ADDIS ABABA UNIVERSITY
SCHOOL OF GRADUATE STUDIES
INSTITUTE OF PSYCHOLOGY
(MEASUREMENT AND EVALUATION)

TEST CONSTRUCTION SKILL AND ASSESSMENT OF FACTORS AFFECTING TEST ANALYSIS AND EVALUATION: THE CASE OF THREE SELECTED HIGH SCHOOLS IN ADDIS ABABA

By
Meseret Kuma

June 2011
Addis Ababa

ADDIS ABABA UNIVERSITY
SCHOOL OF GRADUATE STUDIES
INSTITUTE OF PSYCHOLOGY
(MEASUREMENT AND EVALUATION)

TEST CONSTRUCTION SKILL AND ASSESSMENT OF FACTORS AFFECTING TEST ANALYSIS AND EVALUATION: THE CASE OF THREE SELECTED HIGH SCHOOLS IN ADDIS ABABA

By
Meseret Kuma

June 2011
Addis Ababa

TEST CONSTRUCTION SKILL AND ASSESSMENT OF FACTORS AFFECTING TEST ANALYSIS AND EVALUATION METHODS: THE CASE OF THREE SELECTED HIGH SCHOOLS IN ADDIS ABABA

BY

MESERET KUMA

A Thesis Submitted To The School Of Graduate Studies Of Addis Ababa University In Partial Fulfillment Of The Requirement For The Degree Of Master of Arts In Measurement And Evaluation

APPROVED BY THE BOARD OF EXAMINERS:

_____	_____
Chair person, Psychometric and Educational Psychology Program unit	**Signature**
_____	_____
Advisor	**Signature**
_____	_____
Examiner	**Signature**

Acknowledgements

First and for most, I would like to express my deepest heartfelt gratitude and appreciation to my advisor Ato Daniel Tefera for his critical, insightful and constructive throughout writing this research paper.

I am also greatly indebted to my friends who assisted me morally and materially throughout my study.

Finally, my deepest gratitude and thanks go to all members of my family for their supports and encouragement throughout my study.

Table of contents

List of Tables

Abstract

The purpose of this study was to determine to what extant teacher made tests were effective. The research also focused on identifying the factors that inhibit analysis and evaluation of test by high school teachers of Nifas silk sub city. Moreover, qualitative description of the test was made by analysis of the test papers the teachers set or prepared using the principle of test construction checklist. In addition, by using questionnaire, relevant data were collected from the teachers. Furthermore, interview was carried out to gather data from school principals about the over all condition of the schools and about the factors that limits the teachers from making post test analysis. The finding of this research indicated that teacher made test qualities were not good in terms of the difficulty level and discrimination power. The qualitative data analyses have also shown that most of the test items were not constructed as per the principle of test construction. The finding farther demonstrated that teachers were not in positions to make post test analysis. Finally, it is recommended that to minimize the problems associated with test construction; teachers should be encouraged to engage in test item analysis.

CHAPTER ONE

1. Introduction

1.1 Background to the Study

Tests have significant contributions in education system. Standardized, carefully designed tests and teacher -made tests have an impact on the teaching learning processes. Tests have been used increasingly in recent years to make educational assessment. Teacher -made tests are helpful for learners to acquire adequate knowledge and skills. Thus they are the cornerstones in enabling learners to follow up the teaching learning process. Educational tasks of aptitude and achievement tests greatly improve the precision and efficiency of the observations on which educational assessment rest (Beausay, 1972). The use of tests have been accompanied by a series of critical comments, because tests vary in quality none is perfect and some may be quite imperfect. However, various studies conducted at different levels reveal that there are problems in designing tests to evaluat students' knowledge, proficiency and competence (Beausay, 1972 and Brown, 1983).

Tests can be important part of the teaching learning process if they are integrated into daily classroom teaching and reconstructed to be part of the learning process. They allow learners to see their own progress and teachers to adjust their instruction on a daily basis (Brown, 1983).

Teacher-made test can be are written or oral assessment that are not commercially produced or standardized. In other words, they are tests a teacher design specifically for his or her students. They refer to any kind of activities that result in some type of mark or comment being entered in a checklist or grade book. Besides, it can be a more structured oral or written evaluation of students' achievement. Teacher-made tests uses a variety of formats, embracing multiple-choice items, matching items; true false items, fill in blanks items or essay items.

According to Brown (1981) there are three possible reasons why students do not perform adequately in tests. First of all, they have not learned the material properly. Second teachers used

an in appropriate method of measuring their knowledge. And thirdly students are not motivated to perform in the tests.

Brown (1981) furthermore noticed that from the above three reasons the second can be eliminated by careful construction of test item. Hence, well-defined item writing skills can and should play important role in students' learning.

Related to the above view, Ory and Ryan (1993) described that people spend as much time developing exam and grading papers or they prepare classroom presentation because there are multiple benefits to be gained by their efforts, for example, well written exams can be used by students to determine what content they have mastered and what knowledge they have failed to understand.

Brown (1981) stated that the best known and probably most widely used method of measuring achievement are paper and pencil test constructed by the classroom teachers. This may include comprehensive end of term examination, a unit test, or weekly quiz. Items may be presented in any of a variety of formats multiple choice, true false, short answer, completion and/or easy questions. The distinctive features of such exams are that they are constructed by the classroom teachers and measure objectives specific to a particular course.

Bertrand and Cebula (1980) explained that, in the hands of good teacher test become valuable tools for the diagnosis of individual strengths and weakness. They also added that test is merely a tool and like any tool, it must be used appropriately to serve its purpose appropriately. They also farther tell us appropriate test can provide not only about what children know or don't know but also about how they learn.

Bertrand and Cebula (1980) again tell us that most teachers have their favorite kinds of tests and use them frequently once they become proficient in constructing and administrating them. However, whether the favorite test format is true false, multiple choice or essay, they should have knowledge and skill constructing and properly using tests.

Developed counties mostly use standardized achievement tests to supplement teacher made tests and to maintain the quality of educational decisions made by educators and to minimize problems associated with teacher made tests (Hoge and Buthcer, 1984). Since as to the present researcher's

experience standardized tests are not used efficiently in the Ethiopian school systems, people are largely relying on teacher made test to make different educational decisions. Hence, teachers have to show a great concern and bother about the improvement of their tests.

It is a common source knowledge that even though teachers may not be blamed for the level of the knowledge acquired by their students; they may be blamed for not making proper evaluation and providing feedback; for not improving the quality of test items; and for not trying to improve the teaching methodology that they are employing based on prior test results.

1.2 Statement of the Problem

According to the MOE reports (cited in Hailu 1990) the evaluation made based on classroom tests depended 52.2% of the time on teachers' subjective judgment, while 47.8% of the time the decisions were reliable. This indicates that teacher made tests have extensive level of weakness. Constructing a good teacher- made test is very time consuming and difficult; moreover, it is hard to understand why something so essential to the learning process has been virtually ignored by teacher.

However, it could be argued that in the case of teacher made tests, it is the teacher himself/herself who should be responsible for improving his/her tests and items by making post test analysis. However, as observed by the researcher and as her experiences as a teacher tells most teachers do not seem to analyze their tests.

Theoretically, it is widely believed that teachers do not set test items without taking into account the improvement components and procedures of good test construction such as the objectives set to be achieved, the content area covered, the methodology employed, the interest and attitude of students toward the subject, the teacher-students interaction and the time spent on that particular content area. (Ory and Ryan , 1993; Bertrand and Cebula, 1980; Nunnally, 1964).

Adams (1964) stated that teacher developed a test to aid in ranking students with respect to their level of achievement in some area, the test should be based on a representative sampling of the content studied and the abilities or skills emphasized in the course and should contain a sufficient number of questions with a wide range of difficulty so that the test will have adequate reliability.

Adams (1964) also indicated that constructing teacher made test is becoming very critical for its effects on the instruction and school programs also it concerns both practitioners and researchers in education. Hence, to provide tests which enable us predict the achievement of students, empirical procedures have to be followed during the process of item construction and use.

However, as far as Ethiopian educational system is concerned, a study was not made in how well teachers had made use of tests and test results to improve the method of instruction that they employed. At the same time, it is not well known to what extent item analysis were used by teachers to improve the quality of their test item.

Besides this, even though teachers are expected to carry out statistical analysis of tests for the improvement of instruction and test quality, the researcher's personal experience help to have information that teachers are not acquainted themselves with such educational tasks. In light of this idea, the present study will assess the way how they construct their tests and examine factors that may affect teachers unable to make item analysis.

Moreover, making a survey of the barriers which deter teachers from analyzing their own exam after they have administered and used them is the main concern of this research.

The study was intended to answer the following basic research questions.

- Do teacher-made tests have the desired quality according to the rules of test construction?
- Do teachers have the required skills to carryout test item analysis and evaluation?
- Do teachers have enough time to make post test analysis and evaluation?
- Do teachers' experience, training condition on measurement and evaluation and work burden on them affect their tasks on their test construction and item analysis?

1.3 Significance of the Study

A research which deals with the issues of test constructions and test item analysis is expected to contribute some insight in the area. Therefore, this research would have some significance: In general, the results of this study

- Help school teachers to improve the quality of their tests.

- Inform school principals and other decision-making bodies in the education sector to pay particular attention to the problems of teacher-made tests.
- Motivate other researchers to carryout further research in relation to test construction.

1.4 Delimitation of the Study

The study is limited to three schools in Addis Ababa in Nifas Silk Sub City which are Ginbot 20, Abiyot Fana and Frehiwot as representative of secondary schools. Only examinations in two subjects were taken and assessed for their psychometric qualities.

1.5 Operational Definition of Terms

Test item: - is the smallest independent unit of test.

Item analysis: - is a set of procedures for determining item discrimination and difficulty.

Secondary School: - the school division which has grade 9 and above.

Teacher made test: - test constructed by classroom teachers and measures objectives specific to a particular course.

Test construction skill: -the teachers ability to follow all the necessary test construction procedures and test item analysis during their tests.

CHAPTER TWO

2. Review of Related Literature

2.1 The Importance and Usage of Tests

Berdie (1962) stated that to compare one person to another, or to compare the current status of an individual with his previous status, observation must be refined and quantified, that is the reason why teachers usually use tests. Test consists of situation which allows us to observe a carefully selected sample of well-defined behaviors (Thorndike and Hagen, 1965; Dull 1988).

Linn (1983) noted classroom tests have four features first the degree to match between test items and instructional objectives, second the use of test results to provide feedback to students and to the teachers', third the use of tests to flag facts or concepts that are considered important, and fourth to determine grades. According to him, tests help to flag facts or concepts that is, it refer to tests provide an important means of identifying those items that are judged to be particularly important and for providing feedback to teachers and students about the degree to which those items have been learned. Also Roid and Haladyna (1982) indicated that tests have been used to motivate students into studying, assign grades, evaluate instructional programs, evaluate teaching effectiveness, group students for instruction or select students for special programs or assignment. All these indicate that tests guide learners and yield information regarding the effectiveness of the instructional program. So, teachers have to take care about their tests quality.

The most important function of any achievement test is to provide feedback regarding the effectiveness of teaching (Brown, 1981). He also tells us that feedback serves various functions. First, it informs both the teacher and the student of the student's current level of performance in the course. Second, it provides diagnostic information that can be used in planning further, and/or

remedial instruction. Third, by considering the results of several tests, student's progress and improvement can be measured.

In this regard, teacher –made tests are important tools to know whether students have learned the material and developed the desired knowledge. One of the important tasks to teachers is to determine whether students have, in fact, learned the material and developed the desired knowledge (Brown, 1981). In general according to the above scholars the reason why we measure students' achievement are to provide feedback regarding to the effectiveness of learning, to motivate students, to give grade, to establish proficiency and to evaluate instruction. So teachers should have to take care of about their test when they prepare.

According to Ebel (1982) the construction of any test of educational achievement is confronted with two major problems. The first problem is what to measure and second major problem is how to measure. He also gives solution for both problems. For the first problem to focus primarily on testing for knowledge, testing for ability should come next. The solution proposed for the second problem is to present with a series of incomplete propositions accompanied by two or more alternatives only one of which makes the propositions true. In other words, these two problems are talking about content what teachers cover during test construction and the way how they construct the items?

2.2 The Role of Teachers in Improving Quality of their Tests

As the teacher has achieved a more significant role in the planning of the educational experience for his/her class, he/she has become responsible for apprising the worthwhile those experiences which students are achieving educational goals (Adams, 1964).

Brown (1981) furthermore says that in any course one important task for the teacher is to decide what material will be covered and what knowledge and skills students should develop as a function of the course. He also mentions that there is another important task that should be done by teachers during their teaching process. They have to determine how the material can be presented most effectively so that learning will be facilitated. Next they should determine whether students have in fact, learned the material and developed the desired skill. It is this last task of the instructional process, which is measuring the achievement of students effectively

which is the focuses of this research. When teachers measure achievements, the goal is to obtain information that will be useful in planning and evaluating the instruction.

Brown (1981) also reveals that item writing needs practice as well as feedback regarding the success of efforts. According to him the most important source for feedback are students from their performance on items. Another feedback can come from statistical analysis of items and by having colleagues review of items. It is clear that item writing is not an easy task. Teachers need to know the subject well, understand the characteristic and abilities of students, they have also to know what and why to measure and know how to write good items.

Bertrand and Cebula (1980) point out that for the teachers to know the result of a test well, tell whether or not children have learned as much and as well as intends instructional objectives, the temptation to teach for the test can be very strong. They also added that however, many teachers simply review the unit material and then rapidly assemble a number of test items. This usually indicates that they did not teach the unit according to carefully prepared instructional objectives. Such teaching might fail to provide both the teacher and the students with clear evidence of personal and intellectual development.

Ory and Ryan (1993) also reveal that teachers work with test construction should begin with a discussion of Bloom's taxonomy. They also added that this taxonomy encourage teachers to write course objectives at different learning levels and to analyze teaching process, the desired students over comes, and test items.

Ory and Ryan (1993) farther elaborate that testing plans need to be developed before writing the first test item. Also, teachers can begin to develop their testing plan by answering the following questions.

- What is the purpose of the exam?
- How difficult should they make the test?
- Who is taking the exam?
- How many students are taking the test?
- How much time has been provided for testing?
- How many exams are scheduled?
- What type of test (multiple choice, essay, true-false etc…) is desirable?

So these indicate that accumulated test items without a plan can be accompanied with unduly represent informational learning. Adams (1964) indicates that teachers are likely over emphasizing certain areas of content in which items are easily constructed. He also adds that in order to improve the test's representativeness, or its content validity, one should first develop a blue print for the test. He farther states that table of specification or blue print for the test, forces teachers to relate their testing procedures to the material they taught. It is obvious that teacher do not use such a crutch to prepare a test, many teachers simply review the unit material and then rapidly assembles a number of test items.

Ebel (as cited in Adams 1964) identified seven serious errors that teachers frequently make in their evaluation of pupils' education attainments:

1. Teachers tend to rely too much on their own subjective judgments.
2. Teachers tend to put off test preparation to the last minute.
3. Many teachers use tests which are too poorly planned, too short, or too inefficient.
4. Teachers often place too much emphasis on trivial, details to the neglect of basic principles, understanding and applications.
5. Teachers often write questions, both essay and objectives, whose effectiveness is lowered by ambiguity, or by irrelevant clues to the correct response.
6. Many teachers overlook, or underestimate, the magnitude of the sampling errors which affects test scores.
7. Most teachers fail to test the effectiveness of their tests by even a simple statistical analysis of the results from the test.

As pointed out by Teka (1988) classroom tests must be prepared in the light of certain general principles, and teachers must know these principles to avoid their mistakes. He also goes on by telling that teachers must be aware of those factors involved in qualitative test preparations. He also adds that teachers must avoid the use of hastily and haphazardly prepared classroom tests. So according to the writer we can say that teachers must learn and adapt some basic principles in preparing and using tests. In this regard Teka (1988) stated "The use of simple word, avoiding qualitative and absolute terms, and the use of correct grammar are some of the most important principles in preparing a classroom tests"(P. 19).

On the other hand, Adams (1964) stated that there is no better way to develop skill in testing than to analyze systematically the results of previous efforts. He also adds that on an achievement test, efficiency of measurement can be improved by discarding or revising items that do not correlate well with total score, or do not discriminate between high achieving and low achieving students. Such a study of the difficulty and discrimination value of items is called an item analysis.

In general as it is clearly stated by the above writers' teachers need to have skills on test construction as well as item analysis. Otherwise it is difficult to put them in line of a person who carries out his/her responsibility as expected.

2.3 Experiences in Teaching and Ability in Test Construction

Even though test construction is a time consuming pursuit like most other important tasks, teachers are expected to give a considerable degree of attention and to have patience to construct good tests. It is expected that the experience teachers possess in the teaching profession has its own influence on teachers' knowledge of their students test performance (Thorndike and Hagen, 1965).

Nunnally (1964) argued that the art of item writing must be acquired from classroom experience. He believed that most teachers learn to write better items after several years of practice. He further explained that teachers who save copies of tests can often look back to earlier years and see that many of their items were inappropriate, ambiguous, trivial, too difficult or too easy.

Brown (1981) also reveals that for item writing there is one essential requirement what teachers need to know is practice, like any other skill, item writing is perfected only with practice. He also added that the only way to learn how to use tests is to use them, just as the only way to learn how to teach is to teach.

2.4 The need of Statistical Analysis of Tests

There is a great belief among educators in that instruction and tests achievement of students can be improved highly if proper statistical analyses of tests are made by instructors (Adams, 1964; Thorndike, 1971; Brown, 1983 and Mehrens, 1984; Ashwoorth 1982).

This may indicate that the improvement of instructional process is directly linked with making statistical analysis of the item. In other words, it implies that the improvement of process of

instruction can be expected, only when we are in position to carry out the proper statistical analysis of tests (Mehrens, 1984).

Although item analysis and other forms of statistical analysis of the test are found to be important in improving the quality of the tests and suggested that revising items based on item analysis is more useful and efficient than writing new items, unfortunately teachers practice of making statistical analysis of tests were found to be low (Betera ,1990).

In view of this fact, Mehrens (1975) states that classroom teachers often by pass quality control checks on their tests and explained that the probable cause are either they do not know how to make quality control, or they know, but they think it will be too much time consuming.

Adams (1964) shows that teachers should write many more items than needed then they can revise or eliminate item. He also states that, it can also help them to study the difficulty of each item, so that they can select items which meet desired standards with respect to difficulty.

Every test items is in some way like a little test (Thorndike, 1977). Therefore, every item is subject to certain forms of analysis to determine its adequacy. This analysis helps to estimate each item's difficulty and discrimination. Both of these types of analysis required responses to items by students for which the items were intended.

The first question is difficulty level. According to Bean (1953, P.152) "an item is passed by 90 percent of the group or better, it is obviously an easy item which does not discriminate very well except, perhaps, at the very bottom of the scale of that trait or group of traits being measured."

In the analysis, the second problem that arises is the discrimination value of each item. Again Bean (1953, p.153) states the following "By the discrimination value is meant the degree to which the single item separates the superior from the inferior individuals in the trait or group of traits being measured."

Roid and Haladyna (1982) confirm that every item can be related to the total test scores. This correlation between items and test performance is the discrimination index, which can range from – 1.00 to + 1.00 ; any negative index indicates that students who get high scores tend to get the item wrong where as students with low test scores tend to get the item right. They also added that a discrimination index of zero would indicate no pattern. According to them at a minimum, an

item should differentiate between levels of achievement otherwise items need to be examined carefully.

Concerning to test item's difficulty, they also point out that it can be calculated by knowing the proportion of subjects in a representative sample that give correct response. The statistic that reflects this finding, frequently expressed in decimal form, is often called by them "the p-value" which is directly translatable in the probability of a correct response. For example from 100 students try on item and 75 students get the item correct, the p-value is .75 and the probability that any student will get the item correct is also .75 or 75%.

According to Bowen (1985) and Zurawski (2001) an item is said to be easy if its difficulty index is above 90 %, and too difficult if its difficulty index is below 30%. Regarding to discrimination indexes they also mentioned that below 15% the item is not acceptable.

Any item that adequately represented a part of that curriculum and that passed the test of item analysis was retained, while other items were revised or discarded. Nunnally (1964) points out that it is easy to show that highly discriminating items contribute to test reliability. Thus test developers have to select items that had high discrimination indexes.

In general the goal of empirical item review is to investigate and screen each item to determine if instructional or items deficiencies exist and to revise such troublesome items. Although data on test scores give some indication of the nature and quality of the test, additional information can be obtained from analyzing responses to individual test items.

To determine the index of difficulty and discrimination, as indicated by Meherns (1984), we take only test score of 27% from higher achiever and 27% from lower achiever. Thus, by taking into consideration the two groups, index of difficulty and discrimination of every item is computed.

Bean (1953) farther tells us an external criterion of any sort is available, however, there is a question that will probably important to be answered: to what extent does each item measure the same variables or group of variables that the test as a whole measures? He also explained that, each item may be correlated with the test as a whole to determine to what degree success on the item is related to success on the entire test, in case the measure as a whole is not valid, this process may be useless, but some important may be brought about by the study of what is known

as the internal consistency of a test. According to most writers for example, Bean (1953) and Brown (1983) and others a poorly constructed test items does not give consistent results.

To conclude this part, even though it is believed that statistical analysis of tests are the most powerful instruments to maintain the quality of test items, unless the original collecting of items should be carefully constructed, no statistical analysis can turn a poorly conceived and badly written collection of items into a good test (Nunnally, 1964.)

2.5 School Resource and Working Climate

Measurement is an essential but difficult aspect of the educational program (Adams, 1964). He also goes by saying that teachers are to persevere in becoming more informed and effective in his/her aspect of their teaching role; they need leadership, encouragement, and resource materials from their school administrator. Brown (1985) and Nunnally (1964) confirmed that carefully selected tests can provide data that will aid in making many professional judgments and decisions. It has also been recognized that there is a high correlation between good working conditions and high level of performance by workers.

Normally, not only teachers but also other professionals and non-professional workers leave their profession or occupation when the working climate (environment) is not conductive. For the school to be successful, the school climate should be conducive for the learners and for the teachers too. The school climate can be physical, psychological, social and economical setting that motivate or frustrate the learners, the teacher and their competencies (stinnett, 1968).

But for this paper we restrict our assessment on the following possible factors. The class size, teaching load, the amount of clerical work the teacher involved in, the workweek, free periods for the teachers, and too many extra-curricular activities.

2.5.1 Class Size

It is one of school based factors. As indicated by stinnett (1968) there is no well-defined class size that serves for all cases. Instead, the class size is highly influenced by the nature of the teachers, the maturity and ability level of the learners. As a rule of thumb, for high schools 25 students per classroom is moderate number.

2.5.2 Teaching Load

Dougless (1964) stated that from many factors that affect teaching are the number of sections thought daily (weekly), the number of pupils taught, the number of different preparations required, the amount of time required for co-operations, the length of class period etc. He father tell us that the maximum effectiveness of the teaching staff may be realized only when each teacher is, as nearly as possible, give the work to do which he can do best.

2.5.3 The Amount of Clerical Work

School reports and records seem constantly to increase in number and complexity. The clerical work expected of teachers in many schools is almost a full-time load in itself. The loads of teachers are greatly increased by the clerical work they are required to perform. The record keeping, results of testing programs, reports of parents, variety of public relations material, anecdotal evaluation of pupils. But there are several means of relieving the teacher of this load the most common being the employment of clerical workers or non-professional aids (stinnett, 1968).

2.5.4 Work Week

There is no definite answer for a reasonable work week for a teacher but studies consistently report that the typical teacher work week as ranging from forty-two to fifty hours. As Stinnite stated (1968) teaching is not and should never be a clock punching job, but reasonable limitations should be observed.

2.5.5 Free Period

Teachers need sufficient time in the school hours so that they can prepare the lesson they teach, using library, and discussion with their colleagues, sufficient rest and relaxation time. Teaching is a skilled craft, and its creative work is best done in an atmosphere of freedom (Smith and Harrison, 1960).

2.5.6 Too many Extra-Curricular Activities

High school teachers are almost overwhelmed by the duties performed outside class hours. In most high schools, preparation of the school paper and the school annual and clubs of all sorts are

performed on after-school time. Unfortunately, the load is no always equally distributed. The talented and able teachers draw the assignment some teachers carry a very burden of extra curricular activities.

2.6 Classroom Testing in Ethiopia

To the knowledge of the present researcher relevant information on the topic is very scarce on the examination system in general and classroom testing particular, in the Ethiopian educational system therefore, the following is only the highlight of the information compiled from fragmentary sources such as manuals, unpublished materials and others.

In Ethiopia, classroom tests are constructed and administered by classroom teachers at the various levels of the school system. As the school year is divided in to two semesters, the tests are patterned after the semesters. Accordingly, four or more progressive testes and one final examination are mandatory in each semester. The promotion of the pupils will be determined or decided on the basis of the average results obtained in the two semesters of the academic year (MOE, 1986).

Although the share of classroom tests in the over all student evaluations program is considered very difficult, it seems that no much attentions has been paid to the quality of testing, moreover, relevant documents are scarce concerning the nature, strength, weakness and significance of classroom tests in Ethiopia context, or, if at all, the information is not accusable to researchers.

There are general regulations and guidelines set by the Ministry of Education on the practice and process of classroom testing as stated in the guide book for managing and administering of schools (Amharic text) brief descriptions are given concerning to the objective, quality and type of classroom tests. Besides the points to be consider in testing ,weighting to be assigned to homework, progressive tests and final exams are indicated in the manual (MOE ,1986).

Constructing the tests, administering, scoring and reporting the results fall under the discretion of subject teachers. Whether they prepare the tests individually or in group, it is their duty and responsibility. However, teachers are expected to abide by the rules and regulations of the MOE and the directives of their school management in preparing and administering classroom tests.

This is because, finally the promotion policy is to be decided based on the achievement obtained from the teacher made tests.

CHAPTER THREE

3. Methods of the Study

3.1. Research Design

The main goal of this study was to assess teacher made tests quality and identify factors that affecting the analysis and evaluations of tests. In order to achieve this, descriptive research was used. The researcher has employed this design mainly because she believed that it is suitable and relevant for the study which looks into what actually happens in the test construction and analysis in the teaching learning process. To obtain descriptive information, the researcher used questionnaire for teachers, interviews for school principals and document analysis.

3.2. Research Setting, Population and Sampling

The data was collected in Addis Ababa in Nifas Silk Sub City. Three high schools were selected from the sub-city by using simple random sampling techniques and from the three schools samples, were selected by using simple random sampling techniques too. The schools are Ginbot 20 Secondary School, Abyot Fana Secondary School and Firehiwot Secondary School. Also from each sample school, one grade 9 section is randomly selected for the purpose of analyzing test quality. The data was 1st semester marked test papers of English and Mathematics and the responses of teachers and principals.

3.3. Research Instruments

3.3.1. The Questionnaire

Questionnaire was prepared and distributed to teachers. In this study the questionnaire was used to extract data embedded deep in teachers' minds or their attitudes, feelings or reactions. A questionnaire consisted of 24 items, was used to collect the required data. Pilot test was also

taken place by administering open and closed questionnaire to high school teachers, other than the sample schools included in the study. From the pilot study .75 reliability coefficients were seen.

3.3.2. The Interview

Interviews were conducted with the principals of the schools. The interview method enabled the researcher to get in depth views of the key information on issues that are relevant to the purpose of the study. The lists of questions were designed to guide the interview. It helps to secure data about the over all conditions of teachers and the working environment.

3.4.3. Document Analysis

First semester marked test papers were assessed to determine whether the construction of the tests were according to the guidelines of test construction and test item analysis, reliability was also made. Moreover, the time table of the schools and school record format were also assessed to find teachers' period's distributions and the class size.

3.4 Data Analysis Procedures

Both quantitative and qualitative methods of data analysis were used. Quantitative methods of data analysis such as tables were mainly used. However, the researcher, in most cases described in detail the results obtained from the different instruments. Percentages and frequencies were also used to analyze data. Quantitative analysis based on test item was done to determine the difficulty and discrimination level of the test items used. And reliability of test items by using split half method and Kadre Richardson KR−20 was done.

CHAPTER FOUR

4. Results and Discussion

The purpose of this study was to look into the factors that inhibited teachers from making post test analysis and evaluations; it also intended to assess the test construction skill of the teachers in Nifas Silk Sub City in three selected secondary schools.

Accordingly, this chapter deals with the analysis and interpretation of the research participants. In the following sections, the data are presented and discussed

4.1 Results

4.1.1 Results of Questionnaire

Table 1: Personal Characteristics of the Teachers

Items	Options	Ginbot 20	Firhiwot	Abyot Fana	Number	%
1. Sex	Male	15	19	16	50	83.33
	Female	5	1	4	10	16.67
2. Age	25 and below	8	4	5	17	28.33
	26- 35 years	9	8	10	27	45.00
	36-45 years	2	5	4	11	18.33
	Above 45 years	1	3	1	5	8.33
3. experience	Below 6 years	4	5	6	15	25
	6-15 years	10	8	7	25	41.67
	16-25 years	5	4	5	14	23.33
	Above 25 years	1	3	2	6	10.00
	Not responded	-	-	-	-	-

4. Field of study	Language	5	7	3	15	25.00
	Mathematics	5	4	5	14	23.33
	Natural science	7	4	8	19	31.67
	Social science	2	5	2	9	15.00
	Sport	1	-	1	2	3.33
	Not responded	-	-	1	1	1.67
5. Qualification	Diploma	-	-	-	-	-
	Bachelor Degree	20	20	20	60	100
	Mastered and above	-	-	-	-	-

In this study, the population consists of secondary school teachers and the school principals. Male teachers were 85%.The age of the study participants show that 45% of them were between the age of 26 to 35 years, 28% of them were in the age of 25, 18% of them were between the age of 36 to 45 the remaining 8% were in the age of 45 and above and 65% of them have teaching experience from 6 to 25 years. All teachers have bachelor degree.

Table 2: Training Background of the study Participants

Items	**Options**	**Ginbot 20**	**Firhiwot**	**Abyot Fana**	**Number**	**%**
Did you take training courses in measurement and evolution?	Yes	17	18	20	55	91.67
	No	3	2	-	5	8.33
	Not responded	-	-	-	-	-
Are you familiar with measurement and evaluation concepts?	Yes	17	18	16	51	85.00
	No		-	4	4	6.67
	Not responded	-	-	-	-	-
Have you practiced the principles of measurement and evaluation sufficiently?	Yes	12	19	14	45	75.00
	No	8	1	6	15	25.00
	Not responded	-	-	-	-	-

Teachers training condition and the existence of preferable condition to teachers to develop their knowledge and skill on measurement and evaluation were assessed.

As can be seen from table 2 above, nearly 92% of the teachers reported that they took training courses on measurement and evaluation. Only a few teachers (8%) have no training on measurement and evaluation. Also, many of the teachers (85%) are trained and familiar to measurement and evaluation.

Items	Options	Ginbot 20	Firhiwot	Abyot Fana	Number	%
Reviewing and eliminating items based on test tryout.	Yes	7	10	18	35	58.33
	No	13	7	2	22	36.67
	Not responded	-	3	-	3	5.00
Emphasis on details	Yes	15	12	15	42	70.00
	No	5	7	5	17	28.33
	Not responded	-	1	-	1	1.67
Using large number of items in the test	Yes	4	2	2	8	13.33
	No	16	16	18	50	83.33
	Not responded	-	2	-	2	3.33

Table 3: Teachers Skill on Test Construction

Too reliance on text book expression.	Yes	15	16	16	47	78.33
	No	5	4	4	13	21.67
	Not responded	-	-	-	-	-

As table 3 displays, (58%) of the teachers review and eliminate items based on test tryout. Regarding item writing skills, (70%) of the teachers emphasize on details on tests and (78%) largely depend too much on text book expression while writing tests. Finally, the majority of the respondents (83%) responded that they do not use large number of items in the test. in general; the data indicate that majority the respondents' depend on the text book expression, construct small number o items and emphasize on details in their tests.

Table 4: Teachers' Experience on the Pre-Conditions of Test Construction.

Items	Options	Ginbot 20	Firhiwot	Abyot Fana	Number	%
Using table of specification	Always	2	2	4	8	13.33
	Usually	1	2	2	5	8.33
	Sometimes	7	13	10	30	50.00
	Never	10	3	4	17	28.33
The frequency of preparing newly constructed test items in their exam.	Always	3	4	6	13	21.67
	Often	10	11	7	28	46.67
	Sometimes	7	3	6	16	26.67
	Never	-	2	1	3	5.00

As can be noticed from table 4, using table of specification, 50% of the teachers responded that they use it sometimes and 28% of them answered by saying never. Concerning preparing of new items 46% of the teachers responded that they often prepare new items while 26% of them said they sometimes prepare newly constructed test items.

Table 5: Teachers Experiences' on When to Prepare Their Tests

Items	Options	Ginbot 20	Firhiwot	Abyot Fana	Number	%
Did you construct test items little by little?	Yes	6	8	5	19	31.67
	No	14	12	15	41	68.33
	Not responded	-	-	-	-	-
Do you leave test construction to the last date?	Yes	17	12	15	44	73.33
	No	3	8	5	16	26.67
	Not responded	-	-	-	-	-

When do you start constructing test items?	Before the lesson	1	-	2	3	5.00
	During the lesson	9	12	4	11	15.00
	Test approaching	14	14	11	39	65.00
	Immediately day of the test	3	4	2	9	15.00

Table 5 demonstrates that only 31% of the teachers prepare test items little by little while the majority of them 68% responded that they do not. This implies that the majority of the teachers engage in the constructions of tests when the time of the test is approaching.

Similarly, 73% of the respondents replied that they prepare tests just a few days before the test date. Recently it is believed that preparing tests before the lesson provide learners with ample opportunity to meaningful practice in their lesson and motivates teachers to achieve their objective effectively.

Table 6: Teachers' Experience on Test Item Evaluation

Items	**Options**	**Ginbot 20**	**Firhiwot**	**Abyot Fana**	**Number**	**%**
The practice of peer evaluation of tests.	Yes	5	7	8	20	33.33
	No	15	13	12	40	66.67
	Not responded	-	-	-	-	-
Item analysis.	Yes	8	7	4	19	31.67
	No	12	13	16	41	68.33
	Not responded	-	-	-	-	-
How often did you use item analysis?	Always	5	7	3	15	25.00
	Sometimes	14	10	15	39	65.00
	One time	1	3	2	6	10.00
	Other	-	-	-	-	-

As can be shown in table 6, with regard of peer evaluation, the majority of the respondents 67% replied that they do not practice of peer evaluation. Only 33% of them responded that they did practice peer evaluation. As to the issue of test item analysis, similarly the majority of the teachers 68% responded that they did not make item analysis only about 32% of the respondents responded that they carried out item analysis. When asked how often they use item analysis, the majority of the respondents 65% replied that they conducted item analysis sometimes. From the data presented, it can be possible to say that the majority of teachers understudy did not practice peer evaluation of tests. They did not also make test item analysis.

4.1.2 Test Item Analysis

Table 7: English Test item Analysis for Ginbot 20 Secondary School

Item	Diff. index	Remarks	Discr. index	Remarks	Alternative	
1	0.575	acceptable	0.15	acceptable	Group U 20 L 20	A* B C D 13 2 0 5 10 3 1 6
2	0.725	acceptable	**0.05**	**Not acceptable**	**Group** **U 20** **L 20**	**A* B C D** **15 0 0 5** **14 2 3 1**
3	0.50	acceptable	0.20	acceptable	Group U 20 L 20	A* B C D 12 2 3 3 8 3 4 5
4	0.475	acceptable	0.25	acceptable	Group U 20 L 20	A* B C D 12 2 3 3 7 5 4 4
5	0.525	acceptable	0.35	acceptable	Group U 20 L 20	A B C* D 1 0 14 5 10 2 7 1
6	0.35	acceptable	**0.10**	**Not acceptable**	**Group** **U 20** **L 20**	**A B* C D** **4 8 4 4** **6 6 5 3**
7	**0.275**	**difficult**	**0.05**	**Not acceptable**	**Group** **U 20** **L 20**	**A B* C D** **8 6 2 4** **7 5 3 5**
8	0.675	acceptable	0.25	acceptable	Group U 20 L 20	A B C* D 2 0 16 2 3 2 11 4
9	0.55	acceptable	0.40	acceptable	Group U 20 L 20	A B* C D 11 5 2 2 5 7 4 4
10	0.60	acceptable	0.40	acceptable	Group U 20 L 20	A B C D* 1 3 0 16 6 4 2 8
11	0.50	acceptable	0.60	acceptable	Group U 20 L 20	A B C D* 0 2 2 16 4 6 6 4
12	**0.10**	**difficult**	**0.00**	**Not acceptable**	**Group** **U20** **L 20**	**A B* C D** **5 2 7 6** **7 2 8 3**
13	0.50	acceptable	0.20	acceptable	Group U20 L 20	A B* C D 2 12 3 3 4 8 4 4
14	**0.05**	**difficult**	**0.00**	**Not acceptable**	**Group** **U 20** **L 20**	**A B* C D** **6 1 5 8** **9 1 6 4**
15	0.425	acceptable	0.55	acceptable	Group U 20 L 20	A B* C D 2 14 0 4 5 3 6 6
16	**0.25**	**difficult**	**0.10**	**Not acceptable**	**Group** **U 20** **L 20**	**A B* C D** **5 6 6 3** **6 4 7 3**
17	0.625	acceptable	0.35	acceptable	Group U 20 L 20	A B* C D 2 16 2 0 6 9 3 2
18	0.425	acceptable	0.55	acceptable	Group U 20	A* B C D 14 3 2 1

					L 20	3 7 6 4
19	0.50	acceptable	0.40	acceptable	Group U 20 L 20	A B* C D 2 14 1 3 5 6 4 5
20	0.45	acceptable	0.50	acceptable	Group U 20 L 20	A B C* D 1 3 14 2 6 7 4 3
21	0.525	acceptable	0.45	acceptable	Group U 20 L 20	A B* C D 1 15 2 2 7 6 3 4
22	0.525	acceptable	0.35	acceptable	Group U 20 L 20	A B* C D 5 6 6 3 6 4 7 3
23	0.35	acceptable	0.40	acceptable	Group U 20 L 20	A B C D* 2 3 4 11 6 4 7 3
24	**0.275**	**difficult**	**0.45**	discriminate	**Group** **U 20** **L 20**	**A B C* D** **2 4 10 4** **5 7 1 7**
25	0.45	acceptable	0.40	acceptable	Group U 20 L 20	A B* C D 11 3 3 3 6 5 4 5
26	0.425	acceptable	0.55	acceptable	Group U 20 L 20	Correct responses 14 3
27	0.375	acceptable	0.45	acceptable	Group U 20 L 20	Correct responses 12 3
28	0.425	acceptable	0.55	acceptable	Group U 20 L 20	Correct responses 14 3
29	**0.20**	**difficult**	0.20	discriminate	**Group** **U 20** **L 20**	Correct responses **6** **2**
30	**0.275**	**difficult**	0.45	discriminate	**Group** **U 20** **L 20**	Correct responses **10** **1**
31	**0.275**	**difficult**	0.45	discriminate	**Group** **U 20** **L 20**	Correct responses **10** **1**
32	**0.175**	**difficult**	0.35	discriminate	**Group** **U 20** **L 20**	Correct responses **7** **0**
33	**0.075**	**difficult**	0.15	discriminate	**Group** **U 20** **L 20**	Correct responses **3** **0**
34	**0.225**	**difficult**	0.35	discriminate	**Group** **U 20** **L 20**	Correct responses **8** **1**
35	**0.225**	**difficult**	0.25	discriminate	**Group** **U 20** **L 20**	Correct responses **7** **2**
36	0.325	acceptable	0.55	acceptable	Group U 20	Correct responses 12

					L 20	1
37	**0.22**	**difficult**	0.45	discriminate	**Group** **U 20** **L 20**	Correct responses **9** **0**
38	**0.15**	**difficult**	0.30	discriminate	**Group** **U 20** **L 20**	Correct responses **6** **0**
39	**0.275**	**difficult**	0.45	discriminate	**Group** **U 20** **L 20**	Correct responses **10** **1**
40	**0.15**	**difficult**	0.30	discriminate	**Group** **U 20** **L 20**	Correct responses **6** **0**

In the final English exam for Ginbot 20 Secondary School, 40 items were presented. In the above table all the items are included and out of 40 items 23 of them are appropriate items the remaining 17 items which are written in bold are defective.

Out of these 40 items, 25 (1-25) were multiple choices,10 (26- 35) items in which 12 vocabulary options were given to be filled in the 10 blank spaces provided were filling the blank,6 (36-40) items were to be converted from direct speech in to indirect speech. Each item in all questions carried or held one point score.

The above table reveals that 17 (43%) of the items in English exams were found to be defective. Out of these 15 items were very difficult which were correctly answered by only 5 to 20 percent of the test takers. Moreover, four of the seventeen items were defective in both discrimination and difficulty index. From the table the value “0.00” for the discriminations index indicates that lower achievers and higher achievers answered the item similarly.

As we noticed from the above table the items which are defective are not only defective on discriminating and difficulty level but also the options of the items are not plausible, they were not attracted the lower achievers as expected. Also in items 12 and 14 options are almost equally attract both sides but they were the correct answers.

Table 8: Mathematics Test item Analysis for Ginbot 20 Secondary School

Item	**Diff. index**	**Remarks**	**Discr. index**	**Remarks**	**Alternative**

1	0.425	acceptable	**0.05**	**Not acceptable**	**Group** **U 20** **L 20**	**T* F** **9 11** **8 12**
2	0.375	acceptable	0.55	acceptable	Group U 20 L 20	T* F 13 7 2 18
3	0.60	acceptable	**-0.10**	**Not acceptable**	**Group** **U 20** **L 20**	**T* F** **11 9** **13 7**
4	**0.25**	**difficult**	0.20	acceptable	**Group** **U 20** **L 20**	**T* F** **7 13** **3 17**
5	0.575	acceptable	0.25	acceptable	Group U 20 L 20	T* F 14 6 9 11
6	0.40	acceptable	0.30	acceptable	Group U 20 L 20	T F* 9 11 15 5
7	0.475	acceptable	0.45	acceptable	Group U 20 L 20	T F* 6 14 15 5
8	0.325	acceptable	0.35	acceptable	Group U 20 L 20	Correct responses 10 17
9	0.50	acceptable	0.20	acceptable	Group U 20 L 20	Correct responses 12 8
10	0.30	acceptable	**0.10**	**Not acceptable**	**Group** **U 20** **L 20**	Correct responses **7** **5**
11	0.525	acceptable	0.25	acceptable	Group U 20 L 20	A *B C D 3 5 4 8 2 7 5 6
12	0.325	acceptable	0.15	acceptable	Group U 20 L 20	A*B C D 3 5 4 8 2 7 5 6
13	**0.125**	**difficult**	**0.05**	**Not acceptable**	**Group** **U 20** **L 20**	**A*B C D** **3 5 4 8** **2 7 5 6**
14	**0.275**	**difficult**	0.25	acceptable	**Group** **U 20** **L 20**	**A* B C D** **8 3 4 5** **3 4 6 7**
15	0.325	acceptable	0.15	acceptable	Group U 20 L 20	A* B C D 8 3 5 4 5 4 6 5
16	**0.125**	**difficult**	**0.05**	**Not acceptable**	**Group** **U 20** **L 20**	**A B* C D** **4 3 6 7** **5 2 7 6**
17	**0.20**	**difficult**	**0.10**	**Not acceptable**	**Group** **U 20** **L 20**	**A B *C D** **4 5 6 5** **5 3 8 4**
18	0.375	acceptable	0.35	acceptable	Group U 20 L 20	A* B C D 11 3 4 2 4 5 6 5

19	0.35	acceptable	0.60	acceptable	Group U 20 L 20	A* B C D 13 2 3 2 1 5 7 7
20	0.45	acceptable	0.40	acceptable	Group U 20 L 20	A B C D* 1 5 1 13 5 7 6 2
21	0.425	acceptable	0.25	acceptable	Group U 20 L 20	A B C D* 2 4 3 11 6 3 6 5
22	0.425	acceptable	0.35	acceptable	Group U 20 L 20	A* B C D 12 4 2 2 5 5 6 4
23	**0.20**	**difficult**	0.30	acceptable	**Group** **U 20** **L 20**	**A B C* D** **2 6 7 5** **5 9 1 5**
24	0.375	acceptable	0.25	acceptable	Group U 20 L 20	A* B C D 9 3 2 6 4 5 4 7
25	**0.20**	**difficult**	0.20	acceptable	**Group** **U 20** **L 20**	**A B C D*** **5 6 3 6** **7 7 4 2**
26	**0.20**	**difficult**	0.30	acceptable	**Group** **U 20** **L 20**	Correct responses **7** **1**
27	0.30	acceptable	0.50	acceptable	Group U 20 L 20	Correct responses 11 1
28	**0.125**	**difficult**	0.25	acceptable	**Group** **U 20** **L 20**	Correct responses **5** **0**

In the Mathematics final exam for Ginbot 20 Secondary School, 28 items were presented. In the above table all the items are included. Out of the total items 7(1-7) were true or false, 3 (8-10) items were matching, 14 (11-25) were multiple choice, and 3 (26-28) were filling the blank. Each item in all questions held one point score.

Besides, the above table indicates that 12 (43%) defective items, 4 (10%) in both discrimination and difficulty indexes, the remaining 6 items were difficult and 3 items were not discriminating. But generally, the items in the test were not very easy items.

In the discrimination indexes 6 of the items were not acceptable to discriminate among students. Especially item 3 had negative discriminating index, negative discriminating index showed that lower achiever answered correctly than the upper one.

Table 9: English Test Item Analysis for Abyot Fana Secondary School

Item	Diff. index	Remark	Discr. index	Remarks	Alternatives

1	0.525	acceptable	0.55	acceptable	Group U 20 L 20	A* B C D 16 2 1 1 5 7 5 3
2	0.775	acceptable	0.25	acceptable	Group U 20 L 20	A* B C D 18 0 0 2 13 3 1 3
3	0.625	acceptable	0.35	acceptable	Group U 20 L 20	A B C* D 1 2 16 1 4 5 9 2
4	0.75	acceptable	0.40	acceptable	Group U 20 L 20	A* B C D 19 0 0 1 11 2 3 4
5	0.725	acceptable	0.35	acceptable	Group U 20 L 20	A B C* D 1 0 18 1 4 2 11 3
6	0.45	acceptable	0.50	acceptable	Group U 20 L 20	A* B C D 14 2 1 3 4 6 5 5
7	0.50	acceptable	0.60	acceptable	Group U 20 L 20	A B C D* 2 2 0 16 7 6 3 4
8	**0.25**	**difficult**	0.20	acceptable	**Group** **U 20** **L 20**	**A B C D*** **4 6 3 7** **4 8 5 3**
9	**0.05**	**difficult**	**0.10**	**Not acceptable**	**Group** **U20** **L20**	**A* B C D** **2 7 5 6** **0 8 8 4**
10	0.475	acceptable	0.75	acceptable	Group U 20 L 20	A* B C D 12 3 1 4 7 5 3 5
11	**0.20**	**difficult**	0.30	acceptable	**Group** **U20** **L20**	**A B C* D** **5 4 7 4** **6 9 1 4**
12	0.60	acceptable	**0.10**	**Not acceptable**	**Group** **U20** **L20**	**A B C * D** **3 2 3 2** **4 4 11 1**
13	0.50	acceptable	0.30	acceptable	Group U 20 L 20	A* B C D 13 2 3 2 7 4 6 3
14	0.65	acceptable	0.50	acceptable	Group U 20 L 20	A B C* D 2 0 18 0 4 6 8 2
15	0.47	acceptable	**0.05**	**Not acceptable**	**Group** **U20** **L20**	**A* B C D** **10 2 5 3** **9 0 4 7**
16	0.60	acceptable	0.60	acceptable	Group U 20 L 20	A* B C D 18 0 0 2 6 5 6 3
17	0.65	acceptable	0.50	acceptable	Group U 20 L 20	A* B C D 18 0 0 2 8 3 5 4
18	**0.05**	**difficult**	**0.10**	**Not acceptable**	**Group** **U 20** **L 20**	**A* B C D** **2 8 6 4** **0 8 8 4**
19	0.32	acceptable	0.56	acceptable	 U20 L20	A B* C D 7 9 2 2 7 4 5 4
20	0.30	acceptable	0.20	acceptable	Group U 20 L 20	A B C* D 4 5 8 3 5 7 4 4
21	0.325	acceptable	0.35	acceptable	Group	A B* C D

					U 20 L 20	3 10 4 3 5 3 7 5
22	0.30	acceptable	**-0.30**	**Not acceptable**	**Group** **U20** **L20**	**A B C * D** **4 13 3 0** **6 3 9 2**
23	0.60	acceptable	0.50	acceptable	Group U 20 L 20	A B C D* 2 0 1 17 5 4 4 7
24	0.35	acceptable	**0.00**	**Not acceptable**	**Group** **U20** **L20**	**A B C * D** **2 6 7 5** **6 1 7 6**
25	0.60	acceptable	0.70	acceptable	Group U 20 L 20	A B C D* 0 0 1 19 5 4 6 5
26	0.50	acceptable	0.60	acceptable	Group U 20 L 20	A B C* D 2 0 16 2 5 6 4 5
27	**0.15**	**difficult**	**-0.10**	**Not acceptable**	**Group** **U20** **L20**	**A B* C D** **6 2 6 6** **9 4 1 6**
28	0.60	acceptable	0.40	acceptable	Group U 20 L 20	A B C D* 1 2 11 6 5 4 3 8
29	0.57	acceptable	0.15	acceptable	Group U 20 L 20	A* B C D 13 2 1 4 10 3 2 5
30	0.40	acceptable	0.20	acceptable	Group U 20 L 20	A* B C D 12 2 4 2 4 6 5 5
31	0.75	acceptable	0.30	acceptable	Group U 20 L 20	A B C* D 1 1 18 0 3 2 12 3
32		jumped		jumped	jumped	
33	0.30		**0.00**	**Not acceptable**	**Group** **U20** **L20**	**A B* C D** **14 6 0 0** **4 6 5 5**
34		jumped		jumped	jumped	
35	0.50	acceptable	0.50	acceptable	Group U20 L20	A B* C D 3 15 0 2 1 5 7 7
36	**0.20**	**difficult**	**0.10**	**Not acceptable**	**Group** **U 20** **L 20**	**A* B C D** **5 4 2 2** **3 5 6 4**
37	0.425	acceptable	0.35	acceptable	Group U 20 L 20	A* B C D 12 4 2 2 5 5 6 4

In this school for English test, 37 items were presented in the final exam for first semester. In the above table all the 37 items are included, 2 items were not scored, items 32 and 34, moreover some items are defective and the defective items are written in bold.

Among 37 English items used for analysis 11(27%) were found to be defective. From these 11 items 4 were very difficult in that only very small students answered it correctly. Also 4(10%) items were more defective both in discrimination and difficulty indexes.

The defective items were not only in discriminating and difficulty level; the options by themselves were not properly constructed. In items 22 and 27 the correct answers attracted more of the lower achievers than the upper achievers. Also in item 18 option, D; item 19 option, A; item 27 option, D were equally attract the lower achievers and the upper achievers.

Table 10: Mathematics Test Item Analysis for Abyot Fana Secondary School

Item	**Diff. index**	**Remark**	**Discr. index**	**Remarks**	**Alternatives**	
1	0.325	acceptable	**-0.05**	**Not acceptable**	**Group** **U20** **L20**	**T F*** **13 6** **14 7**
2	0.675	acceptable	0.35	acceptable	Group U20 L20	T* F 17 3 10 10
3	0.50	acceptable	0.20	acceptable	Group U20 L20	T* F 12 8 8 12
4	0.55	acceptable	0.60	acceptable	Group U20 L20	T* F 17 3 5 15
5	0.425	acceptable	0.35	acceptable	Group U20 L20	T F* 8 12 15 5
6	0.675	acceptable	0.35	acceptable	Group U20 L20	A B C D* 2 1 0 17 3 4 3 10
7	0.775	acceptable	0.35	acceptable	Group U20 L20	A B C D* 0 0 1 19 3 4 1 12
8	**0.20**	**difficult**	**0.10**	**Not acceptable**	**Group** **U20** **L20**	A B* C D **6 3 5 6** **4 5 6 5**
9	0.55	acceptable	0.90	acceptable	Group U20 L20	A B* C D 0 20 0 0 6 2 7 5
10	0.50	acceptable	0.80	acceptable	Group U20 L20	A B* C D 1 18 1 0 7 7 2 4
11	0.65	acceptable	0.60	acceptable	Group U20 L20	A B C D* 1 0 0 19 5 4 4 7
12	0.65	acceptable	0.60	acceptable	Group U20 L20	A B C * D 1 0 19 0 6 4 7 3
13	**0.20**	**difficult**	**-0.20**	**Not acceptable**	**Group** **U20** **L20**	**A B C D*** **7 5 6 2** **6 1 7 6**
14	**0.18**	**difficult**	**-0.15**	**Not acceptable**	**Group** **U20** **L 20**	**A B C D*** **7 5 6 2** **6 4 5 5**

15	0.75	acceptable	0.40	acceptable	Group U 20 L 20	A* B C D 19 0 0 1 11 2 3 4
16	**0.05**	**difficult**	**0.00**	**Not acceptable**	**Group** **U20** **L20**	**A B C* D** **7 9 1 3** **8 4 1 7**
17	0.60	acceptable	0.15	acceptable	Group U 20 L 20	A* B C D 12 4 2 2 5 5 6 4
18	0.325	acceptable	0.25	acceptable	Group U 20 L 20	A* B C D 12 4 2 2 5 5 6 4
19	0.40	acceptable	0.20	acceptable	Group U 20 L 20	A B* C D 3 10 2 5 4 6 3 7
20	**0.10**	**difficult**	**-0.10**	**Not acceptable**	**Group** **U20** **L 20**	**A B C* D** **7 9 1 3** **7 4 3 6**
21	0.40	acceptable	0.50	acceptable	Group U20 L 20	A B C D 13 3 2 2 3 6 5 6
22	**0.20**	**difficult**	0.20	acceptable	**Group** **U20** **L20**	**A B C D*** **5 6 3 6** **4 5 9 2**
23	0.55	acceptable	0.40	acceptable	Group U 20 L 20	A* B C D 12 4 2 2 5 5 6 4
24		Jumped		Jumped		
25		Jumped		Jumped		
26	0.33	acceptable	**-0.05**	**Not acceptable**	**Group** **U20** **L20**	**A B C D*** **2 6 6 6** **4 2 7 7**
27	0.475	acceptable	0.55	acceptable	Group U 20 L 20	A* B C D 1 2 2 1 4 6 5 5
28	0.45	acceptable	0.30	acceptable	Group U 20 L 20	A* B C D 15 2 1 1 3 5 7 5
29	0.475	acceptable	0.45	acceptable	Group U 20 L 20	A B C D* 3 2 1 14 6 4 5 5
30	**0.23**	**difficult**	**-0.05**	**Not acceptable**	**Group** **U20** **L20**	**A B C* D** **5 5 4 6** **5 6 5 4**
31	0.475	acceptable	0.55	acceptable	Group U 20 L 20	A* B C D 15 2 1 2 4 6 5 5
32	**0.23**	**difficult**	**0.10**	**Not acceptable**	**Group** **U20** **L20**	**A B C * D** **4 6 5 5** **3 6 4 7**

As can be seen from table 10, 32 items are presented, out of that some of them are defective which are written in bold. Again two omitted items are existed. Out of 32 items 10 (34 %) were

found to be defective. Out of these 7 (18%) items were more defective both in discrimination and difficulty indexes.

Again 9 items were defective in discriminating among students. Moreover 6 of the items have negative discriminating indexes which mean that most of the lower achievers answered correctly than the upper one.

Out of the multiple choice items, those items in the above table were not only defective in discriminating and difficulty level but also the options did not attract the lower achievers and upper achiever as expected. In item 8, 13, 14, 20, 30, the correct answers were answered by lower achievers than the upper achievers. Also in item 16, option, C and item 20 option, A are equally attract the lower achievers and upper achievers.

Table 11: English Test Item Analysis for Firehiwot Secondary School

Item	**Diff. index**	**Remark**	**Discr. index**	**Remarks**	**Alternatives**	
1	**0.05**	**difficult**	**0.11**	**Not acceptable**	**Group** **U 18** **L 18**	**A B* C D** **7 2 4 5** **5 0 3 10**
2	**0.27**	**difficult**	**0.11**	**Not acceptable**	**Group** **U 18** **L 18**	**A B C* D** **7 3 6 2** **5 6 4 3**
3	0.416	acceptable	0.16	acceptable	Group U 18 L 18	A B C* D 3 3 4 2 4 6 5 3
4	**0.27**	**difficult**	**0.11**	**Not acceptable**	**Group** **U 18** **L 18**	**A B C*D** **5 4 6 3** **7 6 4 1**
5	**0.14**	**difficult**	0.16	acceptable	**Group** **U 18** **L 18**	**A* B C D** **4 5 4 5** **1 6 7 4**
6	0.416	acceptable	0.277	acceptable	Group U 18 L 18	A B C* D 3 3 10 2 5 5 5 3
7	0.416	acceptable	0.16	acceptable	Group U 18 L 18	A B C* D 2 4 9 3 3 5 6 4
8	0.638	acceptable	0.722	acceptable	Group U 18 L 18	A* B C D 18 0 0 0 5 7 5 3
9	0.44	acceptable	0.66	acceptable	Group U 18 L 18	A B C* D 0 1 1 4 3 6 5 2 5
10	0.44	acceptable	0.66	acceptable	Group U 18	A B C D* 1 2 1 14

					L 18	5 4 3 6
11	0.55	acceptable	0.33	acceptable	Group U 18 L 18	A B C D* 2 3 0 13 4 5 2 7
12	**0.27**	**difficult**	**-0.22**	**Not acceptable**	**Group U 18 L 18**	**A B C D* 4 5 6 3 6 4 1 7**
13	0.66	acceptable	0.44	acceptable	Group U 18 L 18	A B* C D 0 16 2 0 3 8 5 2
14	0.39	acceptable	**0.00**	**Not acceptable**	**Group U 18 L 18**	**A*B C D 7 3 4 4 7 6 3 2**
15	0.61	acceptable	0.55	acceptable	Group U 18 L 18	A B* C D 0 16 2 0 3 6 5 4
16	0.583	acceptable	0.61	acceptable	Group U 18 L 18	A B* C D 0 15 2 1 3 6 5 4
17	0.583	acceptable	0.61	acceptable	Group U 18 L 18	A B C* D 0 1 17 0 5 6 4 3
18	0.50	acceptable	0.33	acceptable	Group U 18 L 18	A B C D* 0 4 2 12 2 5 5 6
19	**0.25**	**difficult**	**0.05**	**Not acceptable**	**Group U 18 L 18**	**A* B C D 5 4 5 4 4 6 5 3**
20	0.305	acceptable	0.277	acceptable	Group U 18 L 18	A B* C D 3 8 2 5 4 3 5 6
21	0.527	acceptable	0.50	acceptable	Group U 18 L 18	A B* C D 0 16 2 0 3 8 5 2
22	0.61	acceptable	0.55	acceptable	Group U 18 L 18	A* B C D 16 2 0 0 6 4 3 5
23	0.50	acceptable	0.77	acceptable	Group U 18 L 18	A* B C D 16 2 0 0 2 8 5 3
24	0.61	acceptable	0.66	acceptable	Group U 18 L 18	A B* C D 0 17 1 0 4 5 5 4
25	0.638	acceptable	0.50	acceptable	Group U 18 L 18	A B C D* 0 0 2 16 3 3 5 7
26	0.694	acceptable	0.611	acceptable	Group U 18 L 18	A B* C D 0 18 0 0 3 7 5 3
27	0.388	acceptable	0.55	acceptable	Group U 18 L 18	A B C D* 1 3 2 12 5 6 5 2
28	0.33	acceptable	0.55	acceptable	Group U 18	A B C D* 2 2 3 11

					L 18	4 8 5 1
29	0.611	acceptable	0.55	acceptable	Group U 18 L 18	A B C D* 0 0 2 16 3 4 5 6
30	**0.22**	**difficult**	**-0.22**	**Not acceptable**	**Group** **U 18** **L 18**	**A* B C D** **2 5 5 6** **6 6 3 3**
31	**0.16**	**difficult**	**0.11**	**Not acceptable**	**Group** **U 18** **L 18**	**A* B C D** **4 4 5 5** **2 6 7 3**
32	0.777	acceptable	0.33	acceptable	Group U 18 L 18	A B* C D 0 17 1 0 2 11 3 2
33	0.583	acceptable	0.277	acceptable	Group U 18 L 18	A* B C D 13 2 3 0 8 3 5 2
34	0.638	acceptable	0.61	acceptable	Group U 18 L 18	A B* C D 0 17 2 0 3 6 5 4
35	0.611	acceptable	0.33	acceptable	Group U 18 L 18	A B C* D 3 11 4 0 4 3 8 3
36	0.33	acceptable	**0.00**	**Not acceptable**	**Group** **U 18** **L 18**	**A B* C D** **4 6 4 4** **3 6 3 6**
37	0.527	acceptable	0.277	acceptable	Group U 18 L 18	A B C* D 2 3 12 1 3 4 7 4
38	0.44	acceptable	**0.11**	**Not acceptable**	**Group** **U 18** **L 18**	**A B C* D** **4 3 9 2** **5 4 7 2**
39	0.50	acceptable	0.33	acceptable	Group U 18 L 18	A B* C D 3 12 2 1 4 6 5 4
40	0.833	acceptable	0.33	acceptable	Group U 18 L 18	A* B C D 18 0 0 0 12 2 1 3

As can be reported in Table 11 above, 40 English test items were taken. Out of 40 items some are defective in both discrimination and difficulty level. The defective items are written in bold in the above table.

Out of 40 items, 11(28%) items were found to be more defective and 7 (18%) of the items were defective both in discrimination and difficulty indexes.

As the data show 8 items were very difficult and were answered correctly only by 5 to 25 percent of the test takers. Again, 10 of the items were defective in their discrimination power. Especially

negative indexes indicate that these items were answered correctly more by lower achiever than upper achievers.

In Firehiwot School, all the English exam items are in the multiple choice form. The defective items were not only in discriminating and difficulty level. The options by themselves were not properly constructed. In items 12, 30, and 35 the correct answers were chosen more by the lower achievers than the upper achievers, also in items 14 and 36 the correct answers were chosen equally by the lower achievers and upper achievers.

Table 12: Mathematics Test Item Analysis for Firehiwot Secondary School

Item	**Diff. index**	**Remark**	**Discr. index**	**Remarks**	**Alternatives**	
1	0.583	acceptable	0.388	acceptable	Group U 18 L 18	A* B C D 14 2 1 1 7 5 3 3
2	**0.05**	**Difficult**	**0.11**	**Not acceptable**	**Group** **U 18** **L 18**	**A* B C D** **2 4 5 7** **0 6 7 5**
3	0.583	acceptable	0.388	acceptable	Group U 18 L 18	A B* C D 0 18 0 0 3 12 2 1
4	0.77	acceptable	**0.11**	**Not acceptable**	**Group** **U 18** **L 18**	**A B* C D** **0 15 3 0** **0 13 5 0**
5	0.472	acceptable	0.722	acceptable	Group U 18 L 18	A B C* D 2 3 12 1 4 6 5 3
6	0.638	acceptable	0.50	acceptable	Group U 18 L 18	A B C* D 2 0 16 0 4 5 7 2
7	0.52	acceptable	**0.05**	**Not acceptable**	**Group** **U 18** **L 18**	**A*B C D** **10 0 5 3** **9 0 4 4**
8	0.50	acceptable	0.33	acceptable	Group U 18 L 18	A B C* D 1 3 12 2 5 4 6 3
9	0.527	acceptable	0.722	acceptable	Group U 18 L 18	A B C* D 1 1 16 0 7 6 3 2
10	0.58	acceptable	**0.05**	**Not acceptable**	**Group** **U 18** **L 18**	**A B* C D** **4 11 0 3** **1 10 0 8**
11	0.44	acceptable	0.33	acceptable	Group U 18 L 18	A B C* D 3 2 11 2 5 4 5 4
12	0.638	acceptable	0.611	acceptable	Group U 18	A* B C D 16 0 1 1

					L 18	7 4 5 2
13	0.50	acceptable	0.44	acceptable	Group U 18 L 18	A* B C D 13 1 3 1 5 4 6 3
14	0.638	acceptable	0.388	acceptable	Group U 18 L 18	A B C * D 1 3 14 0 2 4 8 4
15	0.61	acceptable	**0.11**	**Not acceptable**	**Group** **U 18** **L 18**	**A* B C D** **12 0 3 3** **10 0 5 3**
16	0.694	acceptable	0.50	acceptable	Group U 18 L 18	A* B C D 17 1 0 0 8 3 4 3
17	0.75	acceptable	0.50	acceptable	Group U 18 L 18	A B C* D 0 0 18 0 4 3 9 2
18	0.694-	acceptable	0.50	acceptable	Group U 18 L 18	A* B C D 17 1 0 0 8 3 5 2
19	0.44	acceptable	**0.05**	**Not acceptable**	**Group** **U 18** **L 18**	**A B* C D** **4 10 0 4** **4 9 0 5**
20	**0.27**	acceptable	**0.11**	**Not acceptable**	**Group** **U 18** **L 18**	**A B C* D** **5 5 6 2** **4 7 4 3**
21	0.638	acceptable	0.722	acceptable	Group U 18 L 18	correct responses 18 5
22	0.583	acceptable	0.833	acceptable	Group U 18 L 18	correct responses 18 3
23	**0.19**	acceptable	0.27	acceptable	**Group** **U 20** **L 20**	correct responses **6** **1**
24	0.527	acceptable	0.94	acceptable	Group U 20 L 20	correct responses 18 1
25	0.527	acceptable	0.277	acceptable	Group U 20 L 20	correct responses 17 2
26	0.416	acceptable	0.611	acceptable	Group U 20 L 20	correct responses 13 2
27	**0.19**	acceptable	**0.05**	**Not acceptable**	**Group** **U 20** **L 20**	correct responses **4** **3**
28	0.39	acceptable	**0.03**	**Not acceptable**	**Group** **U 20** **L 20**	correct responses **10** **4**
29	0.333	acceptable	0.55	acceptable	Group U 20 L 20	correct responses 11 1
30	**0.11**	**difficult**	**0.11**	**Not acceptable**	**Group** **U 20** **L 20**	correct responses **3** **1**

For this school 30 items for mathematics were presented for first semester final exam. The above table has both the defective and acceptable items. The defective items are written in bold.
According to the above table Firehiwot Secondary School of grade 9 students out of 30 items mathematics exam 11(33%) of the items were found to be defective. Again out of these items 4(10%) of them were defective in both discriminating and difficulty indexes.

In difficulty level only 5 items were defective. That indicate these items were very difficult. These were answered only by 5 to 25 percent of the test taker. On the contrary no very easy items existed.

Also 10 items were defective in discrimination power. This means these items were responded correctly by lower achiever and upper achiever similarly.

Here in this school 10 items were matching type which was presented together with multiple choices. Again out of them the defective items were seen as we observe in the above table.

4.1.3 Reliability of Tests

From each of the selected high school, grade 9 English and Mathematics tests were taken to estimate reliability of the test. And the reliability of these tests was presented as follows:

Table 13: Reliability of the Tests

School	**Subject**	**Reliability of the test**		**Remark**
		Split half method	**KR-20**	
Abyot Fana	English	0.7752	0.8723	
	Mathematics	0.5557	0.5734	*****
Ginbot 20	English	0.9644	0.9744	
	Mathematics	0.7734	0.7822	

Firehiwot	English	0.8770	0.8841	
	Mathematics	0.9505	0.9642	

*less consistent

The above table reveled that the tests were highly consistence except the Mathematics test of Abyot Fana Secondary School. From this it is possible to conclude that English tests of all the selected schools and Mathematics tests of Ginbot 20 and Firehiwot high schools are homogenous and measure similar characteristics of a person.

4.1.4 Results of Document Analysis

To assess school based factors that may affect or inhibit teachers from carrying out test item analysis and evaluation, documents which indicate average class size and teaching load, were assessed and analyzed accordingly.

With regard to class size the analysis of the documents demonstrated that, the majority of the teachers teach in large class size that is, about 80 students in a class. Regarding the teaching load still the majority of the teachers were required to teach 12-20 periods per a week, from the data obtained from the documents, it is possible to claim that the school based factors such as class size and teaching load could detract teachers from conducting test item analysis

Furthermore, an attempt was made to assess the number of tests teachers prepared for the students with in a semester. As the document showed, the teachers were required to prepare three to five tests in a semester.

English and mathematics first semester final exams were further examined to see whether or not they fulfill other psychometric qualities using checklists as a guide. The analysis uncovered the following points.

- Giving the weight assigned to each item is not usually practiced.
- There is no stated time in some exam papers.
- In most exam papers the instructions are not stated clearly, also they don’t indicate how to respond.

- The item format seems dependent through all exam papers.
- Multiple choice item formats seems the most preferred item format than other. On the other hand almost in all exam papers the options are listed diagonally in one line, this makes the items be crowded in one page and seems difficult to read. Again from item analysis it is possible to identify many vague or ambiguous items.
- True or False items appear in some exam. Papers and they seems clear and haven't trivial statements.
- Matching items also included in some exam papers but they lack clearly stated basis for matching in their instruction.

4.1.5 Results of the Interview

Interviews with the principals were held to secure additional data about the teachers and working environment of their schools.

All of the principals agreed that teacher-made tests are often subject to question because they differ greatly from class to class and their quality is also open to debate. Although standardized large scale assessments command all the media attention, it's the day to day classroom assessments that have the greatest impact on student learning. Nearly all the assessment events that take place in a student's life happen at the hands of the teacher. They align most closely with day to day instruction and are most influential in terms of their contribution to student.

As the researcher tried to mention earlier most of the tests teachers took as students were multiple choices. Teachers have very little practice constructing problem-solving tests which measure the application of skills and higher-order thinking.

The question which asked the school principal in relation to the importance of the teacher–made tests, they said that they are very vital in the teaching learning process if they are integrated to the daily lessons. They also added that the tests motivated learners to check their progress in the subject. Moreover the principals suggest that the tests are very useful for teachers to see their teaching methodology is appropriate for communicating with the students. Teachers can make changes immediately to fulfill the need of students.

Regarding to the quality of teacher-made tests, the principals suggested that the tests were not free from ambiguity for learners. The participants comment that the instruction and the items were not as clear as they should be.

A question was raised to the school principals concerning the manual prepared by MOE (1986) which contains test related issues. The interviews conducted revealed that the school principals were not aware of such manual which comes from MOE. Regarding the examination and evaluation committee, the school principals indicated that there is committee in all schools that would monitor the proper conduct of the examinations. According to MOE (1986) every school should have this committee also these committees' with a clearly described responsibilities and tasks .Concerning the member of this committee the school deputy principals are leaders and department heads are members. The expected duties and responsibilities of these committees according to MOE (1986) are to evaluate tests' strength and weakness, test reliability and validity, test preparations according to the guidelines and stated objective again to check test neatness are some of the expected one. However, the study disclosed that the committees in all schools were engaged in activities other than what has been mentioned above.

Regarding teachers' evaluation and analysis of their test the principals said that it is difficult to say properly all the teachers are evaluating and analyzing their tests. They also added that they only know teachers are trained to be a teacher. Concerning to shortage of time and the teachers responsibility to involve in clubs, the principals said that teachers have enough time to accomplish their duties but they should involve at least in 3 cubs which are organized in the schools.

4.2. Discussion

This section deals with the discussion of results related to the test quality and teachers inability to make post test analysis and evaluation.

To construct better quality test items, the experience of the teachers coupled with conducive environment is very important (Adams, 1964). As indicated by Yeager (as cited in Hailu 1990) experience alone is nothing with out creating a stimulating environment which is capable of encouraging teachers to develop their professional skill. This idea is also confirmed by Stinnett

(1970) that the quality of test with out considering the teacher competency and school based factors is meaningless.

As it was indicated by teachers' responses to the question about their experience too, teachers in the city are well experienced. The majority of them, about 83% have experience ranging from 6 to 15 years. However; the test item analysis reveals that they are defective. This shows that experience alone does not bring constructions of higher quality test items. This might also indicate that, the school environment may not so stimulating for teachers to construct better quality test. Since teachers work is highly related with testing activities, teachers need to have enough time and proper activities of the testing process (Brown, 1983).

In order to know to what extent learning has been taken place, teachers should design test to measure it (Ebel, 1975 and Gronlund, 1981). They also tell us that subject matter knowledge is not enough; teachers should have the skill and knowledge of educational measurement and evolutions. As studies listed above, stress on the importance of teachers' knowledge and skills in area of measurement and evaluations.

In this study context, teachers' responses on their training conditions indicate that 89% of teachers have taken training of teaching and they are familiar with measurement and evaluation skills.

Barkume (1998) explained that teachers, in general, are working in sustained nature of daily schedule, which is the major dissatisfaction of their working condition. It is clear that teachers are usually engaged in different clerical activities like taking students attendance (every day, every period), are preparing summaries and reports, which can be a full day work to many schools. Moreover, many teachers overwhelmingly are occupied by duties and responsibilities of different club and committees. Teachers become stressful when they are dissatisfied in their work as a result of the working conditions that are very taking or that exceed their capabilities Stinnett (Stinnett, 1968). The working conditions which possibly lead teachers to be stressful are, oversized classrooms, disruptive students, teaching time, number of hours devoted to extra instructional activities etc (Stinnett,1970 and Coolabon,2003) All these indicate that stressful teachers are not effectively perform their relevant activities which are required by the profession .

In this study also, teachers in the city have similar time related issues. As indicated by document analysis and responses of principals that the majority of them teach students with class size up to 70 students, even in some schools it can reach up to 80. Majority of the teachers have teaching load from 12 to 20 periods per a week. Regarding the number of tests of the teachers are required to give at least 3 tests and at most 5 tests with in a semester, teachers also required to report at least every month to school officials.

It is obvious that what is stated in the literature part of this paper agrees with the current expected activities of teachers in the city. It seems that most of them are too tight.

Again from test item analysis it has be seen that, tests were not in the position of good quality in terms of item difficulty and discrimination level. All together 117 items were presented for English exam and out of those items 39 of them were defective, also in Mathematics exam 90 items were presented, out of those items 33 of them were defective. This also indicate that about 35% of the total test items were defective

Test in general, and teacher made test in particular should have the direct quality appropriate indices difficulty and discrimination, and appropriate reliability and quantitatively, and in qualitative terms the presence of clear interactions, different item format, clear statements which are free of ambiguity, good arrangement, from simple to complex (Brown,1983).

Similarly Ebel (as cited Adams 1964) has pointed out that weakness of teacher made testes are the following, rely too often of their own subjective judgment, they prepare tests at the last minute of test date, they often trivial details are tested and understanding remembering, creativity and practical application are neglected, despite these limitation, they are used for making critical decisions.

From the test item analysis, which were made in the selected schools about 37% of the total test items seem defective. Some of the items are below the standard of discriminating index. Some of these items among the items are not only below the standard but also discriminates negatively where low scorers perform better than high scorers.

And still the remaining of the defective item is either below or above the standard of difficulty index. That is too difficult or too easy. Reliability of the tests shows high almost in all schools. It indicates that the tests measure similar characteristics of a person.

From the researcher document analysis, it could be learned that the numbers of the test items are not equal through all the three schools. The presence of some non plausible distracters, in multiple choice, and vague items are among common characteristics of the tests. Hence, teacher made test in the city are not different from what is stated in the literature.

CHAPTER FIVE

5. Summary, Conclusion and Recommendation

5.1. Summary

The aim of the study was to assess that quality of the current teacher made test at the 9 grade level, and the factors that affect teachers from post test analysis and evolution.

The type of the study is descriptive survey. To study this, the researcher had raises the following research questions.

- Do teachers made test have the desired quality according to the rules of test constructions?
- Do teachers have the required skills about the usefulness of item analysis and evaluations?
- Do teachers have enough time to make post test analysis and evolution?
- Do teachers' experience and their working conditions have effect on the quality of their tests?

The population of the study was secondary school teachers. In order to get data related to the quality of the current teacher made test, the first semester final test of English and Mathematics were chosen. The test papers and answers sheet, after it had been scored (marked), were taken from the subject teachers and photocopied. Then, based on the responses of the test takers, test item analyses were done for the two subjects in every school totally six subjects. The item analysis focused on difficulty and discrimination index of every item.

Moreover, to see how test or test items are reliable (homogenous) split half method and kader Richard Son KR− 20 formulas were used.

In addition, the tests constructed in every school were examined using checklist. The other data were gathered from teachers by using questionnaire. In each school 20 questionnaires which were distributed and all of them were collected.

In order to secure information about the over all school related data, interviews were conducted with the principals. All the data obtained from every instrument were analyzed and discussed.

5.2 Results

Below are the summary of the major findings:

- From the test item analysis 37% of the total test items were defective. The presence of some non plausible distracters, in multiple choices, and the presence of some non plausible distracters, in multiple choice, and vague items are among common

characteristics of the tests. Hence, teacher made test in the city are not different from what is stated in the literature.

- Teachers' responses on their training conditions indicate that 91% of the teachers have taken training of teaching but in item analysis 68% of them are not practice item analyses after they administer their tests.
- Teachers have problem in time related issues. As indicated by teachers the majority them teach students with class size up to 70, and also have 12-20 periods per a week. Again every teacher has to involve at least 3 clubs in the school.
- Teachers in the city are well experienced, about 85% of them have experience ranging from 6 to 15 years. However, the test item analysis reveals that they are defective.

5.3 Conclusions

Based on the findings the study and review of the related literature, the following conclusions are drawn.

1. No matter how teachers in the sub city are well qualified and have the knowledge and skill of measurement and evaluation, the practice of post test analysis and evaluation seem poor. They don't seem to carry out item analysis and evaluation.
2. Most of the school environments don't seem to encourage teachers to engage in tasks of evaluating and improving the quality of their tests. Due to sustained nature of daily schedule, teachers are too tight to be involved in tiresome activities of the test item analysis.
3. School in general, don't have a practice of analyzing and evaluating test or test item after their administrations.
4. Most of the teachers in the schools are involved in over crowded duties, which are increasing from time to time. This refrain many teachers from committing themselves to their in professional activities. If they are forced to do so, they do it with minimum effort. They hardly think of engaging in activities, which require taking initiative.

5. Considerable numbers of teacher made tests of the schools are defective in terms of difficulty and discrimination indices of the items. Regarding to the reliability of the test, they are marked with better internal consistency. It seems that the test items are internally homogenous, that means they measure similar characteristics of a person.

5.4 Recommendation

1. The quality improving of test items is found to be very important to improve classroom interaction and students academic performance. Since this can be performed by using the empirical procedures of statistical analysis of test items, teachers have to acquaint themselves with the process of item analysis using statistical techniques. As a result planed seminars and workshops have to be organized on this area and experts in psychometrics have to be invited to provide their help acquainting teachers with the empirical procedures of statistical analysis of test items and the principles and procedures of test constructions.

2. Teacher- made test should be constructed to meet the need and interests of all students by providing many opportunities to evaluate what students can do rather than measuring their ability to read, write, and take tests. The tests also should be created before the lessons conducted and approaching to be given. According to the result of the document analysis the tests should have clear instruction and based on the course objective. It should be also arranged appropriately and include various sorts of items such as multiple choices, matching, true /false and so on.

3. Teachers need to understand that constructing and selecting what appears good tests items, with apparently format, at the beginning is not a guarantees for good quality test item, analysis supplemented with post test analysis and evaluation. In order to make this, the school management and department heads need to work hard to encourage teachers conduct the analysis after every test administrations.

4. School principals and other officials should not impose club membership on teachers, so that teachers will have sufficient time to carryout their tasks of analyzing exams effectively as the main part of the professional duties and responsibilities.

References

Adams, G. S. (1964). **Measurement and Evaluation in Educational Psychology and Guidance.** California state collage at Los Angeles.

Ashwoorth, A. E. (1982). **Testing for Continuous Assessment London.** Evans Brothers Ltd.

Bean, K. L. (1953). **Construction of Educational Personnel Tests**. Psychology Department Baylor University. Mc Grew-Hill Book Company, Inc.

Beausay, W. J. (1972). **Outlines and Readings in Educational Tests and Measurement.** Bowling Green State University.

Berdie, R. F. (1962). **A Manual for the Statewide Testing Program of Minnesota.** Association of Minnesota Collages and the students counseling Bureau.

Bertrand, A and Cebula J. P. (1980). **Tests, Measurement and Evaluation.** American international College.

Betere Meshesha (1992**). "Practice and Processes of Classroom Test Testing in Some Senior Secondary School of Addis Ababa"** (unpublished) senior essay AAU.

Bowen, J. D. (1985). **TESOL Technique and Procedure Brounlum:** Rowley Newbury House Publishes.

Brown, F. G. (1981). **Measuring Classroom Achievement.** Lowa State University. Rowley Newbury House Publishes.

Brown, F. G. (1983).3ed. **Principle of Educational and Psychological Testing.** New York. Holt, Rinehart and Winston.

Coolahan, J. (2003). **Attracting, Developing and Retaining Effective Teachers.** Maynooth: National University or Ireland.

Douglees, H. R. (1964).**Modern Administration of Secondary School.** Collage of Education .University of Colorado.

Dull, L.W. (1981). **Supervision School Leadership Handbook.** Columbus. Charles E. Merril Publishing Company.

Ebel, R. L. (1975). **Essentials of Educational Measurement.** (2^{nd} ed.) Englewood clits, Js: - Prentice Hall.

Ebel, R. L. (1982). Proposed Solutions to two Problems of Test Construction**. Journal of Educational Measurement.** 19, (P.125-135)

Gronlund, N.E (1981). **Measurement and Evaluation in Teaching.** (4th ed).New York: Macmillan Publishing Company Inc. London: Collier, Macmillan.

Gronlund, N.E (1985**). Measurement and Evaluation in teaching. (5th** ed), London: Collier, Macmillan.

Hailu Sime, (1990)." **Teachers' Estimations of Test Items Difficulty as an Indicator of their Prior Knowledge of Students' Academic Performance."** (Unpublished Masters Thesis), Addis Ababa University.

Hoge R. D. & Butcheer, R. (1984). Analysis of Teachers Judgments of Pupil Achievement Level. **Journal of Educational Psychology.** 76, (P. 3)

Linn, Robert L. (1983). **Journal of Educational Measurement.** Testing and Instruction: 20, (P. 95-105).

Meherans, W.A (1975). **Standardized Test in Education.** New York: Hot, Rinehart and Winstor.

Mehrens, W. A (1984). **Measurement and Evaluation in Educational and Psychology.** New York Holt, Rinhart and Winston.

MOE (1986) . **"Evaluation Research of the General Education System in Ethiopia."** (ERGSE) Addis Ababa.

Nunnally, J. (1964).**Educational Measurement and Evaluation.** New York: Mc Graw- Hill, Inc.

Ory, J.C and Ryan, K. E. (1993).**Tips for Improving Testing and Grading.** By Sage Publications.

Roid, G. H. and Haladyna T.M (1982). **A Technology for Test-Item writing.** Harcourt Brace Jovanovich polishers Orlando San Diego New York.

Stinnett, T.M (1968). **Professional Problems of Teachers.** (3rd ed,) New York: The Macmillan Co.

Stinnett, T.M (1970). **The Teacher Dropout.** ILLINOIS. F.E Peacock publisher, Inc.

Teka Wolie (1988). **"The Role of Measurement and Evaluation in Making Educational Decisions."** (Mimeographed), Addis Ababa: Kotabe University Collage of Teachers Educations.

Thorndike, R. L. and Hagen, E. P. (1965) .**Measurement and Evaluation in Psychology and Education** (2nd ed.)New York: Jhanwiley and Sons, Inc.

Thorndike, R. L. (1971). **Educational Measurement.** (2nd ed) Washington, D.C. American council on Evaluation.

Thorndike, R. L. (1977). **Educational Measurement.** Washington: America councilor Education.

Zurawski, R. M. (2001). **" Making the Most of Examinations Procedures for item analysis."** St, Norbert Collage. Retrieved from htt://WWW. Featured. Article htm.

•••••• ••••• (1986) • •••••• •• •••• ••••• •••••• ••••• /••• •••/•• ••• •••• •.•.•.•.•

••••••• ••••• (1986) • ••••••• •• •••• ••••• ••••••• •••••/••• •••/•• •••
•••• •.•.•.•.•

Appendix 1

Checklists for English and Mathematics Examinations

Make√ in the given space, where Y= yes, N= no and S = school

Checklist	S1 Y N	S2 Y N	S3 Y N
1. Stating time allotted	≠	≠	
2. Clearly stated instruction	≠		
3. Clarity and neatness of the test paper			
4. Clearly indicate how to respond			
5. Items are Independent			
In multiple choice			
6. Economical use of "All of the above" and "None of the above"			
7. Avoid placing correct alternative in pattern			
8. Keeping the length of the option uniform			
9. Avoiding negatives specially double negative			
10. Avoiding grammatical clue			
11. Not repeating the keywords form item in the options.			
In true False	=		=
12. use of important content and avoid trivial statements			
13. Use more false items than true items			
14. Avoid specific determines like always, never, all.			
15. Avoid ambiguous statements			
In Matching			
16. Homogeneity of the items			
17. The no of item in the premise side and the response site should be un equal			
18. clearly state the basis form marching the premise with the responses			
19. keeping items brief, limiting the list of premise and when possible short phrases or single word			
Fill in the blank or short answer			
20. single word answer preferable			
21. placed blanks at the end of the statement			
22. do not impair the sense of the statement			

4.1.2.2 for Mathematics Test

Make√ in the box given, where Y= yes, N= no and S = school

Guidelines for Constriction of Tests	S1 Y N	S2 Y N	S3 Y N
1. Stating time allotted	≠		≠
2. Clearly stated instruction	≠	≠	
3. Clarity and neatness of the test paper			
4. Clearly indicate how to respond			
5. Items are Independent			
In multiple choice			
6. Economical use of "All of the above" and "None of the above"			
7. Avoid placing correct alternative in pattern			
8. Keeping the length of the option uniform			
9. Avoiding negatives specially double negative			
10. Avoiding grammatical clue			
11. Not repeating the keywords form item in the options.			
In true False	=	=	
12. use of important content and avoid trivial statements			
13. Use more false items than true items			
14. Avoid specific determines like always, never, all.			
15. Avoid ambiguous statements			
In Matching			=
16. Homogeneity of the items			
17. The no of item in the premise side and the response site should be un equal			
18. clearly state the basis form marching the premise with the responses	≠	≠	
19. keeping items brief, limiting the list of premise and when possible short phrases or single word			
Fill in the blank or short answer	=		=
20. Single word answer preferable			
21. Placed blanks at the end of the statement			
22. Do not impair the sense of the statement			

N.B = Refers to the item format not includes in the test

≠ Refers to the quality of the test or test item lacks

Appendix 2

Addis Ababa University

Department of Educational Psychology

This questionnaire is designs to get relevant information regarding exam related issues from some selected high school teachers.

Thank you in advance for your cooperation in giving me your responses genuinely.

Instruction: Make the √ or X mark in the box given, and specify your response wherever needed.

No Need to write your name

1. Sex Male ☐ Female ☐
2. Age _____
3. Years of experience as a teacher _____
4. The subject you are teaching _____
5. Qualification
   a. Diploma ☐ b. Bachelor Degree ☐ c. Mastered and above ☐
6. Have you taken teacher training courses?
   a. Yes ☐ b. No ☐

   6.1. If no, have you had a chance of any training in measurement and evaluation?
   a. Yes ☐ b. No ☐

   6.2. If yes, did you have a course on measurement and evaluation in your collage or university education?
   a. Yes ☐ b. No ☐

   6.3. If yes list some of them _____
7. Do you practice the skill of educational measurement in sufficient detail in your teaching?
   a. Yes ☐ b. No ☐
8. If no state your main reasons _____
9. Do you write more test items than are needed and decide on the basis of preliminary tryout, which items should be revised or eliminated?
   a. Yes ☐ b. No ☐
10. Do you often place too much emphasis on details on your exam
   a. Yes ☐ b. No ☐

11. Do you construct test items bit by bit through a long period of time?

a. Yes ☐ b. No ☐

12. Do you prepare your test just a few days before the test date?

a. Yes ☐ b. No ☐

13. Do you use text book statements to construct your exam? so as to minimize grammatical errors

a. Yes ☐ b. No ☐

14. How often do you use table of specification when you prepare your exam"

a. Always ☐ b. Usually ☐ c. Sometimes d. ☐ver ☐

15. Do you give chances to your fellow teachers to comment your test?

a. Yes ☐ b. No ☐

16. Do you often conclude that students did not learn the material when they can't correctly answer the questions?

a. Yes ☐ b. No ☐

If your answer is No. would you pleas list some of your reasons

_____

17. How often do you supplement newly constructed test item in your exam?

a. Always ☐ b. usually ☐ c. sometimes d☐ver ☐

18. When do you usually prepare your test?

a. Before the lesson is given ☐

b. During the lesson ☐

c. When the testing is approaching ☐

d. Immediately before the day of the test ☐

19. Have you ever evaluated your test or test items after the test are given to the students?

a. Yes ☐ b. No ☐

20. If yes how often?

a. Always ☐ b. Sometimes ☐ c. One time ☐

d. Other specify _____

21. How often are you required to give information regarding your students a result to your immediate school officials?

a. Every week b. Every month ☐ c. Every semester ☐

22. What must be done by the school management to in force the evaluation of assessment procedures particularly the use of tests? _____

23. Is there any regulation and guidelines concerning to test construction and administration in your school?

    a. Yes ☐ b. No ☐

24. What are the main constraints to evaluate your test items after you have administered them? List them from highest to lowest.

Appendix 3

Interview Questions Directed To Principals

1. What do you think about your school teachers made tests usage and quality?
2. Is there any manual prepared by the ministry of education regarding test (exam) related issues to be used in the school?
3. Is there examination and evaluation committee in your school?
4. What are the members of the committee? What are the criteria to be the member of the committee?
5. What is expected form the committee regarding teacher made test (exam)?
6. Does the committee report to teachers how effective the tests they have prepared are?
7. Do you think teachers have enough time, skill and knowledge to analyze and evaluate their tests?
8. Are there different clubs in your school? How is the involvement of teachers?

Declaration

I, the undersigned, declare that this thesis is my work and that all sources of materials used for the thesis have been duly acknowledged.

Name _____

Signature_____

Date of Submission_____

This thesis has been submitted for the school of Graduate Studies with my approval as University advisor.

Name_____

Signature_____

Date_____