speaking skills development if any

The specific objectives of this study are:

- To explore EFL teachers' beliefs about the uses of alternative assessments
- To explore the extent to which EFL teachers practice various alternatives to assess speaking skills of learners
- To identify factors, if any, that have most affected teachers' practices of the alternatives in assessing learners' oral skills
- To identify which alternative(s) most EFL teachers apply in assessing learners' oral language abilities
- To identify which alternatives grade nine students like to be assessed with for their oral skills development.

Therefore the following research questions have been constructed:

1. What do EFL teachers' beliefs look like towards using alternative assessments?
2. To what extent EFL teachers' beliefs reflect their actual practices of alternative assessments?
3. Which of the alternative assessments EFL teachers most frequently apply in assessing speaking skills of the learners?
4. What serious challenges EFL teachers face in implementing alternative assessments in their schools?
5. Which of the alternatives students like to be assessed with for developing speaking skills?



1.4. Significance of the Study

Enormous efforts are being made to bring oral language abilities world-wide for academic, business, or for diplomatic purposes etc. However, the practice of assessing communicative language abilities of learners through alternative assessments lag behind either because of the lack of know-how, beliefs or the existing challenges in schools. Thus, the findings of this study are hoped to contribute valuable information to the concerned stake-holders and English language teachers in particular. On the other hand, Other EFL investigators who will read this paper may be initiated to apply similar studies in solving their students' oral skills problems since this paper provides information on how to do so. Partly because the researcher himself didn't have these information in sufficient amount before joining the MA program, particularly on oral skills rating systems. Thus, there may be similar fellows who want to make use of these procedures in alternative assessment applications. The results may also implicate the reason why many local researchers complain about current learners' oral performance in various fields of studies at higher institutions.

1.5. Delimitation of the Study

This study focuses on speaking skills assessment rather than other skills even though literatures on alternative assessments entail other broad issues. There were about 30 government secondary schools in Jimma Zone. Among these, 7 schools which make 23% of them were purposely selected because the researcher cannot easily access the remote schools in the Zone for financial capacity and lack of transport and safety roads to do so. Therefore, the nearby schools in the radius of 55 kilometers were targeted. Only government secondary schools focused on rather than those private secondary schools because many of the private schools lack the students' textbooks to use in classrooms. The researcher has taken only few students for triangulation of data even though the main subjects of the study are the teachers.

1.6. Limitation of the Study

Students' beliefs are not investigated. Moreover, literatures on alternative assessments present lots of these types and for various purposes, but to corroborate them here for this study will make the statistics difficult as the sample size is also large enough. Therefore, the researcher states the main types in the literature section and more focus on oral production tests as the best alternatives of such assessments because "The more direct the measure, the more useful the results"(Ur, 1991; Moore, 2001; Brown, 2004; Edge, 1993; Underhill, 1987; Heaton, 1988 and 1990; Doff, 1988; Congelosi, 2000; Byrne, 1987; Norrish, 1983; Soler and Martinez, 2008).

The literature section of this study also mainly proceeds from general to specific because the researcher couldn't find more literatures on some alternative assessments that have directly stated speaking skills assessment-for instance, portfolio assessment in spoken language. However, in this regards some ideas are derived through contexts. Since the study primarily emphasizes on EFL teachers' beliefs and the practices of alternative methods in assessing oral skills, only few student population are used to triangulate the actual practices of these teachers. This study also does not discover and publicly declare its new theories and methods of assessing speaking skills but forward various alternatives and procedures that have already been proposed by scholars and language testing agencies in various countries.

1.7. Acronyms and Key terms

Acronyms

AA= Alternative Assessment

DOT= Direct Oral Testing

CLT= Communicative Language Teaching

TSE=Test of Spoken English

EFL=English as a Foreign Language

ESL= English as a Second Language

IELTS=International English Language Testing System

FCE=First Certificate in English

ACTFL=American Council on Teaching Foreign Language

CERF= Common European Framework of Reference

UCLES=University of Cambridge Local Examinations Syndicate

FSI= Foreign Service Institute

OPI = Oral Proficiency Interview

TOEFL=Test of English as a Foreign Language

CBA= Computer Based Assessment

Key terms: alternative assessment, oral skills, oral assessment, beliefs, practices, Oral test,
Oral performance

Several terms, specifically or in their crude form, are related to oral skills assessment, but impossible to mention all of them here. However, the main ones which are important in forwarding this study will be defined below.

Beliefs: As Graves(2000) states a belief is a mental representation of reality which contains meanings, preferences and attitudes that allow the rationalization of complex and different categories of experience that may have some other driven factors. Teachers' assessments of student behavior and performance, among others, are shaped by the theories they have in relation to teaching, assessment, and the nature of learning. Since beliefs and practices are connected in one way or another, teachers may hold beliefs that are compatible with their practices or may not.

Alternative Assessment: For some educators, alternative assessment is a term adopted to contrast with test-based assessment, e.g. professionally-prepared objective tests consisting mostly of multiple choice items especially in the US tradition (Huerta-Macias, 1995).

Alderson and Banerjee (2001 as cited in Chang, 2003) provide the following definition:

Alternative assessment' is usually taken to mean assessment procedures which are less formal than traditional testing, which are gathered over a period of time rather than being taken at one point in time, which are usually formative rather than summative in function, are often low-stakes

in terms of consequences, and are claimed to have beneficial washback effects(p: 228) .

This implies that alternative assessment is any options we use to assess learners performance. This can entail also written tests, journal and diary writing, peer-group evaluations, formal and informal observations, portfolios, direct performance testing and effective questioning in the teaching classrooms etc.

Oral test: is testing the ability of language learners to interact with other speakers to make meaning as distinct from their ability to perform on discrete- point tests of grammatical knowledge. It is directly testing the oral performance of the learners through designed tasks for such purposes-interviews, role plays, narrating, debating and interpreting etc.

Checklists: a list of criteria, or things to look for, on the basis of which a performance or an end product is to be judged. A checklist differs from a rating scale in that it indicates the presence or absence of specified characteristics.

Rubrics: Rubric scoring is becoming a common form of assessment. A rubric can be viewed as a scoring tool that lists the criteria for a piece of work, or “what counts”. It also articulates gradations of quality for each criterion, from excellent to poor.

Testing: is a systematic procedure for measuring an individual’s behavior. This implies that a test must be developed systematically (using specific guidelines) and must provide a procedure for responding, a criteria for scoring, and a description of student performance levels. A test may be either oral or written. They can be conducted with formal and informal procedures. Formal tests are structured assessment procedures with specific guidelines for administration, scoring, and interpretations of results.

They are standardized tests. Informal procedures less structured or structured differently from standardized tests because there is element of subjectivity in administration, scoring(if they are scored) and interpretation (Mcloughlin and Lewis,1994:8).

Summative assessment: is concerned with recording the overall achievements of a learner judged against expectations. It is primarily aimed at determining student achievement for grading purpose which provide the school with a rational for passing or failing students and usually based on a wide range of accumulated behaviors, skills and knowledge. Therefore, both formal and informal assessments and tests are used for summative evaluation purpose.

Thus, teachers have to harmonize both types in their summative judgments. This will be fairly discussed in the later sections.

Validity: is the extent to which the test measures what it is supposed to measure. A test or an assessment must aim to provide a true measure of the particular(the assumed) skill or knowledge which it intended to measure. Validity helps to make inferences from assessment results to what level these results are appropriate, meaningful and useful in terms of the purpose of assessment(Luoma,2001 and 2004; Knight,1992;Weir,2005; claxton,2008;Hughes,2003; Bachman and Palmer,1996; Oller,1979,Kifle,1995).

CHAPTER-TWO

Review of Related literatures

2.1. Reasons for the Assessment of Speaking Skills and Alternative Approaches

Before discussing the possible alternatives in assessing speaking skills, it is necessary to provide rationales for oral language assessment and alternative assessment approaches. It is on this basis that the later discussions be vivid to readers because the discussion goes on from general to specifics as there were few literatures that forward alternative assessment by sticking to particular skills in language science. None the least, some literatures in their sub-topics and sources in some reputable journals try to provide possible ways in applying alternative assessment methods to oral skills. Among these, oral production (direct) tests are the leading alternative methods that many literatures and journals directly raise unlike the rest in assessing speaking.

2.1.1. Reasons for assessing Speaking Skills.

The ideas under this section have been forwarded based on those known authorities like McCarthy(1991), Luoma(2004), Brown and Yule(1983), Soler and Martinez(2008), Aijmer(1996), Cutting(2002) etc.

Speaking is perceived as the process of building and sharing meaning through the use of verbal and non-verbal symbols and it is the heart of language skills. Speaking is both the product and the process of second language acquisition. *Speaking* is the productive aural/oral skill. It consists of producing systematic verbal utterances to convey meaning. Speaking is “an interactive process of constructing meaning that involves producing and receiving and processing information”(Bailey,1998;Byrne,1987;Richards,2008;Joughin,2003,Luoma,2004).

Speaking is an interactive process of constructing meaning that involves producing and receiving and processing information (Brown, 2001; Burns & Joyce, 1997). Its form and meaning are dependent on the context in which it occurs, including the participants themselves, their collective experiences, the physical environment, and the purposes for speaking. It is often spontaneous, open-ended, and evolving. However, speech is not always unpredictable. Language functions (or patterns) that tend to occur in certain discourse situations (e.g., declining an invitation or requesting time off from work), can be identified and charted (Burns & Joyce, 1997). For example, when a salesperson asks, "May I help you?" the expected discourse sequence includes a statement of need, response to the need, offer of appreciation, acknowledgement of the appreciation, and a leave-taking exchange. Speaking requires that learners not only know how to produce specific points of language such as grammar, pronunciation, or vocabulary (linguistic competence), but also that they understand when, why, and in what ways to produce language (sociolinguistic competence). Finally, speech has its own skills, structures, and conventions different from written language (Burns & Joyce, 1997; McCarthy, 1995; Cohen, 1996). A good speaker synthesizes this array of skills and knowledge to succeed in a given speech act.

As a language skill, speaking incorporates many complex features such as speech sounds, speech registers, speech situations and contexts at different levels. When people hear someone speak, they pay attention to what the speaker sounds like almost automatically. On the basis of what they hear, they make some tentative and possibly subconscious judgments about the speaker's personality, attitudes, home region and native/non-native speaker status. As speakers, consciously or unconsciously, people use their speech to create an image of themselves to others. By using speed and pausing, and variations in pitch, volume and intonation, they also create a texture for their talk that supports and enhances what they are saying. The sound of people's speech is meaningful, and that is why this is important for assessing speaking.

The sound of speech is a thorny issue for language assessment, however. This is first of all because people tend to judge native/nonnative speaker status on the basis of pronunciation. This easily leads to the idea that the standard against which learner pronunciation should be judged is the speech of a native speaker. But is the standard justified? And if it is not, how can an alternative standard be defined? The native speaker standard for foreign language

pronunciation is questioned on two main accounts (see e.g. Brown and Yule, 1983: 26–27). Firstly, in today's world, it is difficult to determine which single standard would suffice as the native speaker standard for any language, particularly so for widely used languages. All languages have different regional varieties and often regional standards as well. The standards are valued in different ways in different regions and for different purposes, and this makes it difficult to choose a particular standard for an assessment or to require that learners should try to approximate to one standard only.

Secondly, as research into learner language has progressed, it has become clear that, although vast numbers of language learners learn to pronounce in a fully comprehensible and efficient manner, very few learners are capable of achieving a native-like standard in all respects. If native-like speech is made the criterion, most language learners will 'fail' even if they are fully functional in normal communicative situations. Communicative effectiveness, which is based on comprehensibility and probably guided by native speaker standards but defined in terms of realistic learner achievement, is a better standard for learner pronunciation.

Pronunciation or, more broadly, the sound of speech, can refer to many features of the speech stream, such as individual sounds, pitch, volume, speed, pausing, stress and intonation. An important question is whether all of these can be covered under one rating criterion. Moreover, should the focus be on accuracy of pronunciation or expressiveness of the speaker's use of voice, or both? The solutions depend on the purpose for which the scores will be used and the importance of the sound of speech for that purpose. If there are many other rating criteria besides pronunciation, fitting accuracy and effectiveness into a criterion like 'naturalness of pronunciation' may be the only option. If the sound of speech is a main focus in the assessment, evaluating aspects of it separately gives material for more detailed feedback.

A focus on pronunciation accuracy is attractive because it can be judged against a norm and, even if the norm is not easy to define given the discussion above, gross deviations from it are easy enough to notice. Since accuracy is related to comprehensibility, it is often at least one aspect of a pronunciation criterion, but comprehensibility is much more than accuracy. It often includes speed, intonation, stress and rhythm, all of which may be more important for the overall comprehensibility of the talk than the accuracy of individual sounds. If the emphasis in

the assessment is on ability to create meaning in discourse, the developers might want to evaluate 'interactional efficiency'. This would encompass the examinees' use of stress and intonation to highlight important phrases, or to suggest in what particular way (e.g. ironically) their words should be interpreted. In yet other contexts, they might want to focus on 'expressiveness' as indicated by the general texture of the talk, the speaker's use of speed and pausing, and variations in pitch, tone and volume. This might be especially relevant in tasks such as creative storytelling or certain kinds of role plays, where liveliness of expression is a central element in task performance. Thus, in designing assessment criteria, the developers need to consider the type of information about the sound of speech that they need. They also have to make sure that their tasks give enough material for rating these features, and that they develop the criteria that serve their needs.

Subsequently, both first and second language learners' progress is often tracked according to the grammatical forms that they can produce accurately. Learners are seen to proceed from knowing a few structures to knowing more and more, from using simple structures to using more complex ones, and from making many errors to making few if any at all. Learner grammar is handy for judging proficiency because it is easy to detect in speech and writing, and because the fully fledged grammars of most languages are well known and available for use as performance standards. However, the grammar that is evaluated in assessing speaking should be specifically related to the grammar of speech.

A speech can also be considered to consist of **idea units**, which are short phrases and clauses connected with *and*, *or*, *but* or *that*, or not joined by conjunctions at all but simply spoken next to each other, with possibly a short pause between them. The grammar of these strings of idea units is simpler than that of the written language with its long sentences and dependent and subordinate clauses. This is because speakers are trying to communicate ideas that listeners need to comprehend in real time, as they are being spoken, and this means working within the parameters of the speakers' and listeners' working memory. Idea units are therefore usually about two seconds or about seven words long, or shorter (Chafe, 1985 as quoted in Luoma,2004). The units are usually spoken with a coherent intonation contour, and they are often limited on both sides by pauses or hesitation markers. Many idea units are clauses with a

verb phrase, a noun phrase and a prepositional phrase, but some of them do not contain a verb, and sometimes an idea unit is started by one speaker and completed by another.

None the less, there are of course some situations where complex grammatical features and a high degree of written language influence are not only common but also expected and highly valued. Examples of this include speeches, lectures, conference presentations, and expert discussions where speakers represent their institution or their profession. These situations involve **planned speech**, where the speakers have prepared and possibly rehearsed their presentations in advance, or they express well-thought-out points and opinions, which they may have voiced many times before. **Unplanned speech**, in contrast, is spoken on the spur of the moment, often in reaction to other speakers. It is particularly in unplanned speech that short idea units and 'incomplete sentences' are common, although even in planned speech, idea units are usually shorter than in writing, because the speakers know that their talk has to be understood by listeners in real time.

The concepts of planned and unplanned speech are closely connected to another factor that affects the grammar of speech, namely the level of formality of the speaking situation. Situations that involve planned speech tend to be relatively formal, whereas unplanned speech situations can range from formal to informal. Formal situations require more written-like language with more complex grammar, whereas informal situations call for more oral-like language with strings of short phrases and short turns between speakers.

Many rating scales for speaking include descriptions of vocabulary use, and at the highest levels these often talk about being able to express oneself precisely and providing evidence of the richness of one's lexicon. This can indeed be important in professional contexts or when trying to convey detailed information. Well-chosen phrases can also make descriptions or stories vivid, and learners who can evoke the listener's feelings deserve to be credited for their ability. However, very 'simple' and 'ordinary' words are also very common in normal spoken discourse, and using these naturally in speech is likewise a marker of highly advanced speaking skills. Moreover, there is a core of phrases and expressions that are highly typical for speaking, which contribute to the listener's impression of the speaker's fluency. They work at the interpersonal



level by keeping the conversation going and developing the relationship between the speakers. This aspect of word use should also be rewarded in assessing speaking.

Speaking as a skill is unique in employing generic words. They are fully comprehensible in the speaking situation because they talk about people, things or activities that can be seen or because they are familiar to the speakers. They make spoken communication quick and easy. Generic words may also come naturally to second-language learners, but in a foreign language context where learners have few opportunities to speak the language outside the classroom this feature of spoken language may be harder to notice and learn. Assessment designers can help this by including descriptions of effective use of generic words in rating scales. This sends the message to learners and raters that generic words are important for the naturalness of talk. Such words are that one, there, here you are, good idea, wonderful, well done etc.

Speakers also need to know words, phrases and strategies for creating time to speak. These are sometimes called fillers or hesitation markers, and they include expressions such as *ah, you see, kind of, sort of, and you know, just you know* as well as whole expressions such as “*That’s a good question, or Now let me see*”. Speakers often use repetition of their own words, or of those used by the previous speaker, to achieve the same purpose, i.e. to keep the floor while formulating what they want to say. These expressions are very common in native speaker speech. However, for some reason their appearance in test performances by foreign language learners is sometimes frowned upon. When writing assessment scales, test developers should perhaps consider if examinees who manage to use such expressions successfully in a test situation should be rewarded for it instead.

In general, the above features in oral language make us not to overlook the assessment of speaking skills through alternative methods other than traditional ways of testing for this skill because many of its features cannot be easily measured with paper-and-pencil testing. It is also believed that in communicative language teaching approaches and learning theories, at least five distinctive language uses are expected to be developed by learners (McLoughlin and Lewis, 1994 :415; quoted from Wiig and Semel). These are the informing, controlling, feeling, ritualizing and imagining abilities in language use. Aijmer (1996) also named these functional elements as “conversational routines” in English language. Thus, it is possible to have a view that the main

goal of most English language teachers (either of EFL or ESL) is to enable their students to have these abilities in the target language. By the end of primary schools, as to the above authors, students are expected to accomplish each of these language functions which in few cases depend on the situations in the countries, wars, beliefs about the new language and political situations etc. may affect this.

2.1.2. Reasons for using Alternative Assessments

Alternative assessment is a form of assessment mode which helps the teacher to assess student in a different area. In recent years language teachers have stepped up efforts to develop non-test assessment options that are nevertheless carefully designed and that adhere to the criteria for adequate assessment. As Brown(2004)proclaims, such innovation is referred to as alternative assessment. Alternative assessment is a blanket term that covers any number of alternatives to standardized tests. While the traditional paper and pencil tests may be effective to assess some of the receptive skills such as listening, they are not sufficient to assess the productive skills of speaking and writing. For a long time, student learning was measured only by testing in traditional school settings. Currently, it is realized that there is not only one way of gathering information about student learning. Furthermore, testing is seen as only one part of assessment and a broader concept of assessment is being widely used.

Alternative assessment has been started as a means for educational reform due to the increasing awareness of the influence of testing on curriculum and instruction in U.S.A during the 1980s (Law and Eckes,2007:13). According to Bailey (1998), traditional assessments are indirect and inauthentic. Bailey adds that traditional assessment is standardized and for that reason, they are one-shot, speed-based, and norm-referenced. Law and Eckes(2007) underline the same issue and state that traditional assessments are single-occasion tests. That is, they measure what learners can do at a particular time. However, test scores cannot tell about the progression of a student. Similarly, they cannot tell what particular difficulties the students had during the test.

Law and Eckes(2007) point out most standardized tests assess only the lower-order thinking skills of the learner; focuses on learner's ability of memorization and recall, which are lower level of cognition skills. Additionally, traditional assessment tools require learners to display

their knowledge in a predetermined way. As the scholars state, alternative assessment assesses higher-order thinking skills since it is holistic by nature, or oral skills can be evaluated in integrative ways.

Authors like Brown(2004), Law and Eckes(2007), Brown and Hudson(1998a), Muller(n.d), Watt(n.d), Hamayan(1995), Kubiszyn and Borich(2003) generally stated that alternative assessment:

- provides a continuous flow of information, not intermittent like tests, now - not information after fact. So alternative assessment allows us to diagnose, prescribe, intervene and certify progress in a continuous cycle
- uses real-world contexts or simulations
- is driven by the concept of improvement unlike test which is summative driven
- allows students to be assessed on what they normally do in class everyday
- focuses on both processes and products.
- provides real information about both strengths and weaknesses of students. That is no one kind of assessment can meet every body's information needs
- enables to make appropriate decisions about what to teach next based on which areas need more work and which areas the students have mastered over
- encourages open discourse of standards and rating criteria for final grading/markings
- informs students and teachers about themselves
- paves way to investigate new instructional and assessment roles(e.g. teaching and learning strategies)
- enhances motivations, self-instruction, teaching and learning styles in such a way that encourages learning in supportive mood unlike traditional tests(test-based assessment).
- Helps low achievers more than any other learners compared to snapshot tests

Alternative assessment gives learners the opportunity to demonstrate what they learned. This type of assessment approach focuses on the growth and the performance of the student. That is, if a learner fails to perform a given task at a particular time, s/he still has the opportunity to demonstrate his/her ability at a different time and at different situation. Since alternative assessment is developed in context and over time, the teacher has a chance to measure the strengths and weaknesses of the student in a variety of areas, tasks and situations. More

authentic assessment tools, such as portfolios, independent projects, journals and so on, let learners express their knowledge on the material in their own ways using various styles and strategies.

In the same vein, learners have opportunity to practice the authentic activities that they might encounter in real life. These activities allow them to transfer their skills to various real world related settings. Collaborative working is encouraged. Finally, alternative assessments assist instructors to have a better understanding of student learning. That is, looking at the student product rather than scores can allow instructor to get further insights regarding students' knowledge and skills.

Bailey(1998), Brown(2001), Law and Eckes(2007)etc. put the distinctions between the two [traditional and alternative] assessments as below:

Alternative assessment	Traditional assessment
1. Continuous long-turn assessment	1. One-shot, standardized exams
2. Untimed, free-response format	2. Timed, multiple-choice format
3. Contextualized communicative tasks	3. Decontextualized test items
4. Formative, interact feedback	4. Scores suffice for feedback
5. Criterion-referenced scores	5. Norm-referenced scores
6. Open-ended, creative answers	6. Focus on the "right" answer
7. Formative	7. Summative
8. Oriented to process and product	8. Oriented to product
9. Interactive performance(authentic)	9. Non-interactive performance(inauthentic)
10. Fosters intrinsic motivation	10. Fosters extrinsic motivation

According to the information provided above, assessment tools in traditional tests such as multiple choices, matching items, true or false items and fill- in items seem to create great dissatisfactions to the proponents of alternative assessments. Where the communicative language curriculum as well as syllabi proposed, ideas of such scholars and language experts are reasonable. Let's take some over view of their justifications below particularly from Law and Eckes(2007).

I. True/false tests: True/false items require students to make a decision and find out which of two potential responses is true. Since they are easy to score, it is easy to administer true/false tests. However, guessing might increase the chance of success by 50%. Especially, when the test item is false, it is quite hard to find out whether the student really knows the correct response. One possible solution is to ask student to provide with an explanation for the incorrect item, or rewrite the statement correctly. However, this affects the ease in scoring negatively.

II. Multiple-choice tests: Multiple-choice tests are commonly utilized by teachers, schools, and assessment organizations for the following reasons (Bailey, 1998, p. 130):

1. They are fast, easy, and economical to score. In fact, they are machine scorable.
2. They can be scored objectively and thus may give the test appearance of being fairer and more reliable than subjectively scored tests.
3. They “look like” tests and may thus seem to be acceptable by convention.
4. They reduce the chances of learners guessing the correct items in comparison to true/false items.

Wagari(1995) and others discussed the disadvantages of multiple choice tests. They explained even their preparation needs cognitive effort, and become harder and more time consuming to create in authentic ways.

Hughes (2003) criticizes multiple-choice tests for the following reasons:

1. The technique test only recognition knowledge,
2. Guessing may have a considerable but unknown effect on the test scores,
3. The technique severely restricts what can be tested,
4. It is very difficult to write successful items,
5. Backwash may be harmful,
6. Cheating may be facilitated.

III. Short-answer tests: In short-answer tests items are written either as a direct question requiring the learner fill in a word or phrase or as statements in which a space has been left blank for a brief written answer. Furthermore, the questions need to be precise. Otherwise, the items that are open to interpretations allow learners to fill in the blanks with any possible information.

However, there are facts we do not deny about such traditional standardized tests. To begin with, traditional assessment strategies are more objective, time effective and reliable if they are prepared carefully. Written tests like dialogue completion may be authentically prepared and cover some of the contents. Despite of their short comings, paper-and-pencil tests are also part and parcel of the existing alternatives.

Alternative assessments, on the other hand, carry some concerns in terms of subjectivity, authenticity, reliability and validity (see Brown, 2004; Baron and Boschee, 1995; Underhill, 1987, Alderson, Clapham and wall,1995; Hamayan,1995). Especially they favor about the high validity, authenticity, washback in alternative assessment tools such as portfolios, performance tests and self and peer evaluation tools. However, alternative assessments can be laborious in terms of time and energy spent by the teacher. For example, the diversity of products in portfolios, which is viewed as one of the most important strengths, can lead problems for the teacher in terms of practicality. They might be harder to score and quite time consuming to evaluate the learner's performance unlike multiple-choice tests, matching, true/false, ordering and gap-filling etc. However, Law and Eckes(2007)noticed about this as "Whether or not you are responsible for testing and placement, the challenges of assessing students' real progress is still your job as a teacher" (page.13).

2.1.2.1. Main Categories of Alternative Assessment

2.1.2.1.1. Observational Assessment

Classroom observation is one of the basic methods in assessing learners' performance. It is particularly important for teachers who need to assess the communicative skills of their students during role plays, small group interactions and during question-answer with teacher (Brown, 1994 and 2004; Genesee, 1994; Genesee and Upshur, 1996). All teachers, being aware or not, observe their students in classroom almost in every lesson so we assume that teachers know about their students even without giving a test; this should be supported to some extent empirically for the other professionals, learner themselves, parents and educational agencies may need the assessment outcomes in one way or another. Observation is not new for teachers but knowing about a student in a sense doesn't mean that he or she is backed up on weaknesses such as uttering a word, complex clauses, conversational routines, repairing interactions, hesitations and fillers and turn takings etc.

A major advantage of observational assessment over other methods is that it can be done without interfering with what is being observed. On the basis of their observations, teachers can assess what students could and couldn't able to perform. They may infer the learning strategies students may be using; the teaching strategies they have used; the tasks and materials the students enjoy; and so on. Information derived from such observations is fundamental to the day-to-day functioning of the classroom because it provides a basis for understanding what is happening and for making decisions about both the instructional process and learner performance.

Apart from this, repeated observation may increase the reliability of the information a teacher obtain because student performance on certain occasions may be subject to momentary extraneous influences hence doesn't reflect accurately what the students have really learned. Moreover, records of assessment by observation should be systematic, complete and explicit if they are to be used (Brown and Hudson, 1998a ; Shohamy, 1998; Hamayan, 1995; Brown, 2004; Tsagari, 2004). However, most teachers rely on informal observations without providing formative feedback to learners. Their feedbacks are based on traditional test results and usually

commented as unfair on the side of the learners. Genesee(1994) stated the following about performance observations as:

Observations that are not recorded in writing or some other permanent ways are likely to be forgotten or distorted after some time. It is unreasonable for teachers to think that they can remember the numerous important details of classroom life over time without recording them for later references.

Therefore, teachers need to plan for observation of performance at least by identifying what to observe; for instance individual and peer-group performances, and this time by choosing and devising techniques for recording the assessments. Many authors confirm that there are three ways of recording assessment by observations- anecdotal records, checklists and rating scales.

Anecdotal records(or notes) can be made on file cards, adhesive labels, or clipboards with notepaper that are left in strategic locations so that we can record our observations quickly and easily no matter where we are or what we are doing. Alternatively, they can be recorded in a book or journal kept especially for this purpose. It is also important to date each entry and describe briefly the context in which the observation was recorded. Genesee (1994) remarks that teachers have to organize their observation notes in file folders or a note book organized according to student names and instructional objectives. Thus, anecdotal records that are not identified and stored systematically and quickly become mere pieces of paper with random notes on them. What we have to notice here is also organizing observations according to student names, lesson units and tasks requires time, but the results are valuable and can be used with other methods of collecting information on learners' performances.

Checklists and rating scales are discussed here because they are one of the tools with similar features used for recording classroom observations. In comparison with anecdotal records, these consist of pre-designated categories for recording observations, so they require precise and well-articulated categories and criteria for observing and assessing student performances. They may take the forms “Never to Always” or numerical scales like “point-five scales” based on the listed criteria for the observation. They are relatively easy to use in short time than anecdotal records. However, it doesn’t seem that teachers carry out assessment by observations with the effective use of these techniques.

2.1.2.1.2. Self and Peer Assessments

Traditionally, the role of the assessor usually falls to the teacher or lecturer. However, it is often worthwhile to consider involving others in assessment process; or involving students in their own assessment. Self-assessment is the involvement of students in identifying standards and or criteria to apply to their work and making judgments about the extent to which they have met this criteria and standards. Effective and appropriate use of involving learners in the assessment practice can enhance the learning experience, enrich the teaching experience, and reduce the marking burden placed on tutors or lecturers (Harmer 1991).

The development of self-assessment by pupils is still in its early stages, but within the framework of formative assessment as an integral part of learning, it seems a natural, almost essential development, as well as a potentially powerful source for the improvement of learning. Thus improved formative assessment can lead to changes which are of much wider significance-changes which should be a powerful help with students’ personal development and which should also be part of any program to help them to be more effective learners. Moreover, monitoring behavior with carefully planned and prepared self-evaluation checklists can help learners to develop metacognitive skills, enrich their learning strategies, and enable them to become independent, confident learners. In the form of verbal or numerical scales, assessors can carry out learners’ self-evaluation. As Harmer suggests the verbal scales are from ‘Poor to excellent scale’, ‘never to always’ based on statements like the following:

criteria	rarely	some times	very often	always
1.I try to answer teacher's question in classroom				
2.I ask questions when the lesson is not clear				
3.I answer classmate's questions during group activities in classroom				
4.I support friends in pair works and group works in class				
5. I can greet classmates and other peoples effectively in English				
6.I can introduce myself to others effectively				
7.I can ask people for information effectively e.g. Name,address,hobbies, directions, feelings etc.				
8.I can ask for and give directions to others in English language effectively				
9.I can convey greetings and take leave effectively				
10.I can tell my opinions by using basic expressions				
11.I can express my agreements and disagreements to friends				
12.I can listen to others during conversations,role plays and take turns				
13.I can use appropriate and relevant vocabularies to speak				
14.I can say words, phrases and clauses almost correctly				
15.I can read and understand stories, dialogues almost appropriately				
16. <i>I can express my opinions</i>				

One way of increasing the efficiency of assessment is to allow students play a role in assessing themselves or each other. This approach to assessment requires careful planning, agreement of criteria and use of common tools for analyzing marks. In addition to assessment by the teacher, self- and peer assessment are also becoming popular. This is particularly true in classrooms

where teachers wish to encourage learner autonomy and a focus on learning processes as well as learning outcomes. While self-assessment has been criticized on the grounds that not all learners are accurate judges of their own ability, this criticism misses the point to some extent, which is to involve learners in their own learning processes:

The major purpose of self-assessment is to provide the opportunity for learners to develop an understanding of their own level of skill, knowledge or personal readiness for a task in relation to their goals. This level will often be compared with a previously determined level and incorporated either into a summative report of gains made during a course or into a cumulative record of learner achievement(Cram as cited in Nunan,2004)

Underhill(1987) and Nunan(2004) claim that self-assessment approach is the easiest, cheapest and quickest form of assessment. Students can be given materials to guide them and make their own judgments based on the activities they have already carried out with criteria like the above one. Some other writers also believe that self-assessment is one among alternative forms of assessment if they are carefully handled.

2.1.2.1.3. Portfolio Assessment

There is another type of alternative assessment that offers more than a one-time picture of what a learner has accomplished. Its principal purpose is to tell a story of a learner's growth in proficiency, long-term achievement, and significant accomplishments in a given academic area. It is called *portfolio assessment*. The portfolio is a measure of deep understanding like the performance demonstrations covered earlier. But, in addition, it shows growth in competence and understanding across the term or school year.

In general term, a portfolio is a planned collection of learner achievement that documents what a student has accomplished and the steps taken to get there. According to Brown(2004) learners of all ages and in all fields of study are benefiting from portfolio assessments. When we take this to spoken language context, learners' portfolios may entail written forms of oral production skills like role play scripts, dialogues, mini-dramas, vocabulary journals, reflective tasks such as self-evaluations and learning-logs, improvisations of situations, interpretative pictures , glossaries of conversational routines and stories in spoken forms etc. The collection represents

a collaborative effort among teacher and learner, to decide on portfolio purpose, content and evaluation criteria.

Portfolio assessment is based on the idea that a collection of a learner's work throughout the year is one of the best ways to show both final achievement and the effort put into getting there. Portfolio should include samples of both spoken and written language, evidence of reflective learning and creative arts such as dialogue and play constructions etc. If designed properly, portfolios can show a learner's ability to think and to solve problems, to use strategies and procedural-type skills, and to construct knowledge. In addition, they also tell something about a learner's persistence, effort, willingness to change, skill in monitoring their own learning, and ability to be self-reflective (Kubiszyn and Borich,2003:174).One purpose for a portfolio is to give a teacher information about a learner that traditional quizzes and standardized tests cannot provide.

There are other reasons for using portfolios. Portfolios are also means to communicate to parents and other teachers the level of achievement that a learner has reached Report card grades give us some idea of this, but portfolios supplement grades by showing parents, teachers, and learners the supporting evidence. Portfolios are not an alternative to paper-and-pencil tests, essay tests, or performance tests. Thus, if you want to assess both achievement and growth in an authentic context, portfolios are a tool that you should consider. Finally, portfolios are a way to motivate learners to higher levels of effort. Brown says this in other word as “ If portfolios are used with careful plans, effective guidelines and systematic follow up of these, they can raise reliability to respectable level, beneficial washback, validity and authenticity”. More importantly he has compared and contrasted the objective tests and alternative assessments like these on a regression graph and concluded that traditional objective tests cannot offer much validity, authenticity, beneficial washback and the like though they are highly reliable and practicable. As to Brown(2001), one has to flatten these elements through alternative assessments, or able to transform these elements in standardized achievement tests with great creativity and effort.

2.1.2.1.4. Direct Oral Testing

There are two main ways of oral language testing-direct testing and indirect testing. According to Harmer(1991) a test is said to be direct if it takes to perform the communicative skill which is being tested whereas indirect items try to measure student's knowledge and ability by getting at what lies beneath their receptive and productive skills.

Direct test items try to be as much like real life language use as possible, but indirect test items try to find out as student's language knowledge through more controlled items such as multiple choice questions (MCQs) or any form of linguistic competence which are quicker to design scorer reliability. Some professionals name the two types of test as objective [indirect] tests and subjective [direct] tests. To mention some of the objective types are multiple choices, completion items, sequencing or recording ,matching, true-false, form transferring and the like which are most commonly used by school teachers and are traditionally oriented ones. On the other hand, direct tests include structured interview, information gap task, picture cues, role-plays, debates, descriptions, responding to mini-situations, free interviews, telling stories, presentations, impromptu speeches, comparisons, problem solving activities, listening and responding, instructing and questioning etc.

Indirect tests normally cannot accurately and reliably measure the test taker's skill in which the test developer and test user are really interested, and it would be difficult to use results from indirect tests to convince test users (i.e., admissions officers and human resource officers) that such scores can reliably reflect the test taker's oral skill. In other words, this test format entails a problem of test validity. Therefore, in general practice these days, indirect testing as a method has been basically abandoned for speaking assessment.

A second type of oral assessment is called direct testing, or live testing. "Testing is said to be direct when it requires the candidate to perform precisely the skill that we wish to measure" (Hughes, 2003: 17). Direct oral testing is said to be first introduced in the United States in the 1950s when the Oral Proficiency Interview (OPI) test and its accompanying scale were developed at the Foreign Services Institute. Later on, the format was fairly widely adopted and

adapted by testing organizations in other parts of the country as well as in other countries. Now, OPI is generally understood as an alternative name to face-to-face oral assessment.

Direct testing is considered a valid test format because it is generally held that the test result should reflect exactly the level of the type of skill we are interested in. This is also the most common mode for assessing oral proficiency (Luoma, 2004). An oral language assessment in the direct form requires the test taker to perform oral tasks that can demonstrate his or her oral language proficiency. Such a test format is believed to offer a much better test validity because, in a properly designed oral test in the direct testing mode, every effort is supposed to be made for the tasks to be as authentic as possible. In a direct oral assessment, the test taker is required to participate in face-to-face communicative interactions involving one or more examiners, among whom one is the interlocutor. Presumably, results from such a test will be more acceptable to test users because the tasks are supposed to measure exactly what test users are interested in, namely, live oral language ability.

According to Heaton(1988) oral tests are clearly where we should devote a high proportion of class time as the main goal in teaching oral skills of speaking is to develop learners' ability of expressing themselves intelligibly. This author further explains that language testing validity can be achieved at the expense of reliability in oral tests. "Objective tests are clearly not subjected to high degree of unreliability, but does this mean that all forms of subjective should be abandoned in favor of objective tests?", and he asserted " Language learning behavior cannot be demonstrated solely by means of an ability to select correct options from alternatives"(p.164-165). Thus, the motto in preparing communicative based textbooks for language curriculum; particularly for secondary school students in our country is to enable them to use this language for academic and communication purposes. What is so ever the plan is, assessing the communicative language ability is largely being carried out with traditional objective tests which are vastly inferior to validity issues in language assessment (Kifle, 1995 ; Seifu,1997; Wagari,1995).Currently there is a movement toward testing practices of spoken language in some of the secondary schools in the country, but it shouldn't be in an intermittent way.

Oral tests can have excellent backwash effect on the teaching that takes place prior to achievement tests in schools. Consequently it is essential to devise a way to valid oral tests and then to establish ways of maximizing their reliability through alternative ways and systematic rating scales. Many of Scholars and language testing agencies in the western and United Kingdom(UK) suggest some of the following techniques in testing of oral skills ability at different levels and for various purposes:

- 1. Interviewing:** This is the most common of oral tests for many people in the form of conversation or questioning and answering. An oral interview may range from controlled, guided or freer kinds. In spite of its subjectivity, an extremely good test is the oral interview(Heaton,1988:89). Students can be interviewed about their hobbies, families, favorite foods, impressions, fames and opinions they have about something etc.
- 2. Giving instructions, descriptions and explanations:** with minimal preparation, the learner describes a known object, place and a person, a process, a system or daily routines at some length. In this regard choosing something that is familiar to learners is a good way of getting them to produce connected discourse on a given topic.
- 3. Giving Picture cues:** simplified pictures, maps and diagrams can be used in oral production tests. This time students are given a picture, a diagram or a statistical information, brochures and the like for a few minutes and describe or explain them. Advertisement posters, cartoons and chained story pictures are very important as these create authenticity of tasks. Moreover, teachers can integrate some basic oral languages such as greetings, introductions and opening and closing presentation in the form of elicitation.
- 4. Role plays, reporting and group discussions:** students are given one's role and expected to perform as that person though not exactly the same way after they have read in dialogue form- doctor and patient role, a father and a son's role, a policeman and a suspected man's role etc.
- 5. Providing mini situations or improvisation technique:** students are asked to respond with the target oral skills in given context or situation. For instance, a teacher may give a student this situation "You introduce your sister/brother to your classmate
- 6. Interpreting themes of a text:** short stories, proverbs, sayings and quotations are helpful with this technique of testing oral abilities. Literary materials that induce real-life situation encourage even shy students to forward their opinions and beliefs unexpectedly.

7. Reading aloud: learners are given some selected texts to read loudly for testing pronunciation abilities. Key sounds, words or phrases can be chosen and constructed into texts which may have minimal pairs, dialogue to utter with conversational fillers.

8. Effective classroom questioning during lesson: questions that help more to explore learners' oral abilities and needs teacher's creative ability or mental power are deployed to students. There are effective teachers at this technique, soliciting, particularly in interviews and reading comprehensions. Abilities such as comprehending, vocabulary and grammar can be assessed in this way. As Kisko and Iyortsuun(1982) have listed, teachers have ask students yes/no prompts, true/false, inversion and alternative questions, inferential and WH-types of questions effectively in a lesson class.

Tasks for the above techniques are also categorized by Brown(2004) and Luoma(2004) as a whole below:

- Imitative Speaking Tasks such as phonemes, words, phrases and sentence levels focusing on specific criteria, or the majorities are meant to test intonation.
- Intensive Speaking Tasks such as short stretches of discourse like completing the elicited parts of the predictable but omitted phrases, dialogues to fill in. Picture cued tasks can also create interactive opportunity for learners either at intensive or extensive levels. Pictures at the intensive task level should be very simple, designed to elicit a word, a phrase, a clause or a key expression of the target language even though pictures at the sentence level can be best in testing tenses and vocabulary.
- Responsive Speaking Tasks like questioning and answering, describing someone, telling functions and process, asking for and giving directions, classroom reflections and instructions etc.
- Interactive tasks: are those what someone would describe as interpersonal and longer but less than extensive kinds. For example, structured interviews, games or warmers, reaction to dairies, guided role plays, mini oral presentations, impromptu speech and information gap tasks.
- Extensive Speaking tasks which entail complex, relatively lengthy stretches of discourse. Telling narratives, prose summaries, debates, public opinions or panel discussions arranged for classroom purpose, statistical information, book report, compare and contrast tasks etc.

2.1.2.1.4.1. TECHNIQUES OF COPING WITH CHALLENGES IN DIRECT TESTINGS

“Challenges are gifts that force us to search for a new center of gravity. Don’t fight them. Just find a different way to stand” O. Winfrey. Many scholars forward the followings as main challenges in oral language assessments:

- The nature of spoken language itself
- Linguistic background of the learners
- The time available to manage assessment in large classes and workloads
- Teachers’ lack of sufficient scoring skills of oral skills/use of alternatives
- Teachers’ lack of proficiency of the ESL/EFL
- Lack of logistic materials in the schools for oral language assessment
- Reluctance on setting and using common rubrics among school teachers
- Subjectivity(on reliability and validity) issues.

(For more on the above see Louma, 2004; Hughes, 2003; Bachman and Palmer,1996; Weir,2005; Brown and Yule,1983; Heaton,1990; Alderson et al ,1995; Madsen,1983; Oller, 1979; Baker, 1989; Underhill,1987; Genesee and Upshur,1996).

Despite of the above challenges many leading authorities believe that direct oral testing is unavoidable. For examples:

- “If we want to encourage students to speak, we should give oral tests from time to time; otherwise, students will always regard speaking as less “serious” than the other skills”(Doff,1988:266)
- “I think that oral testing is worth the investment: not so much for the sake of overall validity of the proficiency test of which it is part, as for the sake of backwash”(Ur,1991:135)
- “Communication is a subjective affair, and testing this ability should include a subjective element if they are to be valid”(Edge,1993)
- “We suggest that teachers should continue to assess features of spoken language not in isolation but as the part of communicative abilities”(Brown and Yule,1983; Luoma,2004)
- “Language is taught not primarily as mental discipline, but in order to give a learner to use for different purposes. Assess the uses that the learner can make of the code not just the code...

Teachers who are confident of their own ability to pick out the appropriate situation can give students the oral tests in the class”(Norrish,1983:108).

- “Instead of just offering paper-and- pencil single answer tests of possibly hundreds of decontextualized items, performance based assessment or direct testing worth a lot though the testing itself is time-consuming and expensive”(Brown,2001:416).

A great deal of authors who were not mentioned above also believes in assessing and testing oral production abilities regardless of the existing challenges. Some suggests that it is easier in English as Foreign Language classroom rather than in Native Speakers classroom say for example in pronunciations and accents. Thus, many authorities forwarded that we have to do all we can to overcome these challenges as long as there is no absolute validity and reliability achievement at present and there is also no one best method, best assessment and test in itself. For instance, an objective test on itself is at least liable to psychometric effects on the side of the learners. These can be tackled with the help of performance assessment scales proposed at various times in language assessment or performances test rubrics, checklists, authentic tasks and related trainings concerning outshining problems. For instance, rubrics answer the following questions:

Which criteria should be used to judge and/or evaluate performance?

Where should we look and what should we look for to judge performance success?

What does the range in the quality of performance look like?

How do we determine what score should be given and what the score means?

How should the different levels of quality be described and distinguished from one another?

Since rubrics contain the performance objectives, range of performance, and performance characteristics indicating the degree to which a standard of performance has been met, they enable teachers to provide feedback to learners about their progress as well as to evaluate performance. Further, they provide some clues as to what good performance might look like even before learners perform an assessment task (Kubiszyn and Borich,2003). Of additional assistance to learners in demystifying performance expectations is seeing exemplars or models of the performance expected, together with the rubrics.

Rubrics can be used for both impressionistic marking and analytic marking scheme of oral production skills as we see in the next sections.

2.1.2.1.4.1.1. Holistic Ratings

This method is used when an assessor or tester wanted to evaluate oral abilities in general; sometimes emphasizing one among them. J. Harmer, H.D, Brown, Hughes and Alderson give an example of a holistic scale, encompassing many areas of communicative competence in one scale. These scales are very wide bands-dichotomous rubrics and band rubrics are of main types for this purpose.

Holistic(also known as integrative, communicative or impressionistic) approach combines all aspects of language competence and performance into one assessment criteria. According to Heaton(1988)it is designed to assess the learner’s ability to use two or more skills simultaneously such as grammars, vocabularies ,pronunciation, interaction ability, fluency, accent, listening skills and other forms of registers. That means the holistic approach does not rely only on discrete points but all language elements. Thus, the approach involves the testing of language in context and primarily concerned with meaning and the total communicative effect.

Such an approach can be carried out by using verbal or numerical score value on preplanned topic say for example ‘expressing one’s view or own view’ to conversant. So, teachers can use small card, note pad or outlined paper to record individual learners’ oral performance(Underhill,1987; Brown and Yule,1983; G.D Brown,1996; Baldwin,2006;Itkonen, 2010; Baker,1989; Bachman and Palmer,1996;Madsen,1983; Edge,1993 Barker and Westrup,2003). As it was written in the authors’ works, the following assessment format for the holistic marking can be used in communicative classrooms if the task for the oral testing is authentic and proposed by the tester for this purposes.

Name/I .D	Interaction Comprehension						Pronunciation						Grammar usage						Vocabulary appropriacy					
	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6
A																								
B																								
C																								
D																								
E																								

According to Fulcher and Davidson(2007) a speech sample is awarded a single number, and the meaning of that number is intended to generalize to a large number of other possible speaking tasks. Another alternative is to record with verbal expression like “poor to excellent” which can be interpreted into numerical value. For example, Barker and Westrup(2003) and Heaton(1988) suggested the following holistic rating scale for oral interview:

Rating	Oral ability descriptors
6	Completely at ease in use of English on all topics discussed. Very good communication skills.
5	No difficulty in understanding English and there are no problems communicating with students. Good strategies for keeping the conversation flowing.
4	Students make a limited number of errors of grammar, vocabulary and pronunciation but is still at ease in communicating on every day subjects. Knows his/her own mistakes and can correct him/her-self.
3	Occasional difficulties in communicating. Several errors which sometimes make it difficult to communicate with the student.
2	Student’s understanding is severely limited, but communicate on everyday topics is possible. a large number of errors.
1	Cannot understand adequately and cannot him/herself understood.

A six-point scale is important to avoid a middle level. If, for example, a five point scale were used, there would be a tendency for many teachers’ to place a lot of students at level three (the middle level because of the median score system). A six-point scale compels us to decide whether a certain student is slightly above average (level four) or slightly below average (level three), thus enabling us to divide the middle range of students into two levels. Level three (slightly below average) should generally wish to set our fail/unsatisfactory level at level 1 or 2, leaving level 3 and above to indicate pass grades (J.B Heaton,1990.p.71).

Dichotomous rubrics or those similar to checklists can be scored on the bases of verbal scales – yes/no criteria, poor to excellent, unsatisfactory to more than satisfactory etc. After setting

descriptors, we can also use band rubrics ranging as 2,4,6,8,10 for each oral components we want to measure; ten being the maximum and two being the least on the rating scale.

2.1.2.1.4.1.2. Discrete (Analytic) Ratings

Even though assessing and testing oral production skills commonly carried out in integrative ways, teachers test their students on discrete points at a time. Language elements like grammar, vocabulary, listening comprehensions, pronunciations and conversational elements are assessed or tested separately either objectively or subjectively. However, it is also used for summative evaluation purpose. If grammar is tested at one time, then vocabularies or pronunciations are tested at other times. The rating scale can be similar to that of holistic system, but this approach requires high numerical scale, most often from ten to twenty, for they are relatively easier to follow up analytic elements.

1. Pronunciation.

scores	Descriptions
6	Pronunciation good. Only 2 or 3 grammatical errors. Not much searching for words very few long pauses fairly easy to understand. Very few interruptions necessary but mastered all oral skills on course.
5	Pronunciation slightly influenced by L1. A few grammatical errors but most sentences correct. Sometimes searches for words, not too many long pauses. General meaning fairly clear but few interruptions necessary.
4	Pronunciation is influenced a little by L1. A few unnatural pauses, convey general meaning fairly clearly. A few interruptions usually exist, though almost clear in message.
3	Pronunciation influenced by L1 pronunciation and grammar errors-several, errors cause serious confusion, longer pauses to search for word or meaning, fairly limited expression, has mastered only some of oral skills on course.
2	Several serious pronunciation errors. Basic grammar errors; unnaturally long pauses very limited expression, needs some effort to understand much of it. Interruptions often necessary and sometimes has difficulty in explaining or making meaning clearer. Only few of oral skills on course mastered.
1	A lot of serious pronunciation errors, many basic errors. Full of unnaturally long pauses, very halting delivery -extremely limited expression. Almost impossible to understand interruptions constantly necessary but cannot explain or make meaning clearer. Very few of oral skills on course mastered.

2. Grammar

Rating	Descriptions
6	No more than a few minor errors during the interaction
5	Few errors, with no failure of patterns
4	Occasional errors showing imperfect control of some patterns but no weakness that causes misunderstanding
3	Frequent errors showing some major patterns uncontrolled and causing occasional imitation and misunderstanding
2	Constant errors showing control of very few major patterns and frequently preventing communication
1	Grammar almost entirely inaccurate except in stock phrases

3. Vocabulary

rating	Descriptions
6	Excellent cohesion-rarely searches for words to say or misses. Vocabulary apparently as accurate and extensive as that of a native speaker.
5	Vocabulary broad and precise, adequate to cope with more complex problems. Good cohesion, seldom misses or searches for words.
4	Adequate words to participate in interaction, with some circumlocutions.
3	Choices of words sometimes inaccurate, limitation of vocabulary prevent discussion at some stages of the interaction.
2	Vocabulary limited to basic personal and survival areas(time, food, travel, family, body parts...) frequently misses or searches for words and confusing because of moderate cohesion of ideas.
1	Vocabulary inadequate for even the simplest conversation, not possible to discuss any topic for no cohesion at all because of extremely limited word knowledge.

4. Interaction Comprehension

Rating	description
6	Understands everything in both formal and colloquial speech to be expected of a native speaker.
5	Understands everything in normal conversation except for few low colloquial or low frequency items, or exceptionally rapid or slurred speech.
4	Understands quite well normal speech directed to him/her, but requires occasional repetition and rephrasing.
3	Understands careful, somewhat simplified speech directed to him/her with considerable repetition and rephrasing.
2	Understands only slow, very simple speech on most basic topics. Require constant repetition and rephrasing.
1	Understands too little for the simplest type of conversation.

5. Fluency

rating	Description
6	Speech on all topics is as effortless and smooth as a native speaker.
5	Speech effortless and smooth, but perceptively non-native in speed and evenness.
4	Speech is occasionally hesitant, with some un evenness caused by rephrasing and groping for words.
3	Speech is frequently hesitant and jerky, sentences may be left un completed.
2	Speech is very low and un even except for short routine sentences.
1	Speech is so halting and fragmentary that conversation is virtually impossible.

Whatever the criteria chosen, the brief descriptions can be made much more specific at each level in order to reflect the contents of the course being followed. The most important point to bear in mind is that for most classroom purposes the rating scale should not have native speaker performance as the desired goal. Instead it should be based on realistic expectations of what successful learners can achieve at a particular stage in their development(see Luoma,2004; Marshall,2010;Abedi,2010;C.Tsagari,2004;Heaton,1988;Luoma,2001;Black,1998;Irons, 2008; Richards,2008; CBA,2010). According to Baron and Boschee(1995) the integration of the

above marking approaches to oral skills can yield both reliability and high validity if they supplemented the summative evaluation. This procedure is known as **Multiple Validation**. Yet, there are scholars who suggest that if the teachers couldn't practice these, they have to develop high content validity of the oral skills in their achievement tests, but still they are expected to authenticate the exam.

Surely sufficient information are depicted in the above sections that alternative assessments are by far better than multiple-choice items of quizzes, mid-term and final examinations in which many teachers cannot maintain the validity types suggested by scholars. That makes such examinations traditional approach to language assessment, or they must be harmonized with the alternative assessment approach that encompasses portfolios compiled by learners, self-evaluations, formal observations, peer-group evaluations and direct oral tests, journal and critical reflections etc. The doubt is that whether beliefs about and practical implementation of alternative assessments in the proposed schools match or not. It's also well stated that teachers who want to conduct oral tests as alternative to traditional norm-referenced tests can utilize either the holistic or the analytic scoring approaches and overcome the related challenges.

Regardless of their challenges, oral skills are more liable to alternatives methods and tools other than some other macro skills of foreign language. When teachers set criteria and descriptors, they can make it to the level of their learners as the purpose in assessing speaking in classroom is to enhance the communicative ability of learners, styles, strategies, autonomy, both inner and external motivation. Teachers also may have strong beliefs on the raised rationales but lack practices or vice-verse.

CHAPTER-THREE

Design and Methodology

3.1. Research Design

This study has focused on EFL teachers' beliefs and practices of alternative assessments in developing learners' speaking skills, and it was an exploratory study of both abstract behaviors and observable behaviors. Such studies can be carried out with descriptive survey method which in turn invites analysis of frequencies, mean values, percentages and average manipulations (Gray, 2004; Dornyei, 2007; Koul, 1988). Therefore, descriptive survey method was employed to this study to have an opportunity to snapshot on some single items but that still have significant information. Since the major tools were questionnaires, descriptive statistics such as mean, frequency, percentage and standard deviations were used for presenting the data as well as for analyses in order to obtain main findings.

3.2. Samples of the Study and Sampling Techniques

It is common that researchers in the fields of social or natural studies visit either probability or non-probability sampling methods. Probability sampling sub-classified as random, stratified random, cluster and systematic sampling (Dornyei, 2007: 97-99). There were about 30 government secondary schools in Jimma Zone. However, the researcher couldn't easily access the remote schools in the Zone for financial capacity and lack of transport and safety roads. Therefore, the nearby schools in the radius of 55 kilometers were targeted. Government high schools such as Agaro, Yebbu, Seto, Jiren, Ababuna, Asendabo and Serbo (7 schools) that make 23% of the secondary schools in this zone were purposefully selected. There were 50 EFL teachers in the seven schools and only 34 teachers who were teaching English to ninth graders were selected. From these, 25 teachers were males and 9 were females. All of them were purposefully selected for my study and also manageable for the study.

The students were also selected according to the non-probability sampling rules because they were required only for triangulation purpose on the data of teachers' assessment practices. Concerning this the researcher has primarily identified the existing number of sections and a

total of 101 sections were identified. Then after, two students were randomly selected from different sections of grade nine in each school. The selection included equal proportion of gender among 202 students that were sampled on this basis.

3.3. Data Collection Instruments

Researchers use lots of data collecting tools and methods based on the types of their studies. The most common ones are direct interviews, direct observations, document analysis and questionnaires. These data gathering tools are also prioritized according to their importance to the proposed study; be it quantitative or qualitative or both types at once. Therefore, this study mainly utilized questionnaires (closed and open types) effectively followed with some other tools where it is necessitated to triangulate the results from questionnaires.

3.3.1. Document Analysis

The researcher had assessed documents (mark lists and anecdotal records) that teachers used to record their assessments. This supports to know whether the teachers set places for the types of alternatives in their learner assessment documents either formally or informally. Though convincing the teachers were difficult, the researcher was able to reach on agreement with some of them. Then the teachers' names were coded as it was promised not to use their names during research report. They were named as Teacher-one (T1) from Agaro, Teacher-two(T2) from Yebbu, Teacher-three(T3) from Jiren , Teacher-four(T4) from Asendabo and Teacher-five(T5) from Serbo secondary schools.

3.3.2. Questionnaires

Questionnaires are valuable research tools if they are carefully constructed and pilot tested(Gray,2004; Dornyei, 2007 and D. Gall, Borg and P. Gall,1996; Koul,1988). They may range from closed types to open- ended types, hence support to gather wide range of information. Thus, the researcher had prepared such types and administered to all sampled populations. The lead-in questions started the items and the rest questions were suited to descriptive on the bases of likert scales. The lead-in items were used to check whether the issues for the study were known there off by the participants or not. If the participant do not know about those matters, it will be hard for researchers to carry out the study successfully.

The descriptively oriented questionnaires were constructed in the form of multiple choices, “yes or no” checklists, verbal scales such as “never” to “always” or “Very high” to “very low” etc. and from “strongly agree” to “strongly disagree” where they were necessary because this study tries to employ many alternative assessment tools which may need different ways of exploring for information. In order to supplement these types the researcher had also used open response questions. The items were evaluated by advisors, colleagues who had already mastered in EFL and by my intimate class mates.

For the final administration, thirty extra questionnaires were provided to non- sampled students for it was speculated that some personal problems may encounter members of the sampled population. In general, questionnaire items that were based on frequency, degree and agreement likert scale were used with some lead-in and semi-structured open questions. The questionnaire that included agreement scale was distributed only to the teacher population to see their beliefs. The sampled students were also provided the questionnaires that concerned only about teachers’ practices. Finally, these tools and the above instruments for data collection were triangulated to come up with valid research as we see in the next chapter. The following table displays framework of questionnaire to EFL teachers.

Table.3.1 Questionnaire for EFL teachers

Questions		contents		Question types
Part-I	1-5	Background information	Sex, age, experience, Teaching loads, educational status	selection
Part-II	1-5	Lead-ins	Teaching speaking skills, textbook contents, challenges and assessments	Yes/ No alternatives, Some open types
Part-III	6-25	EFL teachers' beliefs About alternative assessments	Theoretical concepts Relationship to practices	Fixed alternatives
Part-IV	26-41	Assessment practices of EFL teachers	Observations, tests Self-assessments, peer assessments, direct oral testing, rubrics and checklists	Fixed alternatives
Part-V	42-49	Challenges in assessing speaking skills	Teaching loads, logistic materials, large classes, subject and assessment knowledge, the learners' background on English subject etc	Both fixed alternatives and open-ends

All of questions that were administered to students were fixed alternatives and except the last 14th item. This last question was constructed to see the alternatives learners prefer for their oral skills assessment in schools and this item was provided with 11 alternatives to choose the best three they like.

The researcher had also tried to calculate the standard deviations of each item on the teachers' beliefs, practices as well as on items for students' response. This was to see how far the scores on

the data deviate from their means and inform about the relationship of the items instead of correlational analysis (see Gray,2004: pp.285-300; Gall et al, 1996: pp.409-415 and G. Brown,1996: pp.123-129). To test through correlations the items should have equal numbers and levels. It was also speculated that learners may not equally respond to items that based on theories, principles and approaches of assessment like their teachers. The students may respond only to what they can observe during the lesson classes.

On these bases, the standard deviations for items 6 to 25 mainly vary between standard deviations of 0.08 to 0.18(Item 22 the lowest and Item 25 the highest). For the items 27 to 41 the standard deviations vary between 0.09 to 0.20 (Item 28 the lowest and Item 38 the highest).When we see the standard deviations of the items for students, they also vary between 0.074 to 0.096(see Item 1 and Item 10 respectively). Therefore, there is no as such great variation between the items for both teachers and learners for the data collections.

3.4. Methods of Data Analysis

After data were collected, the researcher:

- Checked whether all the instruments, especially questionnaires, were filled and returned to the researcher;
- Checked across whether all the given items on the questionnaire sheets were filled by the respondents.
- Developed data categories(typology) for frequency, degree and agreement based items
- Manipulated the tallied data through statistical procedures mainly with descriptive survey and some inferential statistics
- tallied the items and carried out statistical manipulations like percentage for frequency data, mean calculations, grand mean and standard deviations;
- reported the results of the data based on statistically significant information.

Data reporting was in the form of tabulations. However, the items in the given tables were either paraphrased or summarized in terms of their major concepts or key words because these items were longer enough to clarify the issue to the respondents(see Gray, 2004:380). It is also convenient for displaying means and standard deviations alongside. Thus, one can further read the full information depicted by the original items provided in the appendices section.

Data Presentations and Analyses

This chapter is aimed at presenting and analyzing the research data through tabulations and possible interpretations. When doing so, various sections were developed for those instruments used during data collections and also fixed alternative items that based on agreement scales, frequency scales and degree levels were assigned numerical values of the minimum 1 and maximum 5 in order to manipulate the mean values. The average of this for every item is expected to be 3 on point five likert scales. It means that if the mean value of the item is below this point 3, then the teachers beliefs about the issue raised in that item is weak, very weak or may not exist at all. The same is true for those items of practices and challenges. Besides, the open-ended items were mixed with the fixed alternative items during the analyses because such items were mainly used for technical procedures in the study rather than as independent questions. Researchers may check for irresponsible respondents as well as those who lack knowledge but still completing the items spontaneously.

4.1. Background information about EFL teachers

Teachers' profiles were explored in this section. In total, 34 EFL teachers of secondary schools teaching 9th graders were selected for the purpose of this study. The next table provides us with the teachers' sex, ages, academic status, teaching experience and their loads.

Table 4.1 Teachers' Profiles

Items	Profiles	Frequency(f)	%
1.Sex	M	25	73.5
	F	9	26.5
2.Age classes	21-30	6	17.5
	31-40	8	23.5
	41-50	11	32.5
	Above 50	9	26.5
3.Academic status in language studies	TTC diploma	0	0
	BA degree	15	44
	BE degree	17	50
	MA degree	2	6
	PhD	0	0



4.Experience in teaching EFL in high schools	< 1year	1	3
	1-2 years	4	12
	3-5 years	10	29.5
	6-10	5	15
	>11	14	41.2
5. Teaching loads in a week	5-10	3	9
	11-15	6	17.5
	16-20	20	59
	21-25	5	15
	Above 25	0	0

From this table we observe that 73.5% (25teachers) are males and the rest are females. Only 9(26.5%) of these are in their late adult hood. Among them no one has less than BA/BE degree and 85.7% have also more than two years of experience in teaching English in secondary schools. From these 14 teachers (41.2%; item-4) have more than 11 years' experience. In addition to other duties the teachers carryout in their schools and at home, most of them have teaching loads that range between 16 and 25 per week. In percentage this amounts 74% as it is seen from item number-5 (see also Table 4.2, item-3 and Table -4.5, item -43) what does this indicate? So, these information will be revisited in chapter-five.

4.2. Lead-in Items

Table 4.2

Items	scales	f	%
1. Do you teach speaking skills to grade nine students by using their textbook?	Yes	34	100
	No	0	0
2. Do you think the new textbook of English for grade-9 has sufficient tasks for oral skills(speaking)?	Yes	26	76.5
	No	8	23.5
3. Are there any challenges you face in assessing speaking skills of grade-9 students?	Yes	30	88.2
	No	4	12
4. Do you have your own view on how to assess speaking skills of learners?	Yes	32	94
	No	2	6
5. Do you know some alternatives for assessing speaking skills?	Yes	26	76.5
	No	8	23.5

All teachers in the selected schools are teaching speaking as a macro-skill of English language to grade nine students. Moreover, the teachers reported that there were challenges while assessing speaking skills of the learners and also they have their own views on how to assess oral abilities of learners and 76.5% of the EFL teachers know the alternatives in assessing speaking skills. The same percentage of the teachers under study believed that the students' English textbook has sufficient amount of speaking skill lessons in its contents which is about 60% as the researcher exploited during proposal preparation for this study.

4.3. Data on EFL Teachers' Beliefs about uses of alternative assessments

This data were gathered to know what teachers' beliefs look like about the use of alternative assessments in spoken language classrooms. It was rated by as "Strongly Agree" to "Strongly Disagree" and the teachers have responded to the items as follows.

Parameters: Strongly Agree(SA)-5 ;Agree(A)-4 ;Not Sure(NS)-3;Disagree(D)-2;

Strongly Disagree(SD)-1

Table 4.3 EFL Teachers' Beliefs N= 34

Items Alternative assessment		Scale					Mean	Standard deviation
		SA 5	A 4	NS 3	D 2	SD 1		
6. On maximizing learners' autonomy	f	16	16	2	0	0	4.41	.10
	%	47.05	47.05	5.88	0	0		
7. On developing speaking skills	f	11	20	3	0	0	4.2	.10
	%	32.35	58.82	8.82				
8. On increasing intrinsic motivation	f	17	17	0	0	0	4.5	.08
	%	50	50	0	0	0		
9. On backing up styles	f	10	21	3	0	0	4.21	.10
	%	29.4	61.7	8.8	0	0		
10. Observation as AA	f	9	19	5	1	0	4.10	.12
	%	26.5	55.8	14.7	2.9	0		
11. Observation for feedback	f	12	17	3	2	0	4.15	.13
	%	35.3	50	8.8	5.9	0		
12. On trustworthy of Direct oral test	f	10	18	6	0	0	4.12	.11
	%	29.4	52.94	17.6	0	0		

Almost half and more than half of teachers believed that using alternative assessments in oral classrooms (speaking) maximizes learners' autonomy, promotes learner's speaking skills, increases motivation, maintain learner styles and provide valid information about the oral language abilities of learners. For instance item-6 concerning the role of learners autonomy indicated 16 (47.05%) agreeing and 16(47.05%) strongly agreeing on the concept similarly, 20 (58.82%, item-7) believed that using alternative assessments enhance speaking skills of learners regardless of the types of language they learn. None the less, no one either disagreed or strongly disagreed on the direct oral testing as an alternative. The face value of mean on item 6 to item 12 ranged between agree to strongly agree (between 4.12 and 4.50). Furthermore, the data indicated

that it is possible to use observation method and assess learners' performance in the group(see the mean value of item-10 and 11) hence no one has shown doubt about this for this can also be done in association with effective questioning during teaching.

Contd= continued from previous table

Teachers' Beliefs.... contd N= 34

Item		Scale					Mean	Standard deviation
		SA	A	NS	D	SD		
Alternative Assessment:								
13. For oral task selection opportunity	f	13	16	4	1	0	4.21	.13
	%	38.2	47	11.7	2.9	0		
14. Oral test in reducing speech apprehension	f	3	14	8	7	2	3.26	.18
	%	8.8	41.2	23.5	20.6	5.9		
15. On minimizing complaints	f	4	21	6	2	1	3.74	.14
	%	11.7	61.7	17.6	5.9	2.9		
16. On practicability of AAs	f	15	15	3	1	0	4.6	.14
	%	44	44	8.8	2.9	0		
17. On communicating criteria and rubrics	f	8	21	3	2	0	4.03	.13
	%	23.5	61.7	8.8	5.9	0		
18. On providing frequent feedback	f	14	17	2	1	0	4.30	.12
	%	41.2	50	5.9	2.9	0		
19. Effective questioning as an alternative	f	23	11	2	2	0	4.67	.15
	%	67.6	32.4	5.9	5.9	0		

Equally important, the teachers believed that the use of oral production tests provide them opportunity to choose task materials from many sources. About 61.7% also indicated that using alternative assessment results as parts of summative evaluation minimizes complaints even though 6 teachers (17.6%) were not certain about this. Yet again item 20(> 50%) confirms about the beliefs of these teachers on the same issues. Moreover, half of the EFL teachers strongly agreed that the alternatives listed in question 16 can be used in assessing speaking skills of learners. About 88%, in aggregate of strongly agree and agree, indicated their view on the same item. These alternatives, as shown in the report of item18, enable teachers to provide constructive feedback to students more frequently than traditional tests. Teachers' responses to item 19 also exposed teachers' beliefs on using questioning effectively in lesson classrooms

enable to assess oral abilities of learners. This was the 2nd strong belief (mean 4.67, 67.6% agree) followed by that of item 23 (mean 4.62, 67.6% strongly agree).

Teachers' Beliefs.... contd N= 34

Item		Scale					Mean	Standard Deviation
		SA	A	NS	D	SD		
Alternative Assessment:		5	4	3	2	1		
20. On creating healthy relationship	f	13	17	2	2	0	4.21	.13
	%	38.2	50	5.9	5.9	0		
21.The effectiveness of written tests on speaking	f	1	6	5	10	12	3.76	.33
	%	2.9	17.6	14.7	29.4	35.3		
22. On fitting to levels	f	7	25	2	0	0	4.15	.08
	%	20.5	73.5	5.8	0	0		
23. On enhancing cooperative learning	f	23	9	2	0	0	4.62	.10
	%	67.6	26.4	5.8	0	0		
24.On possibility of using common criteria	f	5	24	5	0	0	4.00	.09
	%	14.7	70.5	14.7	0	0		
25.influence of written tests on AA practices	f	8	16	4	5	1	3.74	.18
	%	23.5	47	11.7	14.7	2.9		
Grand Mean of the Beliefs							4.14	

As it was discriminated by item 21, more than half of (22) the teachers do not believe that written tests (traditionally prepared) assess spoken language abilities effectively. It seems that the rest of the teachers carelessly responded to this item or they might have considered its practices casually. The teachers also believed that it is possible to bring down oral language assessment criteria to the level of their students (item 22, 94% on the Agree and Strongly Agree parameters).

To sum up, the items on teachers' beliefs yielded a mean value of 4.14. This indicates that the EFL teachers' beliefs of using alternative assessment in developing oral skills ranged between "I agree" and "I strongly agree". However the teachers also forwarded their feelings that students are accustomed to paper and pencil tests in their schools (see item-25).

4.4. Data on EFL Teachers' practices of Alternative Assessments

The following table demonstrates what the teachers' practice of alternative assessments look like in oral classrooms. Item-26 is discussed by intermingling here as it was stated at the beginning of this chapter.

Table 4.4 Teachers' practices of Alternative Assessments

N=34: Never (NV), Rarely(RR), Sometimes(ST),Frequently(FR),Always(AL)

Items on		Scale					Mean	Standard deviation
		NV 1	RR 2	ST 3	FR 4	AL 5		
27. Observational assessment	f	10	18	4	2	0	1.94	.13
	%	29.4	52.9	11.7	5.8	0		
28. Effective oral questioning	f	0	4	23	7	0	3.08	.09
	%	0	11.7	67.6	20.5	0		
29. Accrediting oral questioning	f	4	18	5	4	3	2.52	.19
	%	11.7	52.9	14.7	11.7	8.8		
30. Using AA results (100%)	f	11	15	5	3	0	2.00	.15
	%	32.3	44.1	14.7	8.8	0		
31. Oral productions portfolio	f	26	3	4	1	0	1.41	.13
	%	76.4	8.8	11.7	2.9	0		
32. Self-assessment	f	25	4	3	2	0	1.47	.15
	%	73.5	11.7	8.8	5.8	0		
33. Peer assessments	f	23	4	7	0	0	1.52	.13
	%	67.6	11.7	20.5	0	0		

This Table displays important information about teachers' practices of various alternatives in assessing their students' oral skills. For the sake of orderly discussion, item 26 was designed as multiple-choice where the teachers can select more than one alternative against the given letters. On this item almost all teachers selected written tests. As to information from item 27 (mean=1.94), teachers were not recording their students' oral ability through formal observation. To the contrary, about 67.6% (23 teachers) of them reported that they were using classroom questioning strategies in order to assess the learners' oral abilities. However, the mean value of item 28 indicated that they were practicing this sometimes. As the data obtained from item-31 shows, 26 teachers that accounted 76.4% never let their students to compile portfolios of oral productions.

Learners' self- assessment and peer-group assessment indicated by items 32 and 33 respectively were almost ignored by the teachers in the selected schools though these have got high theoretical consideration in many literatures. On the former item, 25(73.5%) indicated that they never practiced self-assessment as an alternative, and on the later item, 23(67.6%) teachers have informed that they never allow peer-group assessment in oral interaction classes.

Teachers' practicescontnd

N=34

Items		Scale					Mean	Standard deviation
		NV 1	RR 2	ST 3	FR 4	AL 5		
AA Practices :								
34. Using rubrics for speaking	f	22	6	6	0	0	1.52	.13
	%	64.7	17.6	17.6	0	0		
35. Setting common criteria	f	19	9	5	1	0	1.64	.14
	%	55.8	26.4	14.7	2.9	0		
36. Using written tests	f	2	3	2	21	6	3.76	.17
	%	5.8	8.8	5.8	61.7	17.6		
37. Using AAs for motivations	f	1	20	11	1	1	2.44	.12
	%	2.9	58.8	32.3	2.9	2.9		
38. facing challenges in Oral testing	f	1	5	4	10	14	3.91	.20
	%	2.9	14.7	11.7	29.4	41.1		
39. need for support on Oral testing	f	5	5	16	6	2	2.85	.18
	%	14.7	14.7	47	17.6	5.8		
40. supporting with authentic tasks	f	16	9	6	1	2	1.94	.19
	%	47	26.4	17.6	2.9	5.8		
41. Document setting for AAs	f	5	18	5	4	2	2.41	.18
	%	14.7	52.9	14.7	11.7	5.8		
Grand Mean of the practices							2.29	

In the second chapter, a lot was discussed about the importance of direct oral test as one type of alternative assessment (see section 2.1.2.1.4, PP: 30-40). However, responses to item 34 comprised that EFL teachers (28 among 34 of them) never or rarely practiced this alternative. Of course the teachers forwarded that they were always or frequently challenged in testing oral production of their students(see item 38).

As a whole the grand mean (2.29) obtained from items on the teachers' practices of alternative assessment methods that were responded by EFL teachers in the seven high schools portrayed that the teachers rarely practiced the proposed methods.

4.5. Data on the main challenges teachers face

One of the main objectives of this study was to identify if there were the most serious challenges that affected teachers' applications of alternative assessments in speaking classrooms. Therefore the following table is used with scale of the degree to which teachers' practices were affected.

Table 4.5 Assumed challenges the Teachers faced N=34

Items		Scale					Mean
		VH 5	H 4	M 3	L 2	VL 1	
42. The large number of students in a class	f	16	15	3	0	0	4.38
	%	47	44.1	8.8	0	0	
43. Maximum teaching load	f	9	10	10	3	2	3.61
	%	26.4	29.4	29.4	8.8	5.8	
44. Logistic materials for assessing speaking skills	f	12	4	8	6	4	3.41
	%	35.2	11.7	23.5	17.6	11.7	
45. Your knowledge of subject matter	f	4	6	3	6	15	2.35
	%	11.7	17.6	8.8	17.6	44.1	
46. Agreement on criteria and rubrics for assessing speaking	f	4	9	10	8	3	3.08
	%	11.7	26.4	29.4	23.5	8.8	
47. Learners' background on English language abilities	f	16	3	3	5	7	3.47
	%	47	8.8	8.8	14.7	20.5	
48. Compulsory mid-tests and final examinations	f	8	10	10	1	5	3.44
	%	23.5	29.4	29.4	2.9	14.7	
Grand Mean of the challenges =							3.39

Notes: Very high(VH) High(H) Medium(M) Low(L) Very low(VL)

Within the above table about seven challenges were assumed as the main factors impeding the assessment of speaking skills in EFL classrooms (see section 2.1.2.1.4.1: pp.36). The teachers were provided these items to see if there were any challenges they faced. From these most teachers replied that they faced three most serious challenges in implementing alternative assessments. The first serious challenge was large number of students in a classroom they teach English (indicated by item 42, mean 4.38, very high = 47% and high=44.1%) None of the teachers considered the large classes among the others as a little affecting factor. The second most affecting factor that was explored was their teaching loads (periods). Item 43 informed that only five teachers responded from low to very low on the likert scale whereas the vast majority of respondents replied from medium to very high. The third most challenging issue to implement oral assessment, as replied by the teachers, was the learners' background on English language abilities. This might mean that their students have not passed under effective assessment on this subject in their earlier grades. The data we have seen so far were from teachers' responses. The next section also provides us information gathered from their students through questionnaire.

4.6. Data Collected from grade-9 Students about their teachers' practices of AA

This data were gathered with questionnaire that prepared for students in their native languages(Afan Oromo and Amharic languages) and then transferred to its English version for the sake of data analysis. The respondents were 202 and 92 have preferred the Amharic version whereas 110 preferred the Afan Oromo version. The rating scale was from "Never to Always". As it was conveyed in the sampling method section, these students were mainly selected for triangulation purpose; just like that of the other tools to be used in research. So, the information from these students is treated in the next table.

Table 4.6 Learners' Responses to Teachers' Practices

N=202: Never (NV), Rarely (RR), Sometimes (ST), Frequently (FR), Always (AL)

Items		Scale					Mean	Standard deviation
		NV 1	RR 2	ST 3	FR 4	AL 5		
The English teacher:								
1. Communicates criteria and rubrics to students	f	110	35	45	7	5	1.82	.07
	%	54.4	17.3	22.2	3.4	2.4		
2. lets students to produce oral portfolios	f	124	27	38	7	6	1.73	.07
	%	61.3	13.3	18.8	3.4	2.9		
3. observes and records oral performance	f	76	62	32	18	14	2.17	.08
	%	37.6	30.6	15.8	8.9	6.9		
4. Tells results of observation	f	120	19	29	21	13	1.95	.09
	%	59.4	9.4	14.3	10.3	6.4		
5. uses direct oral test	f	109	27	34	15	17	2.00	.08
	%	53.9	13.3	16.8	7.4	8.4		
6. uses written test for assessing speaking	f	12	14	18	53	105	4.10	.08
	%	5.9	6.9	8.9	26.3	51.9		
7. renders self-assessment	f	116	24	39	11	12	1.91	.08
	%	57.4	11.8	19.3	5.4	5.9		

Students indicated that their teachers almost do not communicate about the criteria and rubrics for oral assessment in classrooms (see mean value of item-1 above). This information strengthens what were raised by item 34 and 35 in table 4.3. On the other hand, it contradicts the teachers' beliefs explored with item 17 "It is necessary to communicate oral assessment criteria and rubrics to our students"(in table 4.3: p.53). It was also explored that teachers never let their students produce oral portfolio as we see from the second question rendered to students. 124 (61.3%) students responded "Never" to this item.

Still more than half of students responded that their teachers do not test on oral production skills. As an evidence, we can read the mean value of item 5 (M=2.00) compared to the mean value of item 6 (M=4.10) these items were one among the technical questions the researcher used to identify quality respondents to validate his study more. Still if we consider the frequency of the respondents on both items, it can be interpreted that teachers are using written tests more frequently than direct oral testing. However, item 21 in Table 4.3 informed that the teachers most probably were in strong disagreement with assessing learners' speaking skills with paper-and-pencil tests (see mean value=3.76).

Items		Scale					Mean	Standard deviation
		NV 1	RR 2	ST 3	FR 4	AL 5		
The teacher:								
8. asks effective questions, solicits to assess ability	f	56	43	79	13	14	2.48	.08
	%	27.7	21.2	39.1	6.4	6.9		
9. allows peer-group assessment	f	112	31	28	14	17	1.97	.09
	%	55.4	15.3	13.8	6.9	8.4		
10. uses authentic tasks	f	104	28	34	16	20	2.11	.09
	%	51.4	13.8	16.8	7.9	9.9		
11. gives feedback based on alternative assessments	f	123	23	22	21	15	1.95	.09
	%	60.8	11.3	10.8	10.3	7.4		
Grand Mean of students' report on teachers' practices of AAs							2.20	

Furthermore, the learners reported that their teachers rarely or never provide them the opportunity to assess their own oral abilities or some body's else in their classrooms (Item 7, mean= 1.91, item 9=1.97) .The frequencies and percentages on these items almost speculated these alternatives as never practiced by the EFL teachers. If you see the response rates to item number-11, you come up with that teachers never provided feedback to their students that clearly based on the varieties of oral assessment. Huge number of students, in responding to question 14, preferred to be assessed their speaking abilities with direct oral production tests and followed by other types rather than paper-and- pencil tests. With this item about eleven types of alternative assessments were provided to the students to choose, at least three that they best like to be assessed with, and list in order of their preferences. In sum, the selected students have indicated that their English teachers were not fairly implementing alternative assessments to enhance their speaking skills(grand mean=2.2).

Teachers' document (mark lists and jots) didn't show clear recording of assessment on oral skills. In most of the selected schools for this study, there was high tendency of using written tests to evaluate learners' speaking skills. As some teachers informed me in informal chat, schools are using frequent tests written on black board as continuous assessment and these almost range between 40% and 60% in all schools. Only one school, in Serbo, is rendering oral production test once to their students but with no clear criteria of oral tests. In general, analysis from teachers' assessment document was perceived as there were no pre-set places for oral language assessment. Even those teachers who informed that they were using rubrics for oral tests, this was not observed in their documents and notes.

CHAPTER - FIVE

Summary, findings, Conclusions and recommendations

This chapter presents a summary of the study and a summary of what have been done and major findings obtained from those instruments used in data collection. The chapter also provides conclusions, recommendations and future studies to be conducted by others.

5.1. The summary

The main objectives of the study were to explore EFL teachers' beliefs and practices of alternative assessments in developing their students oral (speaking) skills and to identify what serious challenges; if any, the teachers faced in implementing alternative assessments in oral classrooms.

Literatures were also exploited and depicted under chapter two about the possible alternatives in assessing oral skills in EFL classrooms. The main ones were portfolios of oral production, self and peer assessments, observation of performance, effective questioning during lessons, direct oral test and so on. These main categories were discussed in detail preceded by rationales for alternative assessments and speaking skills.

In order to explore teachers' beliefs about these alternatives and their actual practices, questionnaires were developed for both teachers and students and administered to them. In sum, 61 items including lead-in, document analysis checklist and background exploring items were constructed and clustered into parts to forward the study. Then frequencies, percentages and mean values were manipulated on various likert scales to obtain the major findings.

5.2. Major findings

- EFL teachers have fairly strong beliefs about the concepts and uses of alternative assessments in oral skills. This was approved by those items prepared on teachers' beliefs particularly those constructed on agreement likert scales (grand means of items 6 up to 25=4.14, table 4.3).
- EFL teachers in the sampled seven schools hardly practice alternative assessments in their schools though they have strong beliefs. The grand mean of the items for teachers on this issue yielded 2.29 (Table 4.4) whereas the grand mean of the data from students read 2.2 (Table 4.6). The summation of the two on average provided a grand mean of **2.24**, which means on the frequency scale given from "Never" to "always" indicated "rarely". Though the means indicated this information here, the inter data interpretations as have been discussed in those previous tables displayed between "never" and "rarely". (see percentages and cumulative frequencies on some items). It was also confirmed from teachers' assessment documents.
- Unfortunately teachers had also reported that it was difficult for them to implement the alternative assessments because of the challenges they encountered from large class size, teaching loads and learners back ground on the English subject respectively. The means of these challenges were rated high to very high on these three factors among others (see Table 4.5).
- EFL teachers were mainly using written tests for assessing oral performance of learners but 95% of learners confirmed that they would like their teachers to use the available alternatives for speaking skills development. We can infer this from some likert scale and multiple choice items provided to teachers and students; for example, the means of items 21, 25 and 36 that technically used for teachers and that of students (item-6). Furthermore, multiple choice items like on number twenty-six and on number twelve for the teachers and learners respectively indicated the high utilization of test based (traditional) assessment. On item number fourteen 95% of students have chosen other alternative assessment methods like direct test, self and peer evaluations, portfolios etc. whereas only about 5% selected other plus written tests for their oral skills assessments.

5.3. Conclusions

From the above findings it is reasonable to conclude about the EFL teachers under this study as follows:

1. Teachers in the aforementioned schools had strong beliefs on the multiple advantages of alternative assessment approaches to oral language assessment.
2. Even though these teachers have strong beliefs, they hardly use the existing possible alternatives to assess oral abilities of their students.
3. The English language teachers were also highly affected by learners' poor background on English subject, teaching loads they have the number of students in a class respectively.
4. The teachers were most frequently using written tests for assessing language abilities.
5. Those secondary school students who participated on the study showed that test-based assessment is not their preference for oral skills development.

5.4. Recommendations

Based on the above findings, the following suggestions and recommendations were made:

1. Even though teachers have strong beliefs on using alternative assessments in developing oral skills of learners, they hardly implemented the alternatives. Thus, the teachers should try to practice alternative assessment within the existing challenges. For instance direct oral production tests can be handled with critically prepared checklists, both holistic and analytic types of rubrics and by cooperating with staff members.
The teachers shall consult web sources and published books on how to use rubrics for oral tests and the like.
2. It's also stated in the literature that students pass from one grade level to the next after fulfilling the reasonable development of the required skills or knowledge. For this, it is better for primary schools teachers to hand over students to their clients after effective assessments so that it wouldn't be a great challenge to secondary school teachers. The concerned body have to effectivize this there.
3. None the less, it is possible to increase the number of EFL teachers in schools by maximizing the intakes of EFL trainees at higher institutions. This should also be done by concerning officials and institutions in the country. This is because the teachers presented the teaching loads as the second leading challenges among others to use alternative assessment for the betterment of learners' spoken language proficiency. Surprisingly more than 95% of the students preferred other alternative forms of assessment than written tests for English language particularly speaking skills. Thus, it is better to think of this problem ahead. English as an international language may serve for different purposes: diplomacy, academic, businesses and developments.
4. Number of students in EFL classrooms should be suited to the teachers as much as possible. Because English as a subject demands practical actions than the knowledge of theories and concepts. In fact, leaners may require some grammatical formations but the new book for 9th graders in the country has already been designed on communicative basis. At least 60% of the contents in the book were identified as lessons that are based on speaking skills by the researcher during the proposal development for this study.

5. Future research should investigate more on learners' beliefs and assessment preference on oral skills development because this study didn't sufficiently treated, or targeted on these areas from available literatures. Furthermore, the researcher studied only some selected EFL teachers in a particular zone, and this can be disseminated to other schools with ethical considerations about this study.

Bibliography

- Abedi, J. (2010). *Performance assessments for English language learners*,
Stanford, CA: Stanford University-SCOPE
- Aijmer, K.(1996). *Conversational routines in English*, London: Longman
- Alderson, Clapham & Wall.(1995).*Language test construction and evaluation*.
Cambridge: C.U.P
- Anderson, L.(2003).*Classroom assessment: enhancing the quality of teacher
decision making*. London: LEA
- Animaw, A.(2011). Oral corrective feedback: an exploratory case study of the interplay
between teachers' beliefs, classroom practices, and rationales(PhD thesis)
- Bachman & Palmer.(1996). *Language testing in practice*. Oxford: O.U.P
- Baker, D.1989. *Language testing: a critical survey and practice guide*. London: British
Council library
- Bailey, K.(1998).*Learning about Language Assessment*. Boston: Heinle publisher
- Bailey, K.(1999). Washback in language testing. <http://www.toefl.org>
- Baldwin, Ch. A.(2006). Development of rating scale for a spoken English end course test
for Italian middle school students (MA thesis). <http://www.chrisb-baldwin.com>.
- Barker & Westrup.(2003).*Essential speaking skills*. London: VSO Continuum

- Baron, H & Boschee, D.(1995).*Authentic assessment: the key to unlocking student success*.
U.S.A:Techomic company
- Black, P.(1998).*Testing: friend or foe? Theory of assessment and testing*
- Bloxham, S.& Boyd, P. (2007). *Developing effective assessment in higher education: a practical guide*. Open University Press
- Brown, H.D.(2001). *Teaching by principles: an interactive approach to language pedagogy*
(2nded.). New York: Addison Wesley Longman, inc.
- Brown, H. D.(2004). *Language assessment principles and practices* . New York:
Pearson education
- Brown, G.(1996).*Testing in language programs*. Upper Saddle River: Prentice Hall
- Brown, J. & Hudson, T (1998a).The Alternatives in language assessment. University of
Hawai: *TESOL quarterly* 32, 4: 653-675. <http://www.JSTOR.org/stable>
- Brown, G & Yule, G.(1983).*Teaching the spoken Language*. Cambridge:
Cambridge University Press.
- Byrne, D.(1987).*Teaching oral English*. London: Longman
- Chang, Ch.(2005). Teachers' beliefs towards oral language assessment in Taiwan Collegiate
EFL classrooms(PhD. abstract). fllcccu.edu.tw/conference

Claxton, N.(2008).*Using Deliberative Techniques in English as a Foreign Language Classroom : manual for teachers*. New York: IDEBATE Press.

Congelosi, J.(2000).*Assessing strategies for monitoring student learning*. Utah State University: Longman

Council of European Union (2010).CBA of foreign language speaking skills

Cutting, J.(2002). *Pragmatics and discourse: a resource book for students*, London: Routledge

Doff, A.(1988).*Teach English: a training course for teachers*. Britain: C.U.P

Dornyei, Z.(2007).*Research methods in applied linguistics: quantitative, qualitative and mixed methodologies* ,Oxford: O.U.P

Edge, J. (1993).*Essentials of English language teaching*, London: Longman

Fulcher, G & Davidson, F.(2007). *Language Testing and Assessment: an advanced resource book* . Routledge: Taylor & Francis group.

Gall, M.D, Borg, W & Gall, J.P.(1996). *Educational research: an introduction*, 6th ed. USA: Longman

Genesee, F.(1994). *Educating second language children, the whole child, the whole curriculum, the whole community*. USA: Cambridge University Press.

Genesee, F & Upshur, G.(1996).*Classroom-based Evaluation in Second Language Education* , New York: Cambridge University Press.

Graves, K.(2000). *Designing language courses: a guide for teachers*, Boston: Heinell

- Gray, D.(2004).*Doing research in the real world*. New Delhi: Sage Publication
- Hamayan, E.(1995). *Approaches to alternative assessment*, Cambridge: C.U.P
- Harmer, J.(1991).*The practice of English language teaching*. London: Longman
- Heaton, J.(1988).*Writing English language tests*. London: Longman
- Heaton, J.(1990).*Classroom testing*. London: Longman
- Hughes, A.(2003).*Testing for Language Teachers*,2nd ed. Cambridge: C.U.P.
- Huerta-Macias, A.(1995).Alternative assessment: Responses to commonly asked questions
TESOL Journal, 5, 8-11
- Irons, A.(2008).*Enhancing Learning through Formative Assessment and Feedback*.
Routledge: Taylor& Francis group.
- Itkonen, T.(2010). Spoken language proficiency assessment: assessing speaking or
evaluating acting? (PhD. thesis) <http://www.doria.fi/bitstream>
- Joughin (2003).Oral assessment from learner's perspective,
(PhD. thesis) <http://www120.secu>.
- Kifle, K.(1995).An assessment of the content validity of high school English language tests
in relation to textbook concerned Addis Ababa University(*Unpublished MA*)
- Kissock, C. & Iyortsuun, P.(1982). *A guide to questioning: classroom procedures for*

teacher. London: Macmillan.

Knight, B.(1992).Assessing speaking skills: workshop for teachers.

Koul, L.(1988).*Methodology of educational research*.2nd ed. New Delhi:Vikas publishing
pvt LTD

Kubiszyn, T& Borich, G(2003).*Educational Testing and Measurement: classroom
and practice*. USA: John Wiley & Sons, inc.

Law, B. & Eckes, M.(2007). *Assessment and ESL: An alternative approach*. Canada:
Manitoba. books.google.com

Luoma, S.(2001). What does your test measure? Construct definition in language
Test development and validation (PhD thesis).<http://www.sdkj.jyu.fi/vahat>

Luoma, S.(2004).*Assessing Speaking*. Cambridge: C.U.P

Madsen.(1983).*Techniques in testing*. Oxford: O.U.P

Marshall, B.(2011).*Testing English: Formative and Summative Approaches to
English Assessment*. Continuum: India.

McCarthy, M.(1991).*Discourse analysis for language teachers*. Cambridge: C.U.P.

Mcloughlin & Lewis.(1994).*Assessing special students*, 4thed. USA: Prentice Hall, inc.

Mebea, F.(2008).The challenges of implementing continuous oral assessment in EFL classes

with reference to Debra-brihan teachers' college (MA)

- Moore, K.(2001). *Classroom teaching skills*
- Muller, M(n. d).Alternatives for English Language Assessment in South Korea
Hankuk University of Foreign Studies
- Norrish.(1983).*Language learners and their errors*. Macmillan
- Nunan, D.(2004). *Task-based language teaching*. Cambridge: C.U.P.
- Oller.(1979). *Language testing in schools*. Oxford: O.U.P
- Overton, T.(2000).*Assessment in special education: an applied approach*. Upper
Saddle River: Prentice Hall
- Richards, J.C(2008).Teaching listening and speaking: from theory to practice.
USA: Cambridge University Press.
- Richards, J.C,& Lockhart, Ch.(1994). *Reflective teaching in second language
classrooms*. New York: Cambridge University Press.
- Riihimaki, J.(2009). Assessment of oral skills in upper secondary schools in Finland:
Teachers' view.(BA thesis):University of Jyvaskyla.
- Seifu, B.(1997).Communicative Language Testing: Its Validity and Reliability at the
Addis Ababa University(MA)
- Simachew, G.(2012).Washback of the University Entrance English Exam (UEEE) on
Teachers' and Students' Practices: The Case of Preparatory Schools in
Amhara National Regional State (PhD. Thesis).
- Soler & Martinez.(2008). *Investigating pragmatics in foreign language learning, teaching
and testing*.UK, Bristol: Multilingual Matters
- Tadesse, G.(1990). An assessment of English oral performance of Diploma students of
Bahir-Dar Teachers' College(MA)

- Tesfaye, G.(1982).The measurement of English language proficiency of high school graduates (MA thesis).
- Teshome, D.(1995).The construction and validation of test in English for tertiary education with reference to Addis Ababa University(PhD.)
- Tsagari, C(2004). Alternative Assessment: some considerations. *Language Testing Updates*, 36:116-124.
- Underhill, N.(1987).*Testing Spoken Language*. New York: C.U.P
- Ur, P. (1991).*A course in language teaching: theory and practice*. C.U.P
- Wagari, N.(1995). The impact of multiple choice and short answer on students' performances : three government senior high schools in Addis Ababa in focus, MA.
- Watt, H. (n.d). *Attitudes to the Use of Alternative Assessment Methods in Mathematics: A Study with Secondary Mathematics Teachers in Sydney*. Australia: Springer. <http://www.jstor.org/stable/25047136>
- Weir, C.(2005).*Language testing and validation: an evidence-based approach*. Palgrave: Macmillan
- Zerihun, S.(2009).Teaching speaking skills in secondary schools: attitudes, perceptions and practices(MA).

Appendices

Jimma University
College of Social Sciences and Law
School of Graduate studies
Department of English

Appendix-I: Questionnaires for Secondary School EFL Teachers

Dear teachers,

This is a questionnaire for academic study about beliefs and practices on assessing speaking skills of learners in secondary schools. Your answers to the items in this questionnaire sheet are very valuable and important to the study. Being so, only your honest responses to the items will be valuable for there are other participants who will respond to similar questionnaires to triangulate the relevance. Nevertheless, feel confident to respond because the researcher will use code system to keep respondents' secret.

Part-I: Background Information

Please mark your answer in the appropriate box

1. Gender: male female

2. Age: 21-30 31-40 41-50 Above 50

3. Your academic Status in language studies

Diploma(of TTC) BA Degree BE Degree MA Degree PhD

4. Experience in teaching English in Secondary Schools:

Less than 1 year 1-2 years 3-5 years 6-10 years More than 11 years

5. Your Teaching Loads(periods) in a week:

5-10 11-15 16-20 21-25 Above 25

Part-II: Please answer the following questions by responding as “Yes/No”.

1. Do you teach speaking skills to grade nine students by using their textbook? _____
2. Do you think the new textbook of English for grade-9 has sufficient oral skills(speaking)? _____
3. Are there any challenges you face in assessing speaking skills of grade-9 students? _____
4. Do you have your own view on how to assess speaking skills of learners? _____
5. Do you know some alternatives for assessing speaking skills? _____

If yes, please list them

Part-III: Teachers’ beliefs

Please tick under one of “I **Strongly Agree**(SA), **Agree**(A),**Not sure**(NS),**Disagree**(D) or **Strongly Disagree**(SD)” to show your own beliefs on the following statements.

Items on teachers’ beliefs only	SA	A	NS	D	SD
6. Using alternative assessment maximizes learners’ autonomy.					
7. Using alternative assessment in oral classrooms enhances speaking abilities of the language learners regardless of the types of language					
8. Alternative assessment increases students intrinsic (inner) motivation					
9. Learner styles can be maintained through alternative assessment					
10. Observational assessment can be used to assess oral skills					
11. Assessment by observation needs both formal and informal records for Providing feedback to students at spot or after time					
12. Direct oral testing provides us with true information about speaking ability Of the learners more than written tests					
13. Oral production (speech) tests provide teachers wide opportunity to select Oral tasks from various sources					
14. Oral production tests can reduce learners’ speech apprehension gradually					
15. Using alternative assessment methods as a part of summative assessment can minimize complains from learners ,educators, parents and stakeholder					

16. Reflections, portfolios, journals, peer-assessments, self assessments, Project presentations, learning- logs and diaries and effective questioning can be used in assessing oral skills of language learners					
17. It is necessary to communicate oral assessment criteria and rubrics to our students					
18. Alternative assessments enable teachers to provide feedback to learners more frequently than written tests.					
19. Effective oral questioning in classroom can assess oral ability of learners and can be an alternative to use.					
20. It is possible to develop positive relationship among students and teachers through alternative assessment practices					
21. Traditional paper-and-pencil tests are more effective for assessing oral skills(speaking) of students					
22. Teachers can bring down assessment criteria to the level of their students while using alternative assessment tools					
23. Students can develop cooperativeness and learn better if teachers use alternative assessments					
24. In alternative assessment practices, it is possible to set common assessment criteria or rubrics with teaching staffs					
25. It is very difficult to practice alternative assessments in oral class in our school because the learners are accustomed to paper-and-pencil tests					

Part-IV: Teachers’ Practices

A. For question -26, please encircle the alternatives **you most frequently use** to assess speaking skills of learners in grade-9. More than one method can be chosen based on your experiences.

- 26. A. Written tests (quizzes, mid-exams and final examinations)
- B. Formal observation records
- C. Self-assessment
- D. Peer-assessment
- E. Group assessment
- F. Direct oral testing(interviews, role plays, storytelling, interpretations etc.)
- G. Portfolio works
- H. Effective questioning
- I. Conferences
- J. Demonstrations

K. Journals

L. Project presentations

M. Reflections on progresses

B. From No.27-41, use the rating scales below to show **how often you** are applying during assessing speaking skills of **grade-9 students**.

Never(NV) ,Rarely(RR), Sometimes(ST),Frequently(FR),Always(AL)

Items on practices	NV	RR	ST	FR	AL
27.I use formal observations with effective recording procedures during oral skills classrooms					
28. I use questioning strategies such as soliciting for the purpose of assessing oral ability of learners					
29. I give value to learners' reply to oral questions in classroom					
30. I use most of alternative assessment results for finalizing learners' assessment(i.e. as out of 100%)					
31. I let my students to develop portfolio of oral skills such as dialogue constructions, compiling phrases of speech registers , interactional stories and short plays etc.					
32. When I assess speaking skills, I give student the chance to assess him/herself with prepared checklist or questionnaires					
33. I give learners the opportunities to assess their classmates speaking abilities for summative evaluation					
34. I practice direct oral testing to my classes by preparing both holistic and analytic procedures					
35. My staff members and I set common rubrics and criteria of oral assessment through discussion.					
36. I use to give feedback to learners from test results more than the other forms(alternatives)					
37. I use alternative assessments to motivate learners towards improving their speaking					
38.When I test oral productions of students, I face challenges					
39. I need others' help to overcome the challenges of assessing speaking skills					
40. I supplement textbook tasks with my own choices to support all students					
41. I prepare clear document that displays places for alternative assessment to handle complaints as well.					

Part-V: The Challenges Teachers face

Please rate how much the following factors have affected your practices of alternative assessments in oral skills classes. Use a tick mark to show the degree of the challenges under the given scales.

Very high(VH) High(H) Medium(M) Low(L) Very low(VL)

Items	VH	H	M	L	VL
42. The large number of students in a class					
43. Maximum teaching load					
44. Logistic materials for assessing speaking skills					
45. Your knowledge of subject matter					
46. Agreement among professional staffs to set criteria and rubrics for assessing speaking					
47. Learners' background on English language abilities					
48. Compulsory mid-tests and final examinations to be used for assessment in the school					

49. Which of the challenges written above do you want to be **solved first**? Please write how you think it should be solved with a paragraph.

THANK YOU A LOT FOR YOUR HONEST RESPOSES!!!

Jimma University
College of Social Sciences and Law
School of Graduate studies
Department of English

APPENDIX-II: Questionnaire for students of secondary schools (English version)

Dear students,

This questionnaire is set to be answered by you only for academic study. Your answers to each question are very important for this study, so please answer all the questions confidently that the researcher will use coding system to keep it secret.

Part-II: Please rate the next questions by using the following scales and tick only once.

Never (NV) Rarely (RR) Sometimes (ST)
Frequently (FR) Always(AL)

items	NV	RR	ST	FR	AL
1.The English teacher tells us criteria and rubrics for assessment of speaking skills					
2. The English teacher lets us to produce oral skills portfolios					
3.The English observes us individually or in group works and Records our oral performance on documents					
4. The English teacher tells us the results of observational assessment to make us ready for next time					
5. The teacher uses direct oral test in learning classroom					
6. The teacher uses written test for assessing spoken language More than direct oral test					
7. The teacher gives me chance to assess my oral English for Scoring purpose					
8. The English teacher asks effective questions, solicit and catch up student's oral language ability					
9. The teacher provide opportunities to peer-group assessment for Their speaking skills abilities					
10. The teacher uses authentic tasks to simplify the oral tests or to Match the levels of the students in classroom					
11. The English teacher gives me feedback based on alternative assessments that used.					

12. Which of the following alternative assessments your English teacher is not commonly Using to assess your oral skills abilities? You can choose as many as you can from the lists.

- A. Written tests
- B. Observational assessment
- C. Self- assessment
- D. Peer-group assessment
- E. Portfolio assessment
- F. Oral production tests
- G. Effective classroom questioning
- H. Reflections
- I. Conferencing
- J. Journals
- K. Project presentation

13. How much your English teacher has used alternative assessments in oral classroom during the first semester compared to quizzes and examination scores that marked of 100%? Please choose only one for this question.

- A. Very few of the them
- B. Few of them
- C. A lot of them
- D. It is difficult to know for no much communication on them

14. From the lists under question number- 12, please write the three best methods you would

Your teacher to use when assessing your spoken language performance:

1st. _____

2nd. _____

3rd. _____

Jimma University
College of Social Sciences and Law
School of Graduate studies
Department of English

Appendix-III: Document Analysis

Date of document Assessment: _____

Code for the Teacher whose document was used: _____

“Yes, “No” and “Unclear” checklist is used to assess documents of the EFL teachers.

Items	Yes	No	Unclear
1. The teacher is confident/ volunteer to allow the observation of the document s/he uses for assessment.			
2. The teacher has places for recording portfolio works in the document			
3. The teacher has places for oral test recording in the document			
4. The teacher has places/ways to record assessment by observation			
5. The teacher has places for self-assessment of learners’ oral skills			
6. The teacher has places for peer-group assessment recording			
7. The teacher has clear checklists that planned for alternative assessment types as a document.			



This work is licensed under a
Creative Commons
Attribution – NonCommercial - NoDerivs 3.0 License.

To view a copy of the license please see:
<http://creativecommons.org/licenses/by-nc-nd/3.0/>

This is a download from the BLDS Digital Library on OpenDocs
<http://opendocs.ids.ac.uk/opendocs/>