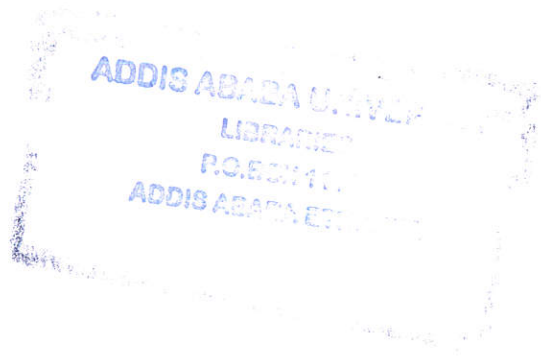




ADDIS ABABA UNIVERSITY
SCHOOL OF GRADUATE STUDIES

**Assessment of Teacher Made Tests and Factors
Affecting Secondary School Teachers from
Analyzing and Evaluating their Tests**

BY
KASSA AMARE



MARCH 2007

Assessment of Teacher Made Tests and Factors Affecting Secondary School Teachers from Analyzing and Evaluating their Tests

**A THESIS PRESENTED TO THE SCHOOL OF GRADUATE STUDIES
ADDIS ABABA UNIVERSITY
IN PARTIAL FULFILMENT OF THE REQUIREMENT FOR THE
DEGREE OF
MASTERS OF ARTS IN EDUCATIONAL PSYCHOLOGY**

**BY
KASSA AMARE**



MARCH 2007

THIS THESIS HAS BEEN SUBMITTED FOR EXAMINATION WITH MY

APPROVAL AS UNIVERSITY ADVISOR

NAME Ayle Mesherka

SIGNATURE Ayle Mesherka

DATE March 27, 2007

ADDIS ABABA UNIVERSITY
SCHOOL OF GRADUATE STUDIES

Assessment of Teacher Made Tests and Factors
Affecting Secondary School Teachers from
Analyzing and Evaluating their Tests

BY

KASSA AMARE

FACULTY OF EDUCATION

Approved by the Examination Board _____

Chairman, Department of Graduate Committee _____

Advisor _____

Examiner, External _____

Examiner, Internal _____ (Dr. P. MOHAN RAJU)

ACKNOWLEDGEMENT

I deeply appreciate the support of my advisor, Dr. Ayele Meshesha, for his follow up, technical advice and invaluable comments and corrections.

I am grateful to Directors Ato Kifle Moges, Ato Tewodros Kelemework, Ato Wolde Michael Bonga, Ato Biruk Bekele and Ato Belay for assisting me in collecting relevant data from their schools.

I am also grateful to Ato Muareta Adeba for his encouragement and cooperation in sharing my responsibilities in office work throughout my study.

I am indebted to W/ro Fikirete H/ Mariam and Ato Solomon Amare for diligently editing and proofreading the whole paper.

Finally, my sincere and heart felt gratitude to my wife, W/ro Aster Negash, who supported and encouraged me to complete my study.

Table of Contents

Contents	Page
Acknowledgement.....	ii
Table of contents.....	iii
List of Table	v
Abstract.....	vi
The Aim of Study	vii
Objective of the study.....	vii
 CHAPTER ONE	
1. Introduction.....	1
1.1. Background of the Study.....	1
1.2 Statement of the Problem	3
1.3 Significance of the Study	4
1.4 Delimitation of the Study	4
1.5 Limitation of the Study	5
1.6 Definition of the Study	6
1.7. Organization of the Study	7
 CHAPTER TWO	
2. Review of Related Literature.....	8
2.1 Usage, Importance and how to	
Improve The Quality of teacher-made test	8
2.1.1 Usage and Importance of teacher –made test	8
2.1.2 Some consideration in test Evaluation	9
2.1.3 Test Item Formats & Guideline.....	11
2.1.3.1. Multiple – choice item formats	12
2.1.3.2. True- false Item formats.....	13
2.1.3.3. Matching Item formats	14
2.1.3.4. Supply Item or short answer formats	14
2.1.4 Quantitative Description of tests	15

2.1.4.1. Reliability	15
2.1.4.2. Validity	15
2.1.4.3. Item Analysis.....	16
2.2. The Role of Teachers in Improving Quality of their Tests.....	16
2.3. School resource and Working Climate	20
2.3.1. Class Size.....	21
2.3.2. Teaching Load.....	21
2.3.3. The Amount of Clerical Work Required.....	22
2.3.4. The Workweek.....	22
2.3.5. Free Period for Teachers	23
2.3.6. Too Many Extra Curricular Duties.....	23
2.3.7. Other Factors.....	24
2.4. Implementation of Teacher-Made Test in Ethiopia	25
2.5. Summary of Review of Related Literatures.....	27
 CHAPTER THREE	
3. RESEARCH DESIGN, METHODS, PROCEDURES AND SOURCE OF DATA	30
 CHAPTER FOUR	
4. DATA PRESENTATION AND ANALYSIS.....	33
4.1. The Current Teachers Made Tests Quality Assessment	33
4.2. Data Obtained from Teachers	41
4.3. The Response of the Principals with the Interview Held with them	50
 CHAPTER FIVE	
5. DISSCUSSION	54
 CHAPTER SIX	
6. SUMMARY, CONCLUSION & RECOMMEDATIONS.....	59
REFERENCES	64
APPENDIX 1.....	68
APPENDIX 2.....	76
APPENDIX 3	92

LIST OF TABLES

Table	Contents	Page
Table I	Reliability of the tests	35
Table II	Correlation of teacher made test of different Schools National Exam score.....	36
Table III	Personal characteristic of Teachers.....	41
Table IV	Teacher Training background.....	43
Table V .A.	Practices of construction and evaluation of test or test items	45
Table V.B .	Practices of construction and evaluation of test or test items.....	47
Table VI	School based factor for teachers.....	49
.	.	.

Abstract

The purpose of this study was to determine to what extent teacher-made tests are effective and to identify the factors that inhibit high school teachers of Addis Ababa from analyzing and evaluating their exams after they have administered them. To check the quality of the test, two tests of final exams from five sample schools were taken. This helped to check the quality of the test items from the response of the test takers by computing difficulty and discrimination indices. To see the internal consistency of the tests, KuderRichardson KR-20, and Spearman Brown, split -half methods were employed.

Moreover, qualitative description of the test was made by observing the test papers based on the principles of test construction. In addition, the correlation of average teacher-made tests of students' score and the national exam score of the students were computed.

By using questionnaire relevant data from teachers, and by conducting interview with principals about the overall condition of the schools data, that can limit teachers from making posttest analysis, were secured.

The finding demonstrated that test qualities currently are not as such good in terms of their difficulty level, discrimination power and internal consistency. The qualitative data analyses have also indicated that most of the tests are not constructed as per the principle of test construction. Teachers are also found not in a position to make posttest analysis. To minimize the problems associated test or test items evaluation the need to encourage teachers to engage in such practice is highly recommended.

The Aim of the Study

The aim of the study is to investigate teachers' problems that hinder them from making analysis and evaluation of their exams after they have administered them. The researcher believes that knowing the quality of teachers made tests helps him to underline or stress the essentiality of the study. Therefore, current test quality of high schools is described qualitatively and quantitatively in some details.

Objectives of the Study

This research is believed to serve the following purposes

1. To create awareness about the importance of making posttest analysis.
2. To identify the problems that hinder teachers from analysis of posttest administration.
3. To indicate the current level of test quality and devise remedial measures
4. To examine the extent to which our teacher made tests results are erroneous.
5. To show the school with relatively better teachers made tests, and to reveal the wanted experience of that school.
6. To be a stepping-stone and give information for other researchers who intend to make further study.

CHAPTER ONE

1. INTRODUCTION

1.1. Background

The quality of the test without considering the teachers' competency and school-based factors is meaningless. Therefore, the background study of test quality and importance, and teachers' possible barriers not to practice test analysis and evaluation are reviewed below.

It is known that the main goal of education, for a given country, is to empower citizens, by building their cognitive, affective and psychomotor skills so they can solve their problems. To this end, the roles of qualified and experienced teachers seem highly essential. However with the current trend of rising number of teachers who drop out of their profession, the success of attaining the goal seems highly challenged (NOE, 2004:9).

Based on the type and the level of empowerment required examiners, and other concerned parties make selection and placement for different occupations or positions. The tools or instruments often used to determine the level and the type of empowerment are teacher made tests or examinations.

Thus, success in selecting the right (qualified) personnel for the right position seems to greatly rely on the implementation of quality instruments which measure efficiently the required skill or capability. Not placing the right person at the right place would results in crises throughout the nation. This affects negatively the prosperity of a nation and extends backwardness by devastating the country's resources. Therefore, great attention has to be given to the quality of measuring instrument, examination or teacher- made tests.

If they are implemented without giving due attention to improvement of quality of tests in general and in particular to teacher-made tests,

undesirable effects of tests will be manifested. Hence, these effects according to Keeves (1994):

- a) The effectiveness of instruction may be incorrectly evaluated,*
- b) Students and teachers may be wrongly informed about
the result of assessment,*
- c) Students motivation to learn may be reduced,*
- d) Teachers motivation to teach may be reduced,*
- e) Students may be incorrectly placed in the course of study,*
- f) Students may not be appropriately rewarded for their learning,*
- h) The teachers may not have sufficient information on which
to base effective teaching.*

Another important task in improving tests quality, which is worth noting, is the importance of evaluating our tests with some intervals. In line with this (Jurs and Wiersma, (1985: 289-220) explained that “programs and objectives change; tests that were appropriate at one time may no longer be relevant, or they may not be the best tests available. Teachers should be involved in the review of any test related to instrumental programs.”

Therefore, reviewing and evaluating tests is not a onetime activity; it is rather an on going process that should be done regularly by test makers in general and by teachers in particular. i.e. a test item which is best at one time may become irrelevant at another time, especially when it contains obsolete information.

1.2. Statement of the Problem

The most widely used tests today are teacher-made achievement tests. Such tests are designed to measure the extent of present knowledge in the students of a particular school or section. As discussed in the proceeding section, if we want our tests to serve the desired purpose, they have to be assessed, analyzed, evaluated and reviewed continuously. It is through this process that we can ensure the efficiency of our test in measuring what they are intended to measure. The report of MOE (1986:23) states that, classroom tests are found to be 52.2 percent dependable on teachers' subjective judgment, while 47.8 percent are reliable. This indicates that teacher-made tests have considerable level of weaknesses. Educators should show great concern regarding this problem and devise ways of improving them.

In the case of teacher made tests, it is the teacher himself who should be responsible for improving his tests and /or items by making post-test analysis, i.e., by discarding the weak items and by retaining the competent items in his item bank.

However as observed by the researcher in his long years of experience as a teacher, most teachers do not seem to analyze their tests. Almost all are not practicing the process of analyzing tests so as to learn about their weakness and strengths. Even the researcher himself came to realize this problem and endeavored to make further investigation on the subject after taking some courses on measurement and evaluation.

Besides, what initiated this research project is absence of adequate local research to my knowledge that investigated the extent of deviation of our exams from what is called the ideal instrument. Moreover, making a survey of the barriers, which deter teachers, from making analysis of their exam after they have administered them is the main concern of this research.

Research questions

The following are specific research questions to guide the research.

- Do teachers have awareness and knowledge about the usefulness of item analysis & evaluation?
- Do teachers have enough time to make posttest analysis and evaluation?
- Do our teacher-made tests have the desired quality according to the rules of test construction and test item analysis?
- Do teachers have similar test related problems in all schools regardless of the type of ownership?

1.3. Significance of the Study

- To communicate to teachers training institutes or colleges so that they could include and give more attention to educational measurement and evaluation courses in their curriculum.
- To give highlight to the concerned parties so that they can contribute to the betterment of schools working environment.
- To contribute to the country's poverty reduction by increasing productivity for it places the right person at the right place when proper instrument (test) is used

1.4. Delimitation of the Study

The study is delimited to the ninth grade level in the academic year of 2005/2006 (1998E.C).

Because of limitation of resources, only five schools were taken as representatives of the secondary schools in the city.

The names of the schools included in the study are Shimelis Habte, Abyot Kirs, Delachen, Lideta catholic Cathedral and Assai. Three of them, Shimelis Habte, Abyot Kirs and Delachen are government Secondary Schools. And the remaining two are non-government Secondary Schools.

From each school, one 9th grade were selected randomly, and exams in two subjects' were under study. The subjects were English and mathematics.

In addition, teachers of the selected secondary schools and their principals were made participants of the study.

Although there are other factors that contribute positively to the problem, the researcher restricted him to give solution from the teachers and the principals' perspective. The knowledge of teachers in the subject area, whether the school working environment is conducive, whether they have enough time, whether they have taken courses related to measurement and evaluation are assessed.

1.5. Limitation of the Study

1. The tests used in each school have different item formats, and different teachers prepare them too. Having these constraints, the Pearson product moment correlation coefficient is used to indicate reliability and predictive validity of the tests.
2. The correlation coefficient computed has some shortcomings. The variables taken are: 8th National Examination scores (which is the average score of English, Mathematics, Amharic, Physics, Chemistry, Biology, Geography & History, Civics and Ethical Education) of particular group of students and their 9th grade first semester average score of English, Mathematics, Amharic, Physics, Chemistry, Biology, Geography, History, Civics and Ethical Education, and Physical

Education. The 8th grade National Examination scores are the average scores of eight subjects, where as the 9th grade first semester, which are taken into consideration, are the average scores of ten subjects. Had it been possible to deal with similar subjects in both National and first semester test scores, the study would have been more sensible. But this can give some high light about the relation of two scores.

3. The test item qualities are analyzed by observing the exam and by making test item analysis. The test item quality in terms of its content is not considered, as it requires being the subject specialist.

1.6. Definition of Terms

Evaluation:- is a process by which the effects and the effectiveness of teaching courses are determined (frith and Macintosh, 1984:253)

Measurement:- is the process of assigning numbers to the individual number of set of objects or person for the purpose of indicating differences among them in the degree to which they possess the characteristics being measured (Ebel, 1979:377)

Test:- is any kind of device or procedure for measuring ability, achievement, interest and other traits (Ebel, 1979:380)

Test Item:- is the smallest independent unit of test (Ebel,1979:380)

Reliability:- is the consistency, accuracy or suitability of the results of a given test that has been repeated two or three times with the same population (Ogunnivi, 1984:147)

Validity:- is the extent to which a test measures what its is supposed to measure (Oguinnivi, 1984:147)

Item analysis: - is a set of procedures for determining item discrimination and difficulty, the validity, and the reliability of each test item in a test (Gall, 1996:762)

Testing:- the selection, administration, scoring and interpretation of tests given to measure achievement, award credit, select applicants for admission, and provide information for placement. The Encyclopedia of Higher Education (1977).

1.7. Organization of the Study

The study is composed of six chapters, the first chapter deals with the problem and its approach; the second deals with the review of related literature; where as the third chapter contains research design, method, procedures and source of data. The fourth chapter contains data presentation and analysis; the fifth chapter contains discussion, and sixth summery conclusion and recommendation of the study.

CHAPTER TWO

2. REVIEW OF RELATED LITERATURE

In this section an attempt will be made to examine the work of others in topics related to tests in general, and teacher-made tests in particular.

2.1. Usage, Importance And How To Improve The Quality Of Teacher Made Tests

2.1.1. Usage and Importance of Teacher Made Test

According to Brown (1970:2), “a test is a systematic procedure for measuring a sample of an individual’s behavior”. It is administered and recorded according to certain prescribed rules.

He further tells us that a test is used as an instrument to measure the trait with in the individual and aids in decision making for different purposes, i.e., selection, placement, diagnosis, hypothesis building, etc.” Brown (1970:9).

Proper decision-making is possible if the quality of the test (instrument) could give the desired out put. Whenever the tools or instruments are defective, the decision we make for placement, selection, diagnosis and others would be defective also. Hence, we should bother on the quality of the test. This idea is supported by Thorndike (1977:1-13), which says that giving much emphasis and concern about the quality of test and their wise use has been major a function in educational and psychological measurements.

The most widely given tests from kindergarten school to college or University level are teachers-made tests. Many formative and summative tests are given at all the levels. According to Frey (2005) about 54 teacher-made tests are given in a typical classroom per year worldwide. As a result of this,

much more decision making in placing, selecting and grading rely mainly on teacher-made test than tests constructed by other organization or standardized tests.

In order to keep the quality of the test, if it is a good one, or improve the quality from time to time, the tests given have to be analyzed and evaluated. As stated by Wiersma (1985:289-290):“Tests that are used regularly should be reviewed and evaluated from time to time. Programs and objectives change; tests that were appropriate at one time may no longer be relevant; or they may not be the best test available. Teachers should be involved in the review of any test related to instructional program.”

Similarly, the importance of analyzing tests as explained by Sax (1974:231) is to see ambiguous items, miss keyed items, too easy or too difficult, and non-discriminating items. Madsen (1983:178) has also stated that evaluating exams (tests) can make students use extra effort to understand the subject, and develop positive attitude, motivation and performance. In the contrary, not evaluating the tests might result in adverse effect on the learners.

The time of analyzing the test given is very important. If the tests are analyzed after the tests results are recorded, it penalizes the learners. Therefore, to avoid or discard the bad questions, it is necessary to analyze before the test results are recorded Madsen (1983:178).

2.1.2. Some Considerations in Test Evaluation

The primary points to be considered in evaluating tests or examinations are primarily by referring to simple statistical evaluation. This helps according to Madsen (1983:199) to check on the quality of teacher-made tests and to gain reasonable and accurate information from the assessment instrument. Based on simple statistics roughly we can get information on central tendency

(mean, median & mode) of the test score. The spread of the mean in the given class can help to evaluate the suitability of the test and examination prepared.

The other aspect of evaluation, which is used in evaluating individual item of a test or exam, is item analysis. It is a systematic procedure to evaluate how good or appropriate a test or an examination is, and it produces information on how each question, item or task functions in the whole test (Bowen, 1985:388-389).

Regarding the usage of too difficult items and too easy items by: an instructor may use very easy items that can reflect the fact that some relatively straightforward concepts were taught well and mastered by all students. Similarly, an instructor may choose to include some very difficult items on an exam to challenge even the best-prepared students. Therefore, to use too easy and too difficult items in tests the items difficulty level needs to be known in advance (Zurawski, 2001).

In conducting item analysis, there are guidelines that are forwarded by different scholars. Difficulty indices between 30% to 40% are hard items, and between 85% to 90% are easy items. This means below 30% are very difficult and above 90% are too easy they need to be rejected (Thorndike, 1977: 216). Indices of discrimination below 19% are poor items they need to be rejected (Eble, 1979:267). As noted by (Bowen, 1985:388-389; Zurawski 2001 & Taiwo, 1995:59-61), if 90% of the respondents rightly answered an item, then it is too easy; if fewer than 30% answered the item correctly, then it is too difficult. They also stated that the index of discrimination 15% and over is acceptable.

The remaining and the most important quality that have to be evaluated are test validity, reliability and usability. Reliability refers to the consistency of the test or the ability of the test in giving the same or similar outcome when applied to a particular group at different times. Validity refers to the ability of the test to measure what is intended to measure i.e, a maths test should

measure maths ability but not reading ability. As indicated by Frey (2005) a good classroom test is valid and also reliable. Usability of a test refers to simplicity of the test during scoring.

Different test item formats are used for measuring different sample of behaviors. Thus, based on the intention of the test maker (teacher), i.e. different item formats can be used. Teachers may want to measure knowledge, recall, skill or ability. According to the view of Frey (2005) multiple choice, matching and true-false item formats are used to measure knowledge. Such item formats are used to recognize the correct response from several presented alternatives (Brown, 1970:264).

Items like short answer, completion of items and essay question are often used to measure recall (Brown, 1970:264).

Other item formats are forwarded by Brown (1970:264) like playing instruments, baking, driving, painting a picture, etc. are used to measure skill or ability.

2.1.3. Test Item Formats and Guidelines

The test constructor must always be concerned with item formats, the way in which the items are presented. And the formats are also highly dictated by the purpose of the test constructor wants to measure or assess Brown (1970:32-33). Related to the above view, for measuring simple learning outcome of knowledge, multiple choices, matching and true false items are recommended. For measuring the recall of respondents, multiple choices, matching and true -false item are often used. And the recall item types are recommended for higher learning outcome. This may not always true, as knowledge (recognition) items can be constructed to measure higher learning outcome (Frey, 2005).

Test item formats are classified mainly in to two: essay type and objective type (Mehrens, 1984:61). Essay type items, range from short answer items of about half a page to discussions or extended responses of several pages. Since these item formats are not used at the grade level we discuss, we show more concern to the objective type of items, i.e, short answer, supply item (single word & phrase and even multiple words or phrases), true-false, multiple-choice, and Matching items.

2.1.3.1. Multiple-choice Item Formats

Multiple choices are one of the test item formats that contain a stem and several alternatives. The stem part presents the item in a question or in incomplete sentence form. And it is followed by alternatives, which are possible answers to the question presented. The multiple choices described by Brown (1981:36) as follows:

A multiple -choice item consists of a stem and a number of possible responses. The stem may be either an incomplete sentence or a question; in either case its function is to state the problem presented by the item. If the stem is an incomplete sentence, the students' task is to select the alternative that best completes the statement. If the stem is a question, the alternatives will be possible answers to the question. Students are instructed to select the correct, or best, alternative.

The stem of a multiple-choice question should present the problem without any ambiguity and the alternatives should be plausible.

The characteristics of good multiple choice items are according to Brown (1970: 267-68) are "without doubt difficult and time consuming to construct; therefore, the time saved in scoring the test is generally balanced by the added time needed to construct the item." It means constructing effective multiple choice is not an easy task. The test constructor has to analyze many factors.

Actually scoring of the test is very rapid and also objective.

To construct good multiple-choice items, the following guidelines are stated by Clark and Starr (1991:445):

- Avoid grammatical clues
- Keep the option length uniform
- Don't repeat key word from the stem in the option.
- Use plausible distracters.
- Avoid use of negative when it is essential (specially double negative)
- Use of "none of the above" and "all of the above" rarely.

Clark and Starr also indicate that these items are relatively free from guessing. They are also more versatile as they measures both simple and complex learning outcome. They are also useful if we call for best answer than correct answer.

2.1.3.2. True-False Item Format

True-false test item formats are items that are presented in declarative statements, and students are required to judge by responding as true or false, based on what they have learned (Brown, 1981:46).

As it is indicated by Wiersma, (1985:124), these test items are reliably scored, and allow adequate content sampling. They are also easily constructed and scored (Brown, 1970:267).

He also states the following guidelines:

- Make the statement clearly. (Free of ambiguity)
- Avoid lifting statement from textbooks
- Avoid specific determiners like always, never, all and none of the above.

- Avoid trick questions.
- Do not limit question to memorize facts
- Make the length and number of true-false statement about the same.

2.1.3.3. Matching Item Format

Matching items have the following characteristics, they have two columns, the first column (that is usually on the left side) is called premises and other column is called responses. The premises consist of the item stem and the response side consists of the available options (Weirsma, 1985:141).

Matching and multiple-choice items have some similarities especially when they cover homogeneous contents. As indicated by Brown (1970: 268) “in essence, a matching item is a series of multiple choice items, covering a homogeneous content domain with the same set of responses serving all items”

The column in the premise and the response side should be unequal. The statement or the elements in the response side should be more than elements in the premises side. The statement or element in the response part may be chosen once, more than once, or not at all (Clark and Starr, 1991:447).

2.1.3.4. Supply Item or Short Answer Formats

Short item formats are items that make students respond to a word, a phrase, a sentence or two, a symbol or a number Brown (1981:58).

Guidelines as indicated by Mehrens (1984:135)

- For computational problems, the teacher should specify the degree of precision and unit of expression expected in the answer.
- Have the blanks occur near the end of the sentence

- Avoid excessive blanks in a single item
- All the blanks should be equal regardless of the length of the answer.

2.1.4. Quantitative Description of Tests

2.1.4. 1. Reliability

For this study, the split-half method (Spearman-Brown formula) and KuderRichardson formula is chosen (even though there are other methods of finding tests reliability). Here a test is given for one group, in one administration.

For split half method, Spearman-Brown formula is used to find reliability after getting the correlation of even and odd items (Gronlund, 1985:95).

The other, after one gets the proportion of the pupils' responding correctly or the difficulty index for each item, is Kuder-Richardson KR-20 formula. It is used also to calculate internal consistency (Gronlund, 1985:95).

2.1.4. 2. Validity

The validity of a test can be found using statistical approach apart from curriculum approach. This is possible, specially for calculating predictive and concurrent validity, by taking the correlation between test scores with other criteria test score, provided the tests are given to the same group of students.

Predictive validity can be found between the test score and the score of some measure representing identified criteria. According to Wiersma (1985:103), there must be an intervening period from three months to six months between the first test and the second test. The correlation coefficient between the tests can be computed by using Pearson product moment correlation formula, and the value of correlation coefficient as it is indicated

by Bertrand (1980:138):

$\pm 0.7 - \pm 1.0$ high and very high

$\pm 0.4 - \pm 0.7$ average to high

$\pm 0.2 - \pm 0.4$ present but low

$\pm 0.0 - \pm 0.2$ negligible to very low

2.1.4. 3. Item Analysis

Item analysis helps to check internal validity, which tells the percentage of high achievers and the percentage of low achievers who answered each item correctly. In addition, it gives information on the difficulty level of each item, and discrimination power of each item, among high achievers and low achievers. If higher percentage of high achievers responded correctly than low achievers, then it shows positive discrimination (Grean, 1963: 83).

To determine the index of difficulty and discrimination, as indicated by Meherns (1984:191), we take only test scores of 27% from higher achiever and 27% from lower achievers.

Thus, by taking into consideration the two groups, index of difficulty and discrimination of every item is computed.

2.2. The Role of Teachers in Improving Quality of Their Tests

Teachers activity cannot be separated from exam related activities; they most of the time engage in the routine and cyclical activities of the testing

process. As stated by Chase and Stiggin (cited in Frey 1994-2005) classroom assessment is an integral part of teaching and it occupies more than one-third of the teachers' professional time.

Since learning is a natural process, it can take place whether we like it or not, whether we are ready or not. It can take place without expending any additional effort or without directing any effort to learn. But in educational setting special effort is directed to make learning more effective and efficient and less cost. If special effort is applied in schools setting to know to what extent learning has taken, teachers should design (test) to measure it (Ebel, 1975:73).

We can measure the achievement test itself, i.e., to what extent the intended behavior is measured and predicted by that test. The classroom teacher in addition to being the master of knowledge and skill he has to be the master of "the art of testing". This means not only the skills in quantitative analysis but also qualitative should have to be considered. This idea is forwarded by Gronlund (1981: 87-88) "such factor as unclear direction, difficult and complex terms or sentences, ambiguous statement, poorly constructed test items and improper arrangement of items etc. are supposed to be impediments that lower the validity of tests."

Teacher made tests have some common characteristics. As it is forwarded by Mehrens (1969:135) "they are sometimes of inferior quality as they are constructed hastily, unreviewed and untested." Such tests as they are constructed by a teacher or teachers alone have the following limitation: shortage of time to be devoted for construction and to examine the test and/or the test items, inability of a teacher to tryout his test before hand. Because of this, the tests are not technically as sound as commercial (standardized) tests (Mehrens, 1969:135).

Teacher-made tests are blamed for their poor quality in various respects. As noted by Ashworth (as cited in Betre (1992:31))

- *They are highly influenced by the subjective judgment of the teacher.*
- *They are prepared often in the last minutes as a result they are unreliable and poor in most respect.*
- *They are short and inadequately sample the whole area.*
- *Only trivial details are tested instead of testing understanding, reasoning, creativity, and practical application are neglected.*
- *Ambiguity questions are not discovered until marking stages.*

Even if teacher-made tests have many limitations, practically, they are used for making critical decision. On the basis of the test more and other relevant (pertinent) data such as teachers' estimate, health report, grade status, and the likes, several pupils are given trivial promotion to the next higher grades.

Teacher-made tests importance is not only for classroom teachers, but also for principals, school counselors or clinical psychologists in vocational and educational guidance. The tests serve as a guideline in planning the students future program (Educational or vocational). By using only teacher-made tests without considering other information about one's interest, aptitude and altitude, counselors or clinical psychologist cannot make the best possible decision regarding future program (Mehrens, 1969:192)

Hence, in order to make teacher-made achievement tests meet the desired purpose; the teacher/s must examine his test, and evaluate the test outcomes. It is only when there are frequent evaluations that it would be possible either to improve the test quality or to identify the marked weakness in the learning areas. In line with this to Green (1963: 83) states, "It is thus

through a close working relationship between teaching and measurement that most effective instruction is attained”

Conducting analysis of the test after scoring it is also believed to have immense significance to learn about its quality and thereby improve it. In connection with this point Brown (1970:23) has the following to say:

“After the test has been administered and scored, the teacher may want to analyze the test scores to see how effective a measuring instrument he has constructed. Although in practice only a few conscientious teachers make systematic analysis of their tests, the feedback provided by such analysis will undoubtedly help the teacher improve his test construction skill and provide information regarding the effectiveness of his teaching and learning of students.”

Teacher-made achievement test and standardized achievement tests have some marked differences. Teacher-made tests cover adaptable topics (a unit, a chapter with in a week or month) for a classroom situation. Standardized test cover more topics, and they usually have higher technical (psychometric) qualities than teachers made tests. As the teacher-made test is more applicable to a particular classroom situation, it is more valid than standardized instrument. This idea as confirmed by Green (1963: 87) “a well constructed teacher-made test can be extremely valid. Furthermore, adaptation of these tests to the teachers’ objective may make them even more valid for this purpose than may be possible with standardized instruments.”

The knowledge and the skill of test construction is very important and helpful for teachers. Teachers should not feel inferior by comparing their test with standardized tests that are prepared and tested by organized bodies that are skillful too. Even these tests are not error free or perfect let alone over burdened teacher-made test.

2.3. School Resource and Working Climate

Normally not only teachers but also other professionals and non-professional workers leave their profession or occupation when the working climate (environment) is not conducive. Actually those who manage to work there may not be effective as much as it is needed. As it is put by AACTE (1969:21) teachers leave their job in high rate when the school based factors and environmental factor are not conducive, let alone engaging in tiresome task that require initiation and creativity. Regarding less motivated and non-voluntary teachers as it is forward by Stennett (1968: 23) probably teachers themselves cannot honestly pin point one reason for their decision to teach. It is likely that there are indistinguishable reasons for each decision. Nevertheless, it is widely believed on sociological ground that individuals are even teaching with ulterior motives is unlikely to remain in the teaching.

From this we can understand that as long as teachers enter into the profession with out their interest, let alone performing the tiresome activity they do not want to stay even in the profession. This situation seems affecting constraint to be investigated.

Concerning the school climate James put his idea (as cited in Horn (2003:1002) as follows: "It is an important and main component of successful school and also a major component in predicting the learners' future success." Therefore, for a 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 fascinate (motivate) or frustrate the learners and teacher, and their competencies.

For our purpose, by leaving behind the social, economic and political factors, we restrict our assessment on the following possible factors. The class size, teaching load, the amount of clerical work the teacher involved in, the

work week, free periods for teachers, too many extra-curricular activities, the school physical setting and teachers perception and attitude.

With the increasing number of students and educational institutes, the standard of the school resource is falling rapidly as the budget allocated for expansion of the educational sector is not proportional to the rate of expansion. As a result of this the standard of the teachers and the working environment is declining from time to time physical setting includes lab, library staff, class room remoteness, etc. This condition makes many teachers leave their profession. Still some other teachers stay in the profession, and some others teachers stay in the profession overwhelmingly by overcoming poor physical facility, low payment and irregularly paid salary. There are also some many teachers who go to school to be counted but not to work conscientiously. This given situation makes the sensitive position of teaching to be occupied by incompetent personnel(Majason, 1995:195).

2.3.1. Class Size

It is one of school-based factors. As indicated by Stinnett, (1968:227), 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 learner. As a rule of thumb for high schools 25 students per classroom is moderate number. As it is indicated by Missouri school improvement program (2006), "a class size of 23 is minimum standard and 28 is desirable standard for secondary school in general."

2.3.2. Teaching Load

Like thee class size there is no well-defined standard teaching load. But not all teachers, some teachers have too little time for rest and relaxation during school hours. Stinnett(1968:232). It means they are involved in the

routine activity of the school and mentally they are tight. According to the case study finding, in the Education system in Germany(1999), it is showed that teachers with teaching load of period per week are heaviest teaching load.

2.3.3. The Amount of Clerical Work Required

In some schools the clerical work alone is a full time work for many organizations. Teachers usually taking attendance every day, period and summarizing, reporting (Barkume, 1998).

In order to make the teachers to be involved in the professional activities effectively and efficiently and some of non-professional personnel can be employed as assistance to give relief for teachers.

According to the report of Wilkinson (2005): “planning, preparation, assessment, record keeping were the biggest contributors to workload.” This shows that unless teachers are given reasonable loads, students cannot make progress the public demands.

As it is put by Stinnett (1968:234)

.... “if possible, adequate clerical assistance should be provided. A little money spent for clerical help lighten the teachers burden considerably. Many schools are seeking to provide relief for teachers in routine, clerical work by providing non-professional clerical helps and high school students.”

2.3.4. The Work Week

According to the study made by Stinnett (1968:234) a teacher works per a week 47-4 hours in working hours. This time was divided into 30hr (63%), which was spent with pupils, and 18hr (37%) in other duties. In addition, 11 hours after school day or hours, these are unpaid hours.

2.3.5. A Free Period for Teachers

Teachers need sufficient time in the school hours so that they can prepare the lesson they teach, using library, and discussion time with their colleagues, sufficient rest and/ or relaxation time. As it is forwarded by Stinnett (1968:234) “a major dissatisfaction of teachers working condition is the sustained nature of the daily schedule. In the DCT study, teachers overwhelmingly expressed the need for a free lunch period”.

As stated by NCAE guideline (2006):

“Principal should not unfairly burden a given teachers by making that teacher give up his or her duty free period on an going, regular basis without the consent of the teachers” this shows that free periods of teachers are very important for accomplishing their task effectively. They use these breaks or free periods as opportunities to talk with colleagues, make photocopies, or drink up of coffee, etc.

2.3.6. Too Many Extracurricular Duties

Many teachers overwhelmingly occupied by duties of different clubs and committees out side the class hours. Especially those teachers who are talented and able teacher carry heavy burdens of these duties than other teachers. If they are over loaded in their regular work and in their extra curricular activity, the environment cannot be conducive for other relevant activities, which require initiation (Barkume, 1998).

In addition to classroom activities teachers have many activities to perform like oversee study halls, supervise extracurricular activities, assist students in choosing courses, colleges and career. They are expected to participate in education conference and workshops. In many schools teachers are increasingly involved in making decisions. According to career

development and job search advice for new college graduates (2006).

2.3.7. Other Factors

The following important relationship has been observed among different variables according to the recent study of Stockard and Mayberry (as cited in NOE (2004:17))

- *Teachers' qualification on the student achievement show only 50% confirmation.*
- *Teachers attending in prestigious school show very little relationship with student achievement*
- *The impact of teachers experience is little more than 50%.*
- *The teachers' verbal proficiency and longevity have positive relation.*
- *Teachers' salary has not shown significant impact on learning.*

As it is observed above, each variable has its own contribution to the student better learning or achievement. The aggregate effect of the entire mentioned variable seems that they have great impact on the students' achievement.

The length of instructional time that is allotted for a given subject, and the number of school days with in a year when supplemented with variety of material resource and management practice, it can boost the learners' achievement. Chen (as cited in NOE (2004:17)). It means that the more the time allotted for a given subject to be taught might have positive impact on students' achievement. But it holds true only if it is accompanied with the necessary materials and effective and efficient school management.

2.4. Implementation of Teacher Made Tests in Ethiopia

Teacher-made tests in Ethiopia are not studied properly. As a result of which information related to test, in general and teacher-made test, in particular are found from the general work of educations in a fragmented way. So far it seems that attention was not given to assess the quality of teacher-made tests. For this study, the sources of information were manuals, unpublished source and very few published sources Betre (1992: 31)

An attempt to study classroom tests (teacher-made tests), even though it was not that much significant, was made in ERGESE (Evaluation Research of General Education System in Ethiopia) project by MOE, for the first time. The report of the study indicates that teacher-made tests lack, content validity and dependability, and the items mainly focus on rote memorization of simple facts. This shows what they tend to measure is questionable, and the decision made based on such tests also cannot make the right person at the right place (MOE, 1986:17)

According to the recent study made by NOE (2004:9), the actual quality of schooling in developing nation is getting worse from time to time. This is because of decreasing of educational expenditure in one hand, and rapid increase of enrollment rate on the other. This contributes to the poor quality of teacher-made tests in particular and poor quality of schooling in general.

Even if the above situation is common to the nations schools, school resources affect teacher-made tests qualities, or the students learning achievement. They are school facilities, per pupil expenditure, teachers qualification, classroom size, material input, teachers experience, testing practices, teachers instructional time, classroom organization, teachers verbal proficiency and, teachers` social ground as stated in NOE (2004: 16). Because

of different reasons different schools of our city have difference in resource. Some of the resource may arise from the ownership of the school (government or public or church or mosques), the experience of the school, the personnel involved in the school, the society around the school, etc. And these can be factors that can determine the quality of teacher-made test.

The MOE put the following points in its manual (1981: 63) regarding teacher-made test evaluation at the school level. There must be examination and evaluation committee, which is led by the deputy director of the school. The committee members are subject department heads, and a student council representative.

Among the many activities of the examination and evaluations committee the major are the following:

To evaluate - whether the exam is based on the stated objective

- the test item format together with strength & weakness.
- How the test (exam) is prepared, administered, scored & reported
- Validity and reliability

To check - whether the items are prepared according to the guideline of test construction.

- Whether the correct response is free from guessing.
- Whether the committee reports to the whole staff

And, finally the committee has to present the raw score of students to teachers and makes statistical analysis so that teachers gain something out of it.

To maintain the quality of teacher-made test, in developing nations like Ethiopia, is a difficult task. Having all these problems, we largely depend on the classroom test for different educational decisions. In developed nations, the “standardized achievement tests” are given to supplement teacher-made test. The comparison of teacher-made test and standardized test is very important for the purpose of quality control (Gragson, 2006). Now and then taking standardized test helps to measure the students’ performance across the country (OERI, 2004). By doing this, the quality of teacher made test and educational decision can be maintained. In our case, instead of standardized achievement tests, there are national examinations, which are limited only to three grade levels, 8th, 10th and 12th. The national exams as much as possible help to minimize problem related with classroom test (Hailu, 1990:1)

The MOE (1986:43-56) forwarded how often one subject teacher needs to prepare and use the tests for a particular grade level is about ten times with in a year.

As the school year is divided into two semesters accordingly, about four progressive tests and one final examination are mandatory in each semester.

2.5. Summary of Review of Related Literatures

It is known that testing cannot be separated from the activities of teaching and learning. And it takes considerable amount of teachers’ professional time. Grean (1963:83) Therefore, teachers should be trained to handle and properly use the testing activities and their results. As pointed out by Gronlund (1981:87-88), the classroom teacher should be the master of the "art of testing" together with his subject mater knowledge. Similarly, Chase and Stiggin (as cited in Frey (2005)). Supports this view by saying that teachers must examine their tests and evaluate the outcome, and it is

only when there is frequent evaluation that teacher made tests be improved.

However, having the necessary skills and knowledge of educational testing only is not enough. The school-based factors like the class size, teaching loads the amount of clerical work required, the work hours per Week, lack of free periods for teachers, too many extracurricular duties are also influential.

Therefore, teachers seem to have no sufficient time for rest or for relaxation during school hours. Especially some teachers are involved in heavy clerical activities which can be a fulltime work for many organizations Stinnett (1968-234) and he added also the sustained nature of schools daily schedule has created dissatisfaction among teachers. This idea is supported by the National Association of college and Employers (2006)

Still some teachers specially the talented ones overwhelmingly occupied by duties and responsibilities of different clubs and committees.

The class size and teaching load if it is more than needed or standard it affects negatively the class scores Stinnett (1968:27-32). And the same idea is reflected in Missouri school improvement program.

Teacher- made test are often prepared associated with school-based factors listed above. As a result, they have known weaknesses. According to Meherns (1969:35), "They are sometimes of inferior quality as they are constructed hastily, unreviewed and untested" Similarly Ashworth (as cited in Betre (1992)) noted that they are highly influenced by the teachers' subjective judgment. He also added that besides their being prepared in the last minutes before testing, most of the items are vague and tend to measure trivial details instead of testing understanding, reasoning, creativity and application. In discussing the weakness of teacher made tests in Ethiopian, MOE (1986:17) states that most items simply focuses on rote memorization of simple facts.

According to the manual of MOE (1986:17) teachers at school level should evaluate their test, based on the stated objectives. It is only by making test item analysis that teachers can point out the strength and weakness of their tests. In the manual it is also indicated that this is the main duty and responsibility of the examination and evaluation committee, in general and school teacher, in particular.

In the preceding chapter, it was presented what others have said regarding the study. In the next chapter, it is tried to show the methods and procedures followed to conduct the research.

CHAPTER THREE

3. RESEARCH DESIGN, METHODS, PROCEDURES AND SOURCE OF DATA

The type of research that is employed is descriptive survey. Here the existing problems are investigated and plausible solutions are forwarded to minimize the problem in the recommendation of this paper.

3.1. Population

The populations for the study are secondary school teachers, tests of first semester, and test score of students of the semester.

3.2. Sample

Since government and non-government organization own the schools of the city, the researcher took representative schools from both groups. Because the majority of the secondary schools are governmental, the researcher took more samples from government schools. Three government schools and two non-government schools are taken, purposely for investigation.

From all government schools and non-government schools the samples are selected by using simple random sampling technique. Five secondary schools are selected as sample representatives of secondary schools of the city. And, again from each sample school one grade 9 section is randomly selected for the purpose of analyzing test scores and test quality.

The schools are: **Shimelis Habte, Abyot Kirs, Delachen, Cathedral and Assai**. The first three, Shimelis Habte, Abyot Kirs and Delachen are government secondary schools. And the remaining two Cathedral and Assai are non-government secondary schools.

3.3. Data

The data of the study are first semester marked test papers of English and mathematics, 8th grade national examination score of the 9th grade students, first semester average score of the students, the responses of teachers, and principals.

3.4. Instruments of the Study

The following data collecting instruments were used depending on the nature of the data. Questionnaires were prepared and distributed to teachers. Twenty questionnaires, each containing 26 items, for each school and totally one hundred questionnaires were distributed to all the schools in the study.

Interviews were conducted with the principals of the schools. It helps to secure data about the overall condition of teachers and working environment an interview is to be held with one principal from each school.

Observations were also made to the construction of test and the responses with its answer key and according to the guidelines of test construction. Test score analysis, and test item analysis were conducted.

Quantitative analysis based on test score and test item was done to determine the difficulty and discrimination level of the test items used. And reliability of test items and prediction validity of the 8th grade National exam test score and the 9th grade first semester final exam will be seen.

- Reliability using split half method, and Kuder Richardson # 20 (KR-20) formula.
- Predictive validity by using product moment correlation coefficient method.

After describing the exam items in qualitative and quantitative terms, teachers' problems in analyzing and evaluating their exam after they have administered them would be investigated. In qualitative description, unclear direction, difficult and complex terms or sentences, poorly constructed test items and improper arrangement of items will be seen.

CHAPTER FOUR

4. DATA PRESENTATION AND ANALYSIS

In this chapter data obtained from teachers, principals, test papers of first semester 9th grade English and mathematics exams, 8th grade national examination scores, and 9th grade first semester average test scores students,' from teacher-made tests, are presented and analyzed.

4.1. The Current Teacher-Made Tests Quality Assessment

4.1.1. Qualitative analysis of the tests

4.1.1.1. Test item analysis (refer appendix-2)

The methods used to make item analysis is that if the difficulty indices are above 90%, then the item is too easy, and too difficult if the index is below 30%. When the discrimination indices are below 15%, the item is not acceptable (Bowen, 1985: 388-389; Zurawski, 2001; Taiwo, 1995)

For Abyot kirs out of 40 item of the English test, 12 (30%) items seem defective. Especially 4 of the items seem more defective in both discrimination and difficulty indices. And again these items cannot discriminate high scorers from low scorers. In mathematics out of 32 items test 13 (41%) seem defective. Again out of which 8 (25%) of them seem more defective in both discrimination and difficulty indices.

English for Delachen, out of 32 items of the test 10 (31%) of them seem defective. Again 3 of them seem more defective in both discrimination and difficulty indices. In mathematics test out of 31 items 14 (45%) of them seem defective. 9 of them are very difficult and 3 of them cannot discriminate high scores from low scorers.

English examination for Lideta Catholic, out of 58 items of the test 17(29%) of them seem defective. Especially 8 of them seem more defective in both difficulty and discrimination indices. And again, 4 of the items are ambiguous as they discriminate negatively. In mathematics, out of 30 items 6 (20%) are defective. The items in general seem difficult.

From 30 items of English test of Shimelis Habte, it seem that 15 (50%) of the test items seem defective. Again 6 of them are defective in both difficulty and discrimination indices. In the mathematics test, 12 (40%) of the total the items seem defective, and again out of which 10 of them seem more defective in both difficulty and discrimination indices.

For Assai, out of 60 item tests 13 (22%) items seem defective. Specially 4 items have both discrimination and difficulty indices beyond the standard stated. And the remaining 9 items have at least one defect. In the school mathematics examination, 18 (40%) of the items out of 45 items seem defective. 16 of items are either too difficult or cannot discriminate high scorers from low scorers.

4.1.1.2. Reliability of tests

Table I
Reliability of the tests

School	Subject	Reliability of the Test		Remark
		Split-Half Method	KR-20	
Abyot kirs	English	0.765	0.795	
	Mathematics	0.747	0.783	
Assai	English	0.936	0.913	*
	Mathematics	0.577	0.746	
Lideta catholic	English	0.811	0.816	
	Mathematics	0.852	0.903	*
Delachen	English	0.632	0.683	
	Mathematics	0.740	0.709	
Shimelis Habte	English	0.412	0.450	**
	Mathematics	0.846	0.825	

The tests have moderate internal consistency except the tests marked with *and* * in the remark . It seems most the test items measure similar or common characteristics of the person. Most of the items are homogeneous, exceptionally English test of assai and Lideta catholic. But Shimelis Habte English test lack internal consistency.

4.1.1.3. Correlations of the National Exam scores with first Semester Teacher-Made tests average scores

(For further information refer the appendix 3 at the back of this text.)

TABLE II

Schools	Correlation coefficients	Coefficients of determination	Ranges Of The variables	Remark
Abyot kirs	0.330	0.109	$\Delta X=29$ $\Delta Y=41$	
Lideta catholic	0.394	0.155	$\Delta X=40$ $\Delta Y=40$	
Delachen	0.349	0.122	$\Delta X=37$ $\Delta Y=66$	*
Assai	0.442	0.195	$\Delta X=26$ $\Delta Y=43$	**
Shimels habte	0.408	0.166	$\Delta X=29$ $\Delta Y=37$	

where:

Δx - the range of the national test score of the students

ΔY -the range of teacher-made tests average score of students.

It is known that the larger the range between correlating variables the higher the correlation coefficients, provided other variables are constant.

As there are weak correlation and determination coefficients, the prediction validity is also weak, and at the same time the reliability of the test is also weak. It seems that Assai tests scores are more reliable than others despite the smaller the ranges between the variables.

The correlation coefficients range from 33% to 44.2%. And from the coefficient of determination it is possible to understand that only the variation of scores from 10.9% to 19.5% is reflected on the variation of scores in the national examination scores.

4.1.2. Qualitative Analysis of the Tests

Points observed by the researcher in each teacher-made test

4.1.2.1. For English Tests

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

	S1	S2	S3	S4	S5	
	YN	YN	YN	YN	YN	
1. More than two item formats.	☐	☐	☐	☐	☐	
2. Clearly stated instruction.	☐	☐	☐	☐	☐	
3. Stating time allotted.	☐	☐	☐	☐	☐	
4. Stating the weight for each item.	☐	☐	☐	☐	☐	##
5. Clarity and neatness of the test paper.	☐	☐	☐	☐	☐	

In multiple choice

6. Keeping the length of all option uniform.	☐	☐	☐	☐	☐
7. Avoiding negatives specially double negatives.	☐	☐	☐	☐	☐

8. Avoiding grammatical clue. □□ □□ □□ □□ □□
9. Not repeating the key words from item in
the options. □□ □□ □□ □□ □□
10. Are the distracters plausible? □□ □□ □□ □□ □□ ##

In true- false

11. Making the statement clear
(free of ambiguity) □□ □□ □□ □□ □□
12. Avoid specific determiners like
always, never, all and none □□ □□ □□ □□ □□ ##
13. Making the length and number of
True & False statements about the same. □□ □□ □□ □□ □□ ##

In Matching

14. The items in them are homogenous
so that the option can be plausible. □□ □□ □□ □□ □□ ##
15. The length of the column should be
unequal, and the response should be longer
from the premises side. □□ □□ □□ □□ □□ ##

In supply item or short answer item format

16. Does the test constructors specify
they degree of precision and unit of
expression expected in answer? □□ □□ □□ □□ □□
17. Not having many blanks to occur
at the end of the sentence. □□ □□ □□ □□ □□
18. Avoid excessive blanks in a single item. □□ □□ □□ □□ □□
19. All blanks should be equal regardless
of the length of the answer. □□ □□ □□ □□ □□

For the test in general

20. Are the items arranged from
simple to complex? □□ □□ □□ □□ □□ ##

4.1.2.2. For Mathematics Tests

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

	S1	S2	S3	S4	S5
	YN	YN	YN	YN	YN
1. More than two item formats.	☐	☐	☐	☐	☐
2. Clearly stated instruction.	☐	☐	☐	☐	☐
3. Stating time allotted.	☐	☐	☐	☐	☐
4. Stating the weight for each item.	☐	☐	☐	☐	☐
5. Clarity and neatness of the test paper.	☐	☐	☐	☐	☐

In multiple choice

6. Keeping the length of all option uniform.	☐	☐	☐	☐	☐
8. Avoiding negatives specially double negatives.	☐	☐	☐	☐	☐
8. Avoiding grammatical clue.	☐	☐	☐	☐	☐
9. Not repeating the key words from item in the options.	☐	☐	☐	☐	☐ ##
10. Are the distracters plausible?	☐	☐	☐	☐	☐

In true- false

11. Making the statement clear (free of ambiguity).	☐	☐	☐	☐	☐
12. Avoid specific determiners like always, never, all and non.	☐	☐	☐	☐	☐
13. Making the length and number of True & False statements about the same.	☐	☐	☐	☐	☐

In Matching

14. The items in them are homogenous so that the option can be plausible.	☐	☐	☐	☐	☐ ##
---	--------------------------	--------------------------	--------------------------	--------------------------	-----------------------------

15. The length of the column should be unequal, and the response should be longer from the premises side. ☐☐ ☐☐ ☐☐ ☐☐ ☐☐ ##

In supply item or short answer item format

16. Does the test constructors specify they degree of precision and unit of expression expected in answer? ☐☐ ☐☐ ☐☐ ☐☐ ☐☐
17. Not having many blanks to occur at the end of the sentence. ☐☐ ☐☐ ☐☐ ☐☐ ☐☐
18. Avoid excessive blanks in a single item. ☐☐ ☐☐ ☐☐ ☐☐ ☐☐
19. All blanks should be equal regardless of the length of the answer. ☐☐ ☐☐ ☐☐ ☐☐ ☐☐

For the test in general

20. Are the items arranged from simple to complex? ☐☐ ☐☐ ☐☐ ☐☐ ☐☐ ##

NB

== Refers to the item format not included in the test.

Refers to the quality the test or test item lacks

From the above observation:

- ☞ Giving the weight assigned to each item is not usually practiced.
- ☞ The selection and numbers of item format seen dependent on the choice of the teacher who construct the test.
- ☞ Multiple choice item formats seems that the most preferred item format than others.
- ☞ From item analysis and observations of each test item, it is possible to identify many vague or ambiguous items.

4.2. Data Obtained from Teachers

TABLE III Personal Characteristics of the Teachers

Items	Options	S _{abyo}	S _{ass}	S _{l..cath}	S _{dela}	S _{shem}	In number	In %
1.sex	Male	13	18	15	19	17	82	84.5
	Female	6	2	3	1	3	15	15.5
2.Age	Below 25	2	5	2	2	2	13	13.4
	26-35	5	10	10	14	5	45	46.4
	36-45	10	2	5	1	3	21	21.6
	Above 45	2	2	1	3	10	18	18.6
3.Experience	Below 6	1	12	5	6	3	27	27.8
	6-15	8	4	8	11	4	35	36.1
	16-25	5	2	2	2	7	18	18.5
	Above 25	3	2	1	3	6	15	15.4
	Not responded	2	-	2	-	-	4	4.2
4.Field of study	Language	3	3	5	4	8	22	22.7
	Mathematics	3	3	4	4	3	17	17.5
	Natural Sc.	7	7	4	7	5	30	30.9
	Social Sc.	4	4	3	4	3	18	18.5
	Sport	-	1	-	1	1	3	3.2
	Not responded	2	2	2	-	-	6	6.2
5.Qualification	Diploma	3	1	-	1	4	9	9.2
	Bachelor Deg.	14	19	18	19	16	86	88.2
	Masters Deg.	-	-	-	-	-	-	-
	Other	-	-	-	-	-	2	2.1
	Not responded	2	-	-	-	-	-	-

From item 1 of table III, we see about 84.5% of the teachers are male teachers.

And item 2 shows about 46.4% are adults between the age of 26 and 35, 40% of the teachers are above the age of 35. In item 3, the majority of the teachers, 54.6%, in the city have work experience from 6 to 25 years. It seems that the non-government schools are filled with younger teachers than governments.

From item 4 the majority of the teachers are natural science and language teachers. Item 5 shows the teachers' qualification is sufficient 88.2% have Bachelor degree.

TABLE IV. Teachers Training Background

Items	Options	S_{abyo}	S_{ass}	S_{l.cath}	S_{dela}	S_{shem}	In number	In %
6.The college or university they have graduated from	Bahir dar	2	5	-	4	0	11	11.3
	Kotebe	3	3	3	2	6	17	17.5
	Alemaya	-	1	-	-	1	2	2.1
	AAU	10	9	12	9	12	52	53.6
	Dehub	1	1	-	4	2	8	8.2
	other	3	1	2	1	-	7	7.2
7.The teachers responses regarding taking teachers training courses	Positively	17	17	18	19	20	91	93.8
	Negatively	-	-	-	1	-	3	3.1
	Not responded	2	1	-	-	-	3	3.1
7.1. The teachers with out teachers training courses towards the know how of measurement and evaluation	Positive	-	2	-	2	2	6	100
	Negative	-	-	-	-	-	-	-
	Not responded	-	-	-	-	-	-	-
		-	-	-	-	-	-	-
7.2. Teachers having teachers training courses responses and their familiarity to measurement and evaluation courses.	Positively	17	17	14	18	15	81	89.0
	Negatively	-	-	3	1	5	9	9.8
	Not responded	-	-	1	-	-	1	1.2
7.2.1 The responses of teachers who have taken measurement and evaluation courses in using in their testing activities.	Positively	12	16	12	14	13	67	82.7
	Negatively	5	1	2	4	2	14	17.3
	Not responded	-	-	-	-	-	-	-

In table IV item 6, it seems that the majority, 53.6% of teachers are 17.8% from AAU & others are from Kotebe teachers' colleges. In item 7 more than 93% of the teachers are trained to be a teacher. In item 7.1, all the teachers who have not taken measurement and evaluation responded positively towards the know-how of measurement and evaluation. In item 7.2, teachers taking teachers training course are familiar to measurement and evaluation. And about 82.7% of the teacher with teachers training course feel that they are using measurement and evaluation techniques in their teaching. According to the presentation in items in table V and in items in table IV, usage of post test analysis do not seem to depend on qualification and field of specialization.

Table V. A. PRACTICE OF CONSTRUCTION AND EVALUATION OF TEST

Items	Options	S _{abyo}	S _{ass}	S _{L.cath}	S _{dela}	S _{shem}	In num.	In %
8. Responses of teachers in taking some sentences or expressions directly from students textbook	Positively	11	14	12	15	14	66	68.0
	Negatively	7	6	6	5	5	29	29.9
	Not responded	1	-	-	-	1	2	2.1
9. The time when teachers construct their test	Before the course...	-	-	1	1	1	3	3.1
	During giving the course.....	3	9	7	8	5	32	33.0
	When the testing time is approaching..	15	9	10	10	13	57	58.8
	Immediately before the day of the test...	-	1	-	1	1	3	3.0
	Not responded...	1	1	-	-	-	2	2.1
10. Teachers responses whether they have ever evaluated their test or test items after the have administered	Positively	15	16	17	14	15	77	79.4
	Negatively	1	4	-	6	4	15	15.5
	Not responded	3	-	1	-	1	5	5.1
10.a.the frequency of evaluating teachers made test	Always	10	7	13	3	8	41	53.3
	Sometimes	5	6	2	11	7	31	40.3
	Rarely	0	1	2	-	-	3	3.8
	Once in a year		2	-	-	-	2	2.6
	Not responded	-	-	-	-	-	-	

In Table V A above in item 8, about 68% of the teachers are taking some phrases, sentences or expressions directly, with out paraphrasing, from students textbook or textbook.

And about 59% of the teachers are constructing tests when the time of testing is approaching, and about 33% of them are while giving the course as in item 9. About 80% of the teachers feel that they always evaluated their tests, as in item 10.

Table V.B. PRACTICE OF CONSTRUCTION AND EVALUATION OF TESTS

Items	Options	S_{abyo}	S_{ass}	S_{Lcatc}	S_{Dela}	S_{Sheme}	In number	In %
11. The average class size	Below 40	-	-	-	-	-	-	-
	41-55	-	20	3	-	-	23	23.7
	56-70	1	-	15	-	-	16	16.5
	71-85	4	-	-	16	7	27	27.8
	Above 85	14	-	-	4	13	31	31.9
12. Teaching periods per a week	Below 16	1	5	2	-	3	11	11.3
	16-20	3	11	4	13	4	35	36.1
	21-25	15	4	12	5	13	49	50.5
	26-30	-	-	-	2	-	2	2.1
	Above 30	-	-	-	-	-	-	-
13. The numbers of tests with in a semester	2	15	1	1	1	11	29	30
	3	2	4	-	10	7	23	23.7
	4	2	2	8	6	1	19	19.5
	5 or more	-	13	9	3	0	25	25.8
	Not responded	-	-	-	-	1	1	1.0
14. The frequency of clerical activities in which teachers are involved.	Every week	8	11	10	11	12	52	53.6
	Every month	1	7	1	2	2	13	13.4
	Every semester	5	1	3	6	2	17	17.5
	Other	5	1	4	1	3	14	14.4
	Not responded	-	-	-	-	1	1	1.0
15. numbers of teachers contact with parents	Often	-	1	3	1	-	5	5.2
	Sometimes	4	12	11	7	7	41	42.3
	Rarely	8	6	2	11	8	35	36.0
	Not responding	5	1	-	1	5	12	12.4
		2	-	2	-	-	4	4.1
16. The activities in which teachers are involved in in schools other than teaching	Chatting with friends.....	5	4	5	6	9	29	17.2
	Preparing lesson..	3	4	10	12	8	37	21.9
	Participating in clubs or committee...	6	16	14	22	14	72	42.6
	Scoring test papers.....	3	1	9	9	3	25	14.8
	Other activities...	2	-	2	1	1	6	3.5

It seem that the class size of students are above 85 in government schools,

and in non government school it seems over than this, as indicated in Table V B, item 11. Most teachers, 50.5%, have teaching load from 21 to 25 periods per a week, and 36.1% have teaching load from 16 to 20 periods in item 12. It seems that in majority government schools most teachers give 2 tests but this seems different in some schools specially in non- government schools, in item 13. Teachers in city, most of the time, are required to work clerical work every week through out the year, as indicated in item 14. as indicated in item 15, about 42.3% of the teachers make contact with parents of their students sometimes and about 36% of them rarely.

About 42.6% of them participate in schools at different clubs and committees, and 21.9% are engaged in preparing lesson and its plan, as in item 16.

Table VI School based Factors for Teachers

Items	Options	S _{abyo}	S _{ass}	S _{Lcate}	S _{Dela}	S _{Sheme}	In number	In %
17. The numbers of free periods during School time	Less than 4	- 3	-	-	1	3	7	7.2
	4-8	4	2	1	2	5	14	14.4
	9-12	4	2	11	4	5	26	26.9
	13-16	5	6	3	6	2	22	26.6
	More than16	2	10	3	7	5	27	27.8
	Not responded			-		-	1	1
18. Teacher Responses in Working out side their normal session of teaching	Positive	6	12	12	8	7	45	46.6
	Negative	13	8	6	12	13	52	53.6
	Not responded	-	-	-	-	-	-	
19. The preferred time for those who workout side the normal working hours	Morning or afternoon...	-	1	2	1	3	7	15.5
	Evening	4	6	7	-	2	19	42.2
	Weekend	1	2	2	2	2	8	17.8
	If other	-	3	1	5	1	11	24.8
20. Teachers responses in participating in clubs or committees	Positive	11	18	17	17	12	75	77.3
	Negative	8	2	1	3	8	22	22.7
21.the numbers of clubs or committees in which a single teacher participate	1	3	6	8	4	3	24	32
	2	5	9	5	9	8	36	48
	3	3	3	2	4	1	13	17.3
	4	-	-	-	-	-	-	2.7
	5	-	-	-	-	-	-	-
	Others	-	-	-	-	-	-	-
22 The role teachers play in the clubs or committees	Coordinator	4	7	11	9	3	34	32.2
	Member	12	11	7	20	17	67	63.0
	Other	-	2	-	-	-	2	1.9
	Not responded	3	-	1	-	-	3	2.8

When we refer Table VI, Teachers of the city 49.9% have free time ranging from 9 to 20 periods per a week in item 17, and about 53.6% of the teachers do not work out side the normal working hours, but about 46.7% of teachers work in addition to the normal working time, as in item 18. The most preferred time to work is evening time, item 19.

Regarding teachers' participation in clubs and committees, as it is indicated in item 20 and 21, 77.3% of the teachers are participating in different clubs or committees where as 22% are not participating. About 48% of the teachers are involved in two clubs or committees whereas 32% of them are participating at least in one. About 32% of them are coordinators where as 63% of them are members, item 22.

4.3. The Responses of Principals on the Interview Held with them

1. Regarding the manual, which was prepared by MOE (1986), contains test related issue and other activities that schools should follow. In most schools could not be found easily. Let alone the schoolteachers, the principles do not know where it is. In some schools it was found on the corner of the shelf in principal's office. From this we can understand that referring the manual is not usually practiced.
2. The presence of examination and evaluation committee is common in all schools. But in one of the non-government school, the principal's and department heads implement the activities of examination and evaluation committee, without forming it, through organization's chain. Because there is no unique and tiresome activities to be performed by the committee of examination and evaluation, the presence or absence seems does not have impact on testing procedure.

3. The members of examination and evaluation committee are department heads and other administration workers and are led by the school deputy principals.
4. In all schools, the roles of the committee seem that very much concerned with the procedure that must be followed before and after test administration. The main duties of the committees are scheduling, collecting the test manuscript, duplicating, stapling, dispatching test papers to the examinees and collecting answer sheets, and finally submitting the answer sheet on which the students respond to the subject teacher. The role of the committee in improving the test item quality is not implemented.
5. The report of the examination and evaluation committee is not common. If they report, the report focuses on issues related to test administration and rarely the numbers of students promoted and detained on each subject.
6. Among the typical efforts that are made to improve teacher-made tests, the school managements often insist teachers to prepare at least four different item formats. And very few schools have attempted in giving refreshment courses to their teachers by test experts from outside. Still others have tried to give refreshment course of measurement and evaluation by inviting senior and experienced teachers among teachers.
7. The roles expected to be played, in school compound from teachers other than teaching in classroom, are participating in at least in two clubs or committees, conducting research, preparing lesson and its plan, giving tutorial, scoring test papers, etc.

8. Concerning teachers' knowledge and skill of measurement and evaluation together with the shortage of time. The principals responded that almost all teachers are trained to be a teacher; they confirmed that teachers have the necessary skill and knowledge as to how test analysis is implemented. Regarding time constraint, not all the teachers but some might have shortage of time due to over obligations they are given.
9. In the usage of educational technologies that facilitate teaching and learning, there is no as such unique difference in all schools. In all schools, as it is reported, there is limitation in computer usage including Internet.
10. All principals responded that there is good working relationship between administration and teachers and among teachers themselves.
11. Concerning the length of one period and the frequency of teaching periods for English and mathematics, the principals responded differently. In government schools, the length of one period is 40 minutes with 5 periods per a week. In non-government schools, the length of one period is 45 minutes with 6 periods per a week. This means the learning periods for non-government schools exceeds by 75 minutes per a week.
12. Concerning teachers' attrition rate and their main reasons. The response of most principals is the same that is plenty in most schools. Except in the school where the teachers are very experienced and relatively older. In this school there is a means of earning substantial income by working night schools.

The main cause for the high attrition rate is, reposed by the principals, most teachers seek higher income than they earn currently. To minimize this, non-government schools have tried to improve teachers' monthly salary, and

some of these schools have arranged learning opportunities by minimizing their teaching loads.

Regarding the instruments (questionnaires, interview questions and observed points in the test) used to secure data refer appendix 1 attached at the back of this material.

In addition, to check the current level of teacher-made tests quality, the correlation between the teacher-made tests scores and 8th grade national examination scores of particular group of students are reveals in appendix 3.

CHAPTER FIVE

5. DISCUSSION

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

To construct better quality test items, the experience of the teachers coupled with conducive environment is very important. As indicated by Yeager (as cited in Hailu (1990)) experience alone is nothing without creating a stimulating environment, which is capable of encouraging teachers to develop their professional skills. This idea is confirmed by Nunnally (as cited in Hailu (1990)) the majority of teachers learn to write better items after several years of practice. As it is indicated in table III item 3, of the study teachers in the city are well experienced, the majority of them (about 36%) have experience ranging from 6 to 15 year, and about 38% of them have more then 16 years of experience. However, the test item analysis reveals that 20% to 50% of the total test items are defective. This shows that experience alone doesn't bring construction of higher quality test items. This might also indicate that, the school environments are not stimulating to construct better quality test. Since teachers work is highly related with testing activities, teachers need to know cyclical activities of the testing process (Chase and Stiggin (as cited is Fray (2005))). In order to know to what extent learning has taken places, teachers should design test to measure it. Ebel (1975: 73). And Gronlund (1981: 87-88) further stated subject matter knowledge is not enough, teachers should have the skill and knowledge of educational measurement. All studies, listed above, stress the importance of teachers' knowledge and skill in area of measurement and evaluation.

In this study context, (Table V) item 7.2 indicates that 89% of teachers have taken training of teaching, and have knowledge and familiarity with class measurement and evaluation skills.

Teachers, in general, are working in sustained nature of daily schedule, which is the major dissatisfaction of teachers about their working conditions. (Barkume, 1998; Stinnett, 1968-234). Teachers are engaged usually by different clerical activities like taking students attendance (every day & every period), and preparing summaries and reports, which can be a full day work for many organizations. As a result of this, teachers are often focusing on non-professional activities. To lighten teachers' burden and enable them to concentrate on their professional duties, non-professional personnel can be employed (ACER, 2005). Moreover, many teachers overwhelmingly are occupied by duties and responsibilities of different clubs and committees. All these show that they cannot effectively perform other relevant activities, which are required by the profession (Stinnett, 1968:234). The same idea is presented by career development and job search advice for new college graduates (2006).

In this study also, teachers on the city have similar time related issue as indicated in (Table V) item 8, the majority of teachers (68%) teach students with class size up to 70 students, even in some classes it can reach up to 100 students. About 50% of the teachers have teaching load from 16-20 periods per a week. Regarding the number of tests, 68% of the teachers are required to give at least 3 tests and utmost 5 tests with in a semester for. Teachers are required to report at least every week to school officials (Table VI), Item -14. And teachers are required to make discussion with parents sometimes (42%). Moreover 42% of the teachers are passing with the remaining time by

involving in club or committee participation, as they are required to participate at least in two different committees. 49.5% of teachers have free time with in a week from 9 to 16 periods. In addition, nearly half (47%) of the teachers are working outside the normal working hours. In addition to this, preparing lesson and lesson plan and scoring test papers are mandatory for all teachers.

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

Tests in general, and teacher-made tests in particular should have the desired quality: appropriate index of difficulty and discrimination, and appropriate reliability and validity coefficients, quantitatively. And in qualitative terms the presence of clear instruction, different item formats, clear statements (which are free of ambiguity), good arrangements, from simple to complex. Teacher-made tests in general have the following drawbacks. As stated by Mehrens (1969: 135) they are constructed hastily, unreviewed and even untested and he added also they are not as sound as Commercial tests. Similarly Asworth (as cited in Betre (1992:31)) has pointed out the weakness of teacher-made tests as follows: they are highly influenced by the subjective judgment of the teachers, prepared in last minute of testing as a result of which they are highly susceptible to be unreliability, they inadequately sample the whole area, they measure only trivial details by neglecting the objectives of the test. Despite this limitation, they are used for making critical decisions.

From the test item analysis, in some schools', teacher-made tests about 50% of the total test items seems defective.

Some of the items are below the standard of discrimination indices. Some of these items are not only below the standard but also discriminates negatively (where low scorers perform better than high scores).

And still the remaining of the defective items is either far below or above the standard of difficulty indices (i, e, too difficult or too easy).

Concerning the reliability and predictive validity, the internal consistency (reliability) of the tests seem moderate in most schools when spilt half method and kuder Richardson #20 formula is used. It shows that the tests seem measure similar characteristics of a person (Thorndike, 1977:82).

Moreover the correlations of the test scores with the 8th grade national exam scores are found. This shows that the reliability of the tests as refers to external measure is questionable. From the coefficient of determination, it is possible to deduce that the variation of the test scores associated with the variance of national test score is not more that 25%.

From the researcher's observation the number of test items in some schools are almost twice that of the other schools. The presence of some non-plausible distracters, in multiple choices, and vague items are among common characteristics of the test. Indicating the weight of every item of the test is not much known or practiced.

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

Even though teacher made tests' constraints are common in all schools, there are some differences among government and non- government schools.

In government schools, the numbers of test items are 30 to 40. The numbers of items are more in non-government schools it reaches up to 60.

This may permit higher reliability of the test as it covers wide sample of the whole area (Thornike, 1977:79-80).

The class sizes in non-government school are better as they are from 45-70 students where as government schools they range from 85-100 (TableVI) item 11.

The number of tests given in non - government schools in one subject can vary from 3 to 5 tests, where as in government schools they can vary from 2 to 3 tests. As it is presented in (Table VI) item 13, the frequency and the length of English and mathematics teaching periods are 6 periods per a week, with 45-minute each, where as in government schools there are 5 periods with 40 minute each.

CHAPTER SIX

6. SUMMARY, CONCLUSION AND RECOMMENDATIONS

6.1 Summary

The aim of the study was to assess the quality of current teacher made tests at the 9th grade level, and the factors that affect teachers from posttest analysis and evaluation.

The nature of the study is descriptive survey

To study this, the researcher had raised the following research questions.

- ❖ Do teachers have awareness and knowledge about the usefulness of item analysis evaluation?
- ❖ Do teachers have enough time to make posttest analysis and evaluation?
- ❖ Do our teacher-made tests have the desired quality according to the rule of test construction?
- ❖ Do teachers have similar test related problems in all schools regardless of the type of ownership?

The population of the study was secondary school teachers of Addis Ababa, and teacher made tests of the city.

The samples consist of a total of 5 schools (3 government and 2 non-government schools), which were selected by using simple random sampling technique. In further selection process the researcher randomly selected one grade 9th section for the study from each of the selected schools. In fact the class size differ from school to schools, ranging from 46 to 102.

Regarding the tests, English and mathematics tests were selected as a sample.

In order to get data related to the quality of the current teacher made tests, the first semester final test of English and mathematics were chosen. The test papers and answer sheet, after it had been scored (Marked), were taken from the subject teachers and photo copied. Then, based on the response of the candidates, test item analysis for the two subjects, in every school totally 10 subjects, were carried out. The item analysis focused on difficulty and discrimination index of every item.

Moreover, to see how the test or test items are reliable (homogeneous), split- half method and kuder Richardson-20 formula were used. Since the reliability is the necessary condition for validity, the researcher tried to check the test quality, by correlating it with external criteria, 8th grade national examination score. This correlation coefficient helps to see to what extent the teacher-made tests are reliable, and to what extent 8th grade national exam predicts success in 9th grade.

In addition, the tests constructed in every school were examined using checklist. The other data were secured from teachers by using questionnaire. In each school 20 questionnaires were distributed and 97 were collected from the total 100 given out.

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 discussions were held in light of the finding obtained and the literature reviewed.

6.2. Conclusion

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

1. No matter how teachers in the city are well qualified and experienced, and have the knowledge and skills of measurement and evaluation, the practice of about posttest analysis and evaluation seem poor. Although they make efforts to construct or select good items before testing, they don't seem to carry out item analysis and evaluation after words.
2. Most of the school environments don't seem to encourage teachers to engage in task 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 test item analysis.
3. Schools in general, do not have a trend of analyzing and evaluating test or test items after their administration. Only the subject specialist checks the items in terms of the subject content and grammatical construction at the beginning.
4. Most teachers in the city are involved in over crowded duties, which are ever increasing from time to time. This makes many teachers to refrain from binding themselves 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 initiatives.
5. Considerable numbers of teacher made tests of the city are defective interims of difficulty and discrimination indices of the items. Regarding the reliability and validity of the tests, they are marked with better internal reliability than external reliability. It seems that the test items are internally homogenous, which means they measure similar

characteristic of a person. Besides, its predictive validity is also weak.

6. Even if examination and evaluation committee exists in almost all schools among the practices stated in the manual of MOE (1986), evaluating test items and finding reliability and validity of the test and reporting to the whole staff are not practiced.
7. Teacher made tests in non-government schools are more in numbers than in government schools. And more over the number of item in testing and teaching period are more than those of government schools. This might contribute to wider content coverage of the course, which can bring about better reliability of test together with its validity. But higher reliability of the test scores in non-government schools is not specially refuted on the study.

6.3 Recommendations

1. Even though teachers of Addis Ababa at secondary school level have had some training in measurement and evaluation, they don't seem to have the experience of applying it in evaluating their own exams. Most of the schools have also no such traditions. Therefore, schools ought to provide teachers' short-term refreshment courses to encourage them conduct post test analysis and thereby establish the tradition of this valuable professional practices.
2. The school management and other related officials should encourage teachers by scheduling sufficient time for post-test analysis and evaluation. And they should include the activity as one of the major requirement to promote the teachers' in career structure.
3. Teachers need to understand that constructing or selecting what appear as good test items, with apparently appropriate format, at the beginning

is not a guarantee for good quality test item, unless 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 administration.

4. School principals and other officials should restrict teachers, as much as possible from engaging in a number of clubs or committees, so that teachers will have sufficient time to carryout their task of analyzing exams effectively as the main part of their professional duties and responsibilities.
5. Deputy principals together with department heads need to design a system to check to what extent the teacher- made tests address the learning objectives stated in the syllabus. They also advised to check the test items against the table of specification prepared. This makes the tests measure different characteristic of a person instead of addressing a particular characteristics only.
6. The presence of examination and evaluation committees, in the current trend has very little contribution in the quality of the test items. Therefore, the committees need to be organized and well equipped so that the task of evaluating exams stated in the MOE manual can be implemented. And the school managements are advised to create a means to empower the committees, with the skill and knowledge of measurement, to accomplish its task competently.
7. Regarding time allotment for school subjects' non-government schools, assign more time than government schools which enable them to cover content with sufficient detail. This has positive impact in increasing test reliability. Thus, it is advisable if schools extend their half-day school program to the whole day school time for a better and reliable test.

References

- Ashworth, A.E (1982). Testing for Continuous assessment. London: Evan Brothers Ltd.
- Bishop, B. (1986). Innovation on Education. New York: Macmillan Co.
- Bertrand, A and cebula,J.(1980). Tests,Measurement and Evaluation:a Developmental Approach.Reading. Mass:Addison wesly.
- Barkume, Megan (1998). "Teacher Aides". Occupational Outlook Handbook. Retrieved from <http://states.bls.gov/ovohom.htm>.
- Betre Meshesha (1992) "practice & process of classroom test testing in some senior secondary schools of Addis Ababa" unpublished senior essay. AAU.
- Bowen, J.D.(1985) TESOL Technique and procedure. Brounlum: Rowley Newbury House Publishers.
- Brown F.G (1970). Principle of Educational and Psychological Testing. 3rd ed. Tokyo: Iowa state university: Holt, Rinehart and Winston.
- (1981). Measuring classroom achievement. Iowa State University New York: Holt, Rinehart & Winston.
- Clark, L.and Starr, I. (1991). Secondary and middle school teacher method. 6th ed. New York: MacMillan Publishing Co.
- Eble, Robert. (1972). Essentials of educational measurement. 2nd ed. Englewood Cliffs, JS: Prentice – Hall
- (1975) Measuring Educational Achievement. Englewood Cliff-NF. Prentice- hall.
- (1979). Essentials of educational measurement. 3rd ed. Englewood Cliffs, JS: Prentice – Hall
- Frey, B.B. (2005) . "An introduction to Quality Test Construction " University of Kansas. Retrieved from <http://www.specialconnection.com.et>.
- Gragson, Thomas E. "Note on Students Assessment" Retrieved from <http://www.studentassessment.htm>.
- Frith, D.S. and Macintosh, H,G. (1984). A Teaching Guide to assessment. Glasgow: Stanty Thories Ltd.

- Gall, Borg Gal (1996). Educational Research an Introduction. USA: Longman Publisher:
- Green, J.A (1963). Teacher Made Tests. NewYork: Haspar & Row.
- Gronlund, N.E (1982). Measurement and evaluation in Teaching. 4th ed, New York: Macmillan.
- (1985) Measurement & Evaluation in Teaching. 5th ed, London: Collier, Macmillan.
- Hailu Sime (1990). "Teachers Estimate of test item difficulty an indication of their prior knowledge." Unpublished source. (MastersThesis). AAU
- Heaton, J.B. (1990). Class room Testing. London: Longman.
- Horn, M. Lee van (2003) Educational and psychological measurement.
 "Assessing the unit of measurement for school climate through psychometric and out come analysis of the school climate survey"
 New Delhi: Vol 63, No 6, SAGE publication.
- Keeves, J.P. (1994). "Test: Different Types" The International Encyclopedia of Education. (2nd ed, vol. 11 p-6340-49), Great Britain: PERGAHON.
- The International Encyclopedia of Higher Education. (1977) vol. 1 p-543a
 London: Jossey-Bass Publisher.
- Missouri Department of Elementary and secondary Education. "Missouri School Improvement Program." Retrieved on Dec 15,2006 <http://www.schoolimprovementprogram.com>.
- Madsen. H.H. (1983) Techniques in Testing . Oxford University press.
- Majason J.A. (1995) The Teachers Profession: A manual for professional Excellence. Lagos: Spectrum Book Limited.
- Mehrens, W. A (1969). Standardized Test in Education New York: Hot, Rinehart & Winston.
- (1976). Reading in measurement and evaluation in Education. Holt, Rinhart and Winston: New York.
- ; Lehmann,I.J.(1984). Measurement and evaluation in Education and Psychology. New York: Holt, Rinhart and Winston

- MOE (1986) "Evaluation Research of the general education system in Ethiopia" (ERGSE) A.A.
- (1986). የትምህርት ቤት የውስጥ አመራርና አስተዳደር መመሪያ፡፡ አዲስ አበባ ትምህርት ሚኒስቴር፤ በት.መ.ማ.ማ. ድርጅት ታተመ፡፡
- NACE (2006). "Career Development and Job Search Advice for New College Graduates" National Association of College and Employers. Retrieved on Dec 3, 2006. <http://www.naceweborganization> .
- NCTE Guideline. (1987) "Statement on Class Size and Teacher work load: College" National Council of Teachers of English. Retrieved on Dec 3, 2006 <http://books.google.com>.
- NOE (2004) Ethiopian Secondary National Learning Assessment of Grade 8 Students. Ethiopia: Publisher, USAID/ETHIOPIA.
- OERI (2004). "Help Your child Improve test Taking "Office of educational Research & Improvement (OERI), Retrieved 2/16/2006 <http://www.pmkmonkey.com>
- Ogunnivi, M.B (1984) Educational Measurement and Evaluation. Lagos: Longman.
- Ross, C.C and Stanley, J.C (1954). Measurement in today's school EngleWool Cliff N.J: Prentice -Hall, Inc.
- Sax, Gilbert. (1974). Principles of Educational Measurement & Evaluation California: Wadsworth, Publishing Co., Inc., Belmont..
- (1989) Study guide for principle of Educational & Psychological measurement and evaluates. 3rd ed. California: Wadsworth Publishing Co.
- Stinnett, T.M. (1968). Professional problems of Teachers. 3rd ed, New York: The Macmillan Co.
- Taiwo, A.A. (1995). Fundamentals of classroom testing. Delhi: Vikan publishing House PLC.
- Teka Woldie (1988) "The role of measurement and evaluation on Making Educational Decision" (Mimeographed, Addis Ababa Kotebe College of Teachers Education).

- Thoridike, Robert Ladd (1977). Educational measurement. Washington: American Council on Education.
- Wiersma, W. (1985). Educational Measurement and Testing. University of Teledo, Toronto: Allyn and Bacon, Inc.
- Wilkinson, Jenny; Beavis, A; Ingvarson, L; Kleinhenz, E. (2005). Primary Teacher Work Study Report. Australian Council for Educational Research.
- Zurawski, Raymond M. (2001). "Making the most of Exams: Procedures for Item analysis." St. Norbert College. Retrieved from <http://www.FeaturedArticle.htm>.

Appendix 1

Addis Ababa University Department of Educational Psychology

This questionnaire is designed to get relevant information regarding exam related issues and school based test (exam) constraints from 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: M ☐ F ☐
2. Age (a) below 25 ☐ (b) 26-35 ☐ (c) 36-45 ☐ (d) 46 & above ☐
3. Years of experience as a teacher -----
a) Below 6 ☐ b) 6-15 ☐ c) 16-25 ☐ d) above 25 ☐
4. The subject you are teaching -----
5. Qualification
(a) Diploma ☐ (b) Bachelor Degree ☐ (c) Master Degree & above ☐
(d) other Specify -----
6. College or university you have graduated from:

7. Have you taken teacher training courses?
7.1 yes ☐ No ☐

7.1.1. If No, Have you had a chance of any training in measurement & evaluation? Yes ☐ No ☐

7.1.2 If yes, did you have training or courses of measurement & evaluation in your college or university education?
Yes ☐ No ☐

7.1.2.1 If yes? _____

Are you practicing the skill of educational measurement in sufficient detail in your teaching?

Yes ☐ No ☐

If No, state your main reasons _____

8. Do you take some sentence, expression or phrases directly out of the students' textbook and use them in your exam (test)?

Yes ☐ No ☐

9. When do you usually prepare your tests or exams?

- ☐ Before the course (lesson) is given.
- ☐ During giving the course (lesson).
- ☐ When the testing is approaching.
- ☐ Immediately before the day of the test.

10. Have you ever evaluated your test or test items after the test is given to the students?

Yes ☐ No ☐

10 a. If yes, how often?

(a) Always ☐ (b) sometime ☐ (c) one time ☐

(d) Other specify -----

11. What is the average class size you teach?

(a) Below 40 ☐ (b) 41 - 55 ☐ (c) 56 - 70 ☐

(d) 71- 85 ☐ (e) above 85 ☐

12. What are the numbers of teaching periods per week?

(a) Below 15 ☐ (b) 16-20 ☐ (c) 21-25 ☐

(d) 26-30 ☐ (e) above 30 ☐

13. How many tests do you give within a semester?

(a) 2 ☐ (b) 3 ☐ (c) 4 ☐ (d) 5 or more ☐

14. How often are you required to give information regarding you students or
the lesson you teach to your immediate school officials?

(a) Every week ☐ (b) every month ☐ (c) every semester ☐

(d) if other specify -----

15. In your subject, the numbers for a particular section per a week are:

a) 1 ☐ b) 2 ☐ c) 3 ☐ d) 4 ☐

e) 5 ☐ f) 6 or more ☐

16. How long is one period ?

(a) 30 min ☐ (b) 40 min ☐ (c) 45 min ☐

(d) 50 min ☐ (e) 60 min ☐

15. How often do you discuss with a particular parent about his/ her child?
 (a) Often ☐ (b) sometimes ☐
 (c) rarely ☐ (d) no discussion at all ☐
16. In what activities are you involved in school other than teaching?
 (a) Chatting with friends ☐ (b) preparing lesson ☐
 (c) club or committee activity ☐
 (d) scoring test papers ☐ (e) if other specify -----
17. How many free periods do you have in a week?
 (a) Less than 4 ☐ (b) 4-8 ☐ (c) 9- 12 ☐
 (d) 13-16 ☐ (e) if other specify-----
18. Is there time, when you work outside the working hours?
 Yes ☐ No ☐
19. If yes, at what time?
 (a) Morning or afternoon ☐ (b) evening time ☐
 (c) weekend ☐ (d) if other specify -----
20. Do you participate in extra curricular duties?
 Yes ☐ No ☐
21. If yes, in how many clubs or committees?
 a) 1 ☐ b) 2 ☐ c) 3 ☐ d) 4 ☐ e) 5 ☐
 If other specify _____
22. What is your role in a club or committee you are involved in?
 a. coordinator ☐ b. member ☐ c. if other specify _____
23. What must be done by the school management so that teachers can

evaluate their exam after they have administered them?

1. _____
2. _____
3. _____

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

1. _____
2. _____
3. _____

POINTS OBSERVED IN EACH TEACHER-MADE TEST

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

		S1	S2	S3	S4	S5
		Y N	YN	YN	YN	YN
1. More than two item formats.	☐	☐	☐	☐	☐	☐
2. Clearly stated instruction.	☐	☐	☐	☐	☐	☐
3. Stating time allotted.	☐	☐	☐	☐	☐	☐
4. Stating the weight for each item.	☐	☐	☐	☐	☐	☐
5. Clarity and neatness of the test paper.	☐	☐	☐	☐	☐	☐

In multiple choice

6. Keeping the length of all option uniform.	☐	☐	☐	☐	☐
7. Avoiding negatives specially double negatives.	☐	☐	☐	☐	☐
8. Avoiding grammatical clue.	☐	☐	☐	☐	☐
9. Not repeating the key words from item in the options.	☐	☐	☐	☐	☐
10. Are the distracters plausible?	☐	☐	☐	☐	☐

In true- false

11. Making the statement clear (free of ambiguity).	☐	☐	☐	☐	☐
12. Avoid specific determiners like always, never, all and non.	☐	☐	☐	☐	☐
13. Making the length and number of True & False statements about the same.	☐	☐	☐	☐	☐

In Matching

14. The items in them are homogenous so that the option can be plausible.	☐	☐	☐	☐	☐
15. The length of the column should be unequal, and the response should be longer from the premises side.	☐	☐	☐	☐	☐

In supply item or short answer item format

16. Does the test constructors specify
they degree of precision and unit of
expression expected in answer? ☐ ☐ ☐ ☐ ☐
17. Not having many blanks to occur
at the end of the sentence. ☐ ☐ ☐ ☐ ☐
18. Avoid excessive blanks in a single item. ☐ ☐ ☐ ☐ ☐
19. All blanks should be equal regardless
of the length of the answer. ☐ ☐ ☐ ☐ ☐

For the test in general

20. Are the items arranged from
simple to complex? ☐ ☐ ☐ ☐ ☐

Interview Questions Directed to Principals

1. Is there any manual prepared by the Ministry of Education, regarding test (exam) related issues in your school?
2. Is there an examination and evaluation committee in your school?
3. Who are the members of the committee? What are the criteria to be the member of the committee?
4. What is expected from the committee regarding teacher made tests (exam)?
5. Does the committee report to teachers how effective the tests they have prepared are?
6. What effort is made to improve teacher made tests by the school management?
7. What role is expected from teachers other than teaching in classroom?
8. Do you think teachers have enough time skill and knowledge to analyze and evaluate their test?
9. What about is using educational technologies that facilitate teaching- learning, and simplify teachers' burden, innovate teachers with?
10. Do you think there is good working r/n ship between

administration and teacher, and among teachers themselves?

11. What about teachers attrition rate is your school? What do you think are the main reasons for leaving their profession?
12. What effort you made to minimize the attrition rate of teachers? if there are any.

APPENDIX –2

I) Tests Item Analysis:

Table I
ABYOT KIRS FIRST SEMESTER FINAL TEST
ITEM ANALYSIS

ENGLISH

MATHEMATICS

ITEM NO.	DIFFL. INDEX	REMARK	DISCR. INDEX	REMARK	ITEM NO.	DIFFL. INDEX	REMARK	DISCR. INDEX	REMARK
1	0.71		0.04	Not acceptable	1	0.59		0.59	
2	0.75		0.12	Not acceptable	2	0.35		0.19	
3	0.87		0.12	Not acceptable	3	0.78		0.37	
4	0.81		0.31		4	0.65		0.33	
5	0.54		0.23		5	0.89		0.15	
6	0.69		0.54		6	0.57		0.63	
7	0.46		0.85		7	0.06	V.Difficult	0.04	Not acceptable
8	0.65		0.69		8	0.80		0.04	Not acceptable
9	0.54		0.69		9	0.50		0.56	
10	0.69		0.46		10	0.20	V.Difficult	0.11	Not acceptable
11	0.75		0.50		11	0.06	V.Difficult	0.04	Not acceptable
12	0.73		0.54		12	0.96	V.EASY	0.00	Not acceptable
13	0.42		0.75		13	0.37		0.30	

14	019	V. Difficult	0.00	Not acceptable	14	0.96	V.EASY	0.00	Not acceptable
15	0.27	V.Difficu lt	0.15		15	0.76		0.26	
16	0.44		0.04	Not acceptable	16	0.22	V.Difficult	0.07	Not acceptable
17	0.50		0.69		17	0.43		0.11	Not acceptable
18	0.46		0.08	Not acceptable	18	0.96		0.00	Not acceptable
19	0.12	V.Difficu lt	0.00	Not acceptable	19	0.11	V.Difficult	0.15	
20	0.31		0.08	Not acceptable	20	0.61		0.56	
21	0.85		0.15		21	0.06		0.04	Not acceptable
22	0.54		0.69		22	0.61		0.48	
23	0.48		0.42		23	0.32		0.63	
24	0.63		0.58		24	0.61		0.63	
25	0.62		0.69		25	0.59		0.67	
26	0.62		0.46		26	0.51		0.81	
27	0.58		0.75		27	0.51		0.74	
28	0.83		0.19		28	0.48		0.89	
29	0.06	V. Difficult	0.04	Not acceptable	29	0.50		0.85	
30	0.23	V. Difficult	0.00	Not acceptable	30	0.67		0.59	
31	0.37		0.50		31	0.11	V.Difficult	0.15	

32	0.42		0.75		32	0.59		0.74	
33	0.56		0.58						
34	0.60		0.58						
35	0.58		0.54						
36	0.85		0.31						
37	0.58		0.75						
38	0.50		0.61						
39	0.65		0.69						
40	0.67		0.50						

Table II

DELACHEN FIRST SEMESTER FINAL TEST

ENGLISH

MATHEMATICS

ITEM NO.	DIFFI. INDEX	REMARK	DISCR. INDEX	REMARK	ITEM NO.	DIFFI. INDEX	REMARK	DISCR. INDEX	REMARK
1	0.48		-0.12	Not Acceptable	1	0.66		0.05	Not Acceptable
2	0.87		0.08	Not Acceptable	2	0.58		0.00	Not Acceptable
3	0.69		0.20		3	0.39		0.05	Not Acceptable
4	0.69		0.46		4	0.26	V.Difficult	0.21	

5	0.38		0.33		5	0.63		0.42	
6	0.60		0.29		6	0.37		0.42	
7	0.87		0.25		7	0.74		0.21	
8	0.71		0.50		8	0.42		0.42	
9	0.21	V.Difficult	0.17		9	0.37		0.42	
10	0.25	V.Difficult	0.25		10	0.33		0.32	
11	0.60		0.54 [/]		11	0.37		0.53	
12	0.65		0.46		12	0.36		0.37	
13	0.77		0.38		13	0.47		0.63	
14	0.60		0.20		14	0.36		0.47	
15	0.02	V.Difficult	-0.04	Not Acceptable	15	0.08	V.Difficult	0.16	
16	0.75		0.00	Not Acceptable	16	0.24	V.Difficult	0.26	
17	1.00	V.Easy	0.00	Not Acceptable	17	0.16	V.Difficult	0.00	Not Acceptable
18	0.58		0.83		18	0.42		0.53	
19	0.50		0.33		19	0.39		0.58	
20	0.58		0.58		20	0.37		0.42	
21	0.42		0.33		21	0.33		0.53	

22	1.00	V.Easy	0.00	Not Acceptable	22	0.37		0.63	
23	0.48		0.54		23	0.24	V.Difficult	0.37	
24	0.44		0.46		24	0.13	V.Difficult	0.16	
25	0.54		0.67		25	0.36		0.26	
26	0.71		0.58		26	0.53		0.42	
27	0.66		0.50		27	0.18	V.Difficult	0.37	
28	0.65		0.04	Not Acceptable	28	0.11	V.Difficult	0.21	
29	0.85		0.29		29	0.13	V.Difficult	0.26	
30	0.63		0.42		30	0.05	V.Difficult	0.11	Not Acceptable
31	0.19	V.Difficult	0.20		31	0.13	V.Difficult	0.26	
32	0.10	V.Difficult	0.20						

Table III

LIDETA CATHOLIC CATHEDRAL FIRST SEMESTER FINAL TEST

ITEM ANALYSIS

ENGLISH

MATHEMATICS

ITEM NO.	DIFFL. INDEX	REMARK	DISCR. INDEX	REMARK	ITEM NO.	DIFFL. INDEX	REMARK	DISCR. INDEX	REMARK
1	0.45		0.37		1	0.61		0.67	
2	0.63		0.11	Not Acceptable	2	0.58		0.72	
3	0.55		0.16		3	0.61		0.11	Not Acceptable
4	0.53		0.53		4	0.47		0.39	
5	0.66		0.16		5	0.42		0.72	
6	0.16	V.Difficult	0.21		6	0.61		0.67	
7	0.95		0.00	Not Acceptable	7	0.55		0.78	
8	0.64		0.00	Not Acceptable	8	0.78		0.22	
9	0.73		0.37		9	0.61		0.33	
10	0.84		0.21		10	0.44		0.78	

11	0.63		0.42		11	0.28	V.Difficult	0.22	
12	0.82		0.16		12	0.47		0.28	
13	0.18	V.Difficult	-0.16	Not Acceptable	13	0.31		0.39	
14	0.58		0.11	Not Acceptable	14	0.28	V.Difficult	0.44	
15	0.32		-0.42	Not Acceptable	15	0.55		0.44	
16	0.34		0.16		16	0.58		0.61	
17	0.71		0.37		17	0.61		0.67	
18	0.58		0.32		18	0.42		0.72	
19	0.87		0.16		19	0.53		0.25	
20	0.74		0.21		20	0.53		0.72	
21	0.16	V.Difficult	0.11	Not Acceptable	21	0.69		0.39	
22	0.79		0.32		22	0.25	V.Difficult	0.17	
23	0.84		0.31		23	0.55		0.67	
24	0.73		0.16		24	0.61		0.44	
25	0.18	V.Difficult	0.16		25	0.33		0.56	
26	0.45		0.37		26	0.19	VDifficult	0.17	

27	0.11	V.Diffic ult	0.11	Not Acceptable	27	0.50		0.59	
28	0.13	V.Diffic ult	-0.05	Not Acceptable	28	0.36		0.61	
29	0.79		0.32		29	0.36		0.39	
30	0.73		0.26		30	0.28		0.44	
31	0.92		0.05	Not Acceptable					
32	0.84		0.00	Not Acceptable					
33	0.39		0.47						
34	0.61		0.37						
35	0.26	V.Diffic ult	0.11	Not Acceptable					
36	0.45		0.26						
37	0.53		0.42						
38	0.66		0.58						
39	0.95		0.11	Not Acceptable					
40	0.64		0.53						
41	0.55		0.37						

42	0.50		0.37						
43	0.24	V.Diffic ult	-0.16	Not Acceptable					
44	0.79		0.32						
45	0.29	V.Diffic ult	0.16						
46	0.50		0.26						
47	0.58		0.32						
48	0.87		0.26						
49	0.84		0.32						
50	0.63		0.53						
51	0.47		0.53						
52	0.47		0.53						
53	0.47		0.53						
54	0.34		0.58						
55	0.73		0.37						
56	0.74		0.53						
57	0.64		0.53						
58	0.58		0.63						

SHIMELIS HABTE FIRST SEMESTER FINAL TEST

ITEM ANALYSIS

Table IV

ENGLISH

MATHEMATICS

ITEM NO.	DIFFI. INDEX	REMARK	DISCR. INDEX	REMARK	ITEM NO.	DIFFI. INDEX	REMARK	DISCR. INDEX	REMARK
1	0.87	V.Easy	0.19		1	0.85		0.30	
2	0.12	V. Difficult	0.00	Not acceptable	2	0.76		0.48	
3	0.90	V. Easy	0.12	Not acceptable	3	0.63		0.33	
4	0.58		0.15		4	0.15	V.Difficult	0.15	
5	0.83		0.12	Not acceptable	5	0.56		0.07	
6	0.92	V.Easy	0.08	Not acceptable	6	0.61		0.78	
7	0.56		0.58		7	0.72		0.56	
8	0.63		0.42		8	0.67		0.66	
9	0.15	V.Difficult	0.15		9	0.76		0.48	
10	0.75		0.27		10	0.80		0.41	

11	0.25	V.Difficult	0.12	Not acceptable	11	0.59		0.81	
12	0.19	V.Difficult	0.23		12	0.80		0.41	
13	0.94	V. Easy	0.04	Not acceptable	13	0.52		0.52	
14	0.84		0.12	Not acceptable	14	0.74		0.44	
15	0.54		0.31		15	0.61		0.33	
16	0.23	V.Difficult	0.15		16	0.07	V.Difficult	0.00	Not acceptable
17	0.40		0.42		17	0.63		0.66	
18	0.54		0.69		18	0.59		0.74	
19	0.56		0.27		19	0.72		0.56	
20	0.83		0.35		20	0.59		0.74	
21	0.27	V.Difficult	0.23		21	0.05	V.Difficult	0.11	Not acceptable
22	0.37		0.42		22	0.13	V.Difficult	0.11	Not acceptable
23	0.33		0.50		23	0.17	V.Difficult	0.19	
24	0.46		0.38		24	0.05	V.Difficult	0.11	Not acceptable

25	0.94	V.EASY	0.12	Not acceptable	25	0.02	V.Difficult	0.04	Not acceptable
26	0.77		0.31		26	0.02	V.Difficult	0.04	Not acceptable
27	0.65		0.46		27	0.02	V.Difficult	0.04	Not acceptable
28	0.23	V.Difficult	0.23		28	0.02	V.Difficult	0.04	Not acceptable
29	0.40		-0.12	Not acceptable	29	0.02	V.Difficult	0.04	Not acceptable
30	0.78		-0.04	Not acceptable	30	0.02	V.Difficult	0.04	Not acceptable

ASSAI FIRST SEMESTER FINAL TEST
ITEM ANALYSIS

ENGLISH

MATHEMATICS

ITEM NO.	DIFFI. INDEX	REMARK	DISCR. INDEX	REMARK	ITEM NO.	DIFFI. INDEX	REMARK	DISC R. IND EX	REMARK
1	0.85		0.31		1	0.42		0.85	
2	0.50		0.38		2	0.69		0.46	
3	0.65		0.54		3	0.85		0.00	Not acceptable
4	0.96	V.Easy	-0.08	Not acceptable	4	0.58		- 0.08	Not acceptable
5	0.96	V.Easy	0.08	Not acceptable	5	0.62		0.46	
6	0.77		0.15		6	0.50		- 0.23	Not acceptable
7	0.77		0.31		7	0.73		- 0.08	Not acceptable
8	0.62		0.15	Not acceptable	8	0.58		0.08	Not acceptable
9	0.73		0.38		9	0.58		0.23	
10	0.38		0.62		10	0.85		0.15	

11	0.50		0.25		11	0.69		0.31	
12	0.54		0.62		12	0.69		0.62	
13	0.62		0.46		13	0.31		0.15	
14	0.69		0.46		14	0.46		0.62	
15	0.81		0.38		15	0.58		0.54	
16	0.77		0.31		16	0.54		0.00	Not acceptable
17	0.62		0.46		17	0.46		0.46	
18	0.54		0.31		18	0.62		0.46	
19	0.81		0.25		19	0.38		0.31	
20	0.81		0.25		20	0.23		0.15	Not acceptable
21	0.69		0.62		21	0.42		0.38	
22	0.81		0.25		22	0.35		0.54	
23	0.69		0.62		23	0.38		0.15	
24	0.15	V.Difficult	0.15		24	0.62		0.46	
25	0.96	V.Easy	0.08	Not acceptable	25	0.54		0.15	Not acceptable
26	0.65		0.69		26	0.38		0.31	
27	0.81		0.38		27	0.27	V.Difficult	0.38	

28	0.81		0.38		28	0.50		0.54	
29	0.62		0.46		29	0.38		0.31	
30	0.88		0.08		30	0.19	V.Difficult	0.23	
31	0.73		0.54		31	0.38		0.31	
32	0.50		0.23		32	0.50		- 0.08	Not acceptable
33	0.73		0.38		33	0.38		0.31	
34	0.65		0.38		34	0.23	V.Difficult	0.15	
35	0.96	V.Easy	0.08	Not acceptable	35	0.46		0.31	
36	0.46		0.15		36	0.39		0.46	
37	0.73		0.08	Not acceptable	37	0.62		0.62	
38	0.62		0.62		38	0.23	V.Difficult	0.31	
39	0.54		0.31		39	0.27	V.Difficult	0.54	
40	0.27	V.Difficult	0.08	Not acceptable	40	0.04	V.Difficult	0.08	Not acceptable
41	0.88		0.23		41	0.35		0.54	
42	0.92	V.Easy	0.15		42	0.27	V.Difficult	0.54	
43	0.46		0.46		43	0.15	V.Difficult	0.31	

44	0.92	V.easy	0.15		44	0.12	V.Difficult	0.08	Not acceptable
45	0.54		0.62		45	0.04	V.Difficult	0.08	Not acceptable
46	0.35		0.38						
47	0.62		0.62						
48	0.58		0.69						
49	0.62		0.62						
50	0.38		0.72						
51	0.23	VDifficult	0.31						
52	0.27	V.Difficult	0.54						
53	0.23	V.Difficult	0.46						
54	0.50		0.54						
55	0.31		0.46						
56	0.27	V.Difficult	0.54						
57	0.38		0.62						
58	0.38		0.62						
59	0.30		0.46						
60	0.38		0.60						

Appendix-3

Correlations of the National Exam scores with Teacher-Made tests

r_{xy} is the correlation between the national exam and teacher-made tests by using Pearson Product moment correlation formula.

Where “X” 8th grade National Examination scores and “Y” is 9th grade first semester average score all subjects given

Assai Public Secondary School

Code of students	X	Y	XY	X ²	Y ²	
1	51	51	2601	2601	2601	
2	60	72	4320	3600	5184	
3	66	62.9	4151.4	4356	3956.41	
4	66	56.8	3748.8	4356	3226.24	
5	69	59.1	4077.9	4761	3492.81	
6	66	64.7	4270.2	4356	4186.09	
7	72	64.5	4644	5184	4160.25	
8	67	77.3	5179.1	4489	5975.29	
9	74	67.3	4980.2	5476	4529.29	
10	68	63.3	4304.4	4624	4006.89	

11	54	59.1	3191.4	2916	3492.81	
12	65	69.3	4504.5	4225	4802.49	
13	69	51.2	3532.8	4761	2621.44	
14	72	66.5	4788	5184	4422.25	
15	68	73.5	4998	4624	5402.25	
16	70	72.4	5068	4900	5241.76	
17	74	69.3	5128.2	5476	4802.49	
18	69	61.7	4257.3	4761	3806.89	
19	87	93.8	8160.6	7569	8798.44	
20	73	65.8	4803.4	5329	4329.64	
21	74	86.3	6386.2	5476	7447.69	
22	69	64.5	4450.5	4761	4160.25	
23	74	88.3	6534.2	5476	7796.89	
24	62	62.4	3868.8	3844	3893.76	
25	80	70.3	5624	6400	4942.09	
26	72	63.2	4550.4	5184	3994.24	
27	62	69.3	4296.6	3844	4802.49	
28	61	68.6	4184.6	3721	4705.96	
29	58	71.7	4158.6	3364	5140.89	
30	50	69.6	3480	2500	4844.16	

31	60	67.9	4074	3600	4610.41	
32	67	59.6	3993.2	4489	3552.16	
33	65	49.9	3243.5	4225	2490.01	
34	70	58.7	4109	4900	3445.69	
35	66	71.5	4719	4356	5112.25	
36	65	58.9	3828.5	4225	3469.21	
37	65	53.1	3451.5	4225	2819.61	
38	76	83.5	6346	5776	6972.25	
39	74	56.4	4173.6	5476	3180.96	
40	70	75.6	5292	4900	5715.36	
41	68	57.4	3903.2	4624	3294.76	
42	75	78.8	5910	5625	6209.44	
43	64	60.9	3897.6	4096	3708.81	
44	66	60	3960	4356	3600	
45	65	56.6	3679	4225	3203.56	
46	57	64.6	3682.2	3249	4173.16	

Σ 3095 3049.1 206505.4 210465 206324.79

$\Delta X= 26$ $\Delta Y= 43$

$r_{xy}= 0.4420568$

Lideta Catholic Cathedral Secondary School

Code of students	X	Y	XY	X ²	Y ²	
1	80	57.6	4608	6400	3317.76	
2	61	53.3	3251.3	3721	2840.89	
3	66	54.2	3577.2	4356	2937.64	
4	80	59.2	4736	6400	3504.64	
5	81	64.9	5256.9	6561	4212.01	
6	70	60.2	4214	4900	3624.04	
7	80	56.7	4536	6400	3214.89	
8	85	80	6800	7225	6400	
9	70	62.1	4347	4900	3856.41	
10	84	72	6048	7056	5184	
11	79	58.3	4605.7	6241	3398.89	
12	79	69.6	5498.4	6241	4844.16	
13	71	72.7	5161.7	5041	5285.29	
14	84	71	5964	7056	5041	
15	63	66	4158	3969	4356	
16	66	52.3	3451.8	4356	2735.29	

18	73	58.1	4241.3	5329	3375.61	
19	61	49.6	3025.6	3721	2460.16	
20	76	54.4	4134.4	5776	2959.36	
21	77	52.1	4011.7	5929	2714.41	
22	60	59.5	3570	3600	3540.25	
23	59	48.3	2849.7	3481	2332.89	
24	79	65	5135	6241	4225	
26	63	60	3780	3969	3600	
27	75	51.7	3877.5	5625	2672.89	
28	72	62.4	4492.8	5184	3893.76	
29	82	88	7216	6724	7744	
30	50	64	3200	2500	4096	
31	67	58.3	3906.1	4489	3398.89	
32	71	60	4260	5041	3600	
33	65	68.4	4446	4225	4678.56	
34	72	69.9	5032.8	5184	4886.01	
35	71	75.2	5339.2	5041	5655.04	
36	77	65	5005	5929	4225	
37	59	70.3	4147.7	3481	4942.09	
38	67	72.9	4884.3	4489	5314.41	

39	75	78	5850	5625	6084	
40	73	80	5840	5329	6400	
41	78	72	5616	6084	5184	
43	76	63.5	4826	5776	4032.25	
44	74	58.1	4299.4	5476	3375.61	
45	66	65.2	4303.2	4356	4251.04	
46	81	58.6	4746.6	6561	3433.96	
47	74	56.3	4166.2	5476	3169.69	
48	64	66.4	4249.6	4096	4408.96	
49	73	76.2	5562.6	5329	5806.44	
50	85	80	6800	7225	6400	
51	81	73.5	5953.5	6561	5402.25	
52	78	71	5538	6084	5041	
53	86	90	7740	7396	8100	
54	54	64	3456	2916	4096	
55	60	68	4080	3600	4624	
56	90	88	7920	8100	7744	
57	78	67.6	5272.8	6084	4569.76	
59	63	75	4725	3969	5625	
60	60	61.6	3696	3600	3794.56	

61	64	67.3	4307.2	4096	4529.29	
63	60	61.6	3696	3600	3794.56	
65	59	73	4307	3481	5329	
66	77	79.5	6121.5	5929	6320.25	
67	76	65.3	4962.8	5776	4264.09	
68	63	52.2	3288.6	3969	2724.84	
69	78	71.4	5569.2	6084	5097.96	

Σ **4521** **4146.5** **299662.3** **329359** **278669.75**

$\Delta X=40$ $\Delta Y=40$

$r_{xy}=0.3947169$

ABYOT KIRS SECONDARY SCHOOL

Code of students	X	Y	XY	X²	Y²	
1	50	69.2	3460	2500	4788.64	
2	55	64.3	3536.5	3025	4134.49	
3	42	61.7	2591.4	1764	3806.89	
4	54	62	3348	2916	3844	
6	56	59	3292.8	3136	3457.44	
7	52	53.1	2761.2	2704	2819.61	

8	43	40.5	1741.5	1849	1640.25	
9	45	68.8	3096	2025	4733.44	
10	61	81.1	4947.1	3721	6577.21	
11	52	62.4	3244.8	2704	3893.76	
12	54	61.7	3331.8	2916	3806.89	
13	47	59.4	2791.8	2209	3528.36	
17	46	75.3	3463.8	2116	5670.09	
18	43	66.6	2863.8	1849	4435.56	
19	37	89.6	3315.2	1369	8028.16	
20	50	61.3	3065	2500	3757.69	
22	48	68.8	3302.4	2304	4733.44	
23	42	68	2856	1764	4624	
24	66	80.6	5319.6	4356	6496.36	
26	63	74.1	4668.3	3969	5490.81	
27	51	73.5	3748.5	2601	5402.25	
28	51	68.4	3488.4	2601	4678.56	
29	39	62.9	2453.1	1521	3956.41	
31	45	71.2	3204	2025	5069.44	
32	47	60.8	2857.6	2209	3696.64	
33	52	70.2	3650.4	2704	4928.04	

34	53	59.8	3169.4	2809	3576.04	
36	60	64.1	3846	3600	4108.81	
37	60	62.5	3750	3600	3906.25	
38	50	60.8	3040	2500	3696.64	
39	54	67.9	3666.6	2916	4610.41	
40	51	66.8	3406.8	2601	4462.24	
41	52	72.5	3770	2704	5256.25	
42	54	75.2	4060.8	2916	5655.04	
43	49	70.5	3454.5	2401	4970.25	
44	54	62.7	3385.8	2916	3931.29	
45	45	68.6	3087	2025	4705.96	
48	50	61.8	3090	2500	3819.24	
49	46	53.6	2465.6	2116	2872.96	
50	69	84.3	5816.7	4761	7106.49	
51	45	66.9	3010.5	2025	4475.61	
52	66	89.3	5893.8	4356	7974.49	
53	50	71	3550	2500	5041	
54	43	66.3	2850.9	1849	4395.69	
55	47	74.2	3487.4	2209	5505.64	
56	47	61.2	2876.4	2209	3745.44	

57	44	62.9	2767.6	1936	3956.41	
58	50	61.1	3055	2500	3733.21	
59	42	64.3	2700.6	1764	4134.49	
60	44	73.8	3247.2	1936	5446.44	
61	50	66.7	3335	2500	4448.89	
62	45	60.7	2731.5	2025	3684.49	
63	47	72.4	3402.8	2209	5241.76	
64	54	73.4	3963.6	2916	5387.56	
65	49	65.7	3219.3	2401	4316.49	
66	44	63	2772	1936	3969	
67	44	63.5	2794	1936	4032.25	
68	43	67.3	2893.9	1849	4529.29	
69	46	72.8	3348.8	2116	5299.84	
70	44	62.9	2767.6	1936	3956.41	
71	47	64	3008	2209	4096	
72	45	63.6	2862	2025	4044.96	
73	46	68	3128	2116	4624	
74	50	69.4	3470	2500	4816.36	
75	44	68.6	3018.4	1936	4705.96	
76	45	59.8	2691	2025	3576.04	

77	44	66	2904	1936	4356	
78	41	64.6	2648.6	1681	4173.16	
79	46	68.1	3132.6	2116	4637.61	
80	49	60.8	2979.2	2401	3696.64	
81	49	63.2	3096.8	2401	3994.24	
82	45	57.7	2596.5	2025	3329.29	
83	41	58.8	2410.8	1681	3457.44	
84	50	60.7	3035	2500	3684.49	
86	55	56.8	3124	3025	3226.24	
87	50	70.8	3540	2500	5012.64	
88	58	67.2	3897.6	3364	4515.84	
89	54	73.3	3958.2	2916	5372.89	
90	41	56.8	2328.8	1681	3226.24	
91	49	63	3087	2401	3969	
92	56	73.6	4121.6	3136	5416.96	
93	56	59.6	3337.6	3136	3552.16	

Σ

4043

5439.2

269451.8

202541

365434.86

ΔX=29

ΔY=41

r_{xy}=

0.330096

Dilachin Secondary School

Code of students	X	Y	XY	X ²	Y ²	
1	56	62.4	3494.4	3136	3893.76	
2	53	64.6	3423.8	2809	4173.16	
3	53	60.3	3195.9	2809	3636.09	
4	51	68	3468	2601	4624	
5	60	59.5	3570	3600	3540.25	
6	66	59.2	3907.2	4356	3504.64	
7	52	59.8	3109.6	2704	3576.04	
8	41	57	2337	1681	3249	
9	46	63.6	2925.6	2116	4044.96	
10	42	62.9	2641.8	1764	3956.41	
11	51	57.2	2917.2	2601	3271.84	
12	66	64.9	4283.4	4356	4212.01	
13	51	64.8	3304.8	2601	4199.04	
14	51	68.1	3473.1	2601	4637.61	
15	51	53.6	2733.6	2601	2872.96	
16	66	76.4	5042.4	4356	5836.96	
18	61	55.5	3385.5	3721	3080.25	
19	53	66.8	3540.4	2809	4462.24	

20	44	68.7	3022.8	1936	4719.69	
21	49	62	3038	2401	3844	
22	56	70.2	3931.2	3136	4928.04	
23	54	23	1242	2916	529	
24	50	50.8	2540	2500	2580.64	
25	41	62.8	2574.8	1681	3943.84	
27	63	67.4	4246.2	3969	4542.76	
28	52	73.1	3801.2	2704	5343.61	
29	67	65.2	4368.4	4489	4251.04	
30	43	65.4	2812.2	1849	4277.16	
31	50	67.6	3380	2500	4569.76	
32	48	77.8	3734.4	2304	6052.84	
34	50	67.8	3390	2500	4596.84	
35	63	74.8	4712.4	3969	5595.04	
36	43	59.4	2554.2	1849	3528.36	
37	49	63.3	3101.7	2401	4006.89	
38	54	67.7	3655.8	2916	4583.29	
39	60	62.7	3762	3600	3931.29	
40	71	89.1	6326.1	5041	7938.81	
41	47	61	2867	2209	3721	
43	52	67.6	3515.2	2704	4569.76	
44	62	63.5	3937	3844	4032.25	
45	69	65.9	4547.1	4761	4342.81	
46	57	85.5	4873.5	3249	7310.25	
47	47	71.5	3360.5	2209	5112.25	

48	46	65.7	3022.2	2116	4316.49	
49	47	61.8	2904.6	2209	3819.24	
50	49	68.6	3361.4	2401	4705.96	
51	45	57.4	2583	2025	3294.76	
52	52	66.7	3468.4	2704	4448.89	
53	45	61.7	2776.5	2025	3806.89	
55	54	65.5	3537	2916	4290.25	
56	59	68.1	4017.9	3481	4637.61	
57	61	71.8	4379.8	3721	5155.24	
58	58	81.2	4709.6	3364	6593.44	
59	50	67.8	3390	2500	4596.84	
60	52	60.5	3146	2704	3660.25	
62	48	62	2976	2304	3844	
64	54	63.5	3429	2916	4032.25	
65	46	58.6	2695.6	2116	3433.96	
66	45	67.4	3033	2025	4542.76	
67	53	60.5	3206.5	2809	3660.25	
68	66	71.9	4745.4	4356	5169.61	
69	59	69.6	4106.4	3481	4844.16	
70	53	58	3074	2809	3364	
71	41	70.9	2906.9	1681	5026.81	
72	53	57.3	3036.9	2809	3283.29	
73	53	54.6	2893.8	2809	2981.16	
74	60	69.2	4152	3600	4788.64	
75	82	67.2	5510.4	6724	4515.84	

76	45	66.6	2997	2025	4435.56	
77	42	50.9	2137.8	1764	2590.81	
78	50	75.1	3755	2500	5640.01	
79	63	82	5166	3969	6724	
80	78	89	6942	6084	7921	
81	53	63.5	3365.5	2809	4032.25	

$$\Sigma \quad 3973 \quad 4833 \quad 261473 \quad 218615 \quad 321778.66$$

$$\Delta X= 37 \quad \Delta Y=66$$

$$r_{xy}= \quad 0.3493788$$

Shimels Habte Secondary School

Code of students	X	Y	XY	X ²	Y ²	
1	51	65.8	3355.8	2601	4329.64	
2	60	80.9	4854	3600	6544.81	
3	51	62.2	3172.2	2601	3868.84	
4	48	64.2	3081.6	2304	4121.64	
6	52	63.9	3322.8	2704	4083.21	
7	45	64.7	2911.5	2025	4186.09	
8	54	70.3	3796.2	2916	4942.09	
9	49	64.8	3175.2	2401	4199.04	

10	58	50.9	2952.2	3364	2590.81	
11	47	59	2773	2209	3481	
14	55	70.8	3894	3025	5012.64	
15	56	66.7	3735.2	3136	4448.89	
16	46	61.2	2815.2	2116	3745.44	
17	48	63.4	3043.2	2304	4019.56	
18	68	83	5644	4624	6889	
19	48	55.3	2654.4	2304	3058.09	
20	47	68.5	3219.5	2209	4692.25	
21	45	64.7	2911.5	2025	4186.09	
22	47	67.8	3186.6	2209	4596.84	
24	49	57.1	2797.9	2401	3260.41	
25	51	57.1	2912.1	2601	3260.41	
26	43	73.4	3156.2	1849	5387.56	
27	43	60	2580	1849	3600	
28	45	66.5	2992.5	2025	4422.25	
29	56	68.4	3830.4	3136	4678.56	
30	41	56.9	2332.9	1681	3237.61	
31	70	72.2	5054	4900	5212.84	
32	54	60.2	3250.8	2916	3624.04	

33	42	58.4	2452.8	1764	3410.56	
35	47	60.9	2862.3	2209	3708.81	
36	42	56	2352	1764	3136	
37	60	71.5	4290	3600	5112.25	
38	41	52.9	2168.9	1681	2798.41	
41	44	50.3	2213.2	1936	2530.09	
42	43	68.5	2945.5	1849	4692.25	
43	44	67.4	2965.6	1936	4542.76	
44	43	53.3	2291.9	1849	2840.89	
46	44	67.3	2961.2	1936	4529.29	
48	41	55.1	2259.1	1681	3036.01	
49	41	58.4	2394.4	1681	3410.56	
51	51	50.1	2555.1	2601	2510.01	
52	54	65.4	3531.6	2916	4277.16	
53	55	56.2	3091	3025	3158.44	
55	41	53.4	2189.4	1681	2851.56	
56	52	52.9	2750.8	2704	2798.41	
57	55	52.6	2893	3025	2766.76	
68	49	49.6	2430.4	2401	2460.16	
59	57	65.5	3733.5	3249	4290.25	

60	53	62.4	3307.2	2809	3893.76	
61	59	69	4071	3481	4761	
62	55	54.6	3003	3025	2981.16	
63	43	52.2	2244.6	1849	2724.84	
64	58	58.7	3404.6	3364	3445.69	
65	46	66	3036	2116	4356	
66	48	60.1	2884.8	2304	3612.01	
67	46	58	2668	2116	3364	
68	46	61.5	2829	2116	3782.25	
69	44	67.5	2970	1936	4556.25	
71	50	58.1	2905	2500	3375.61	
72	59	69.5	4100.5	3481	4830.25	
73	51	64.1	3269.1	2601	4108.81	
74	48	55.6	2668.8	2304	3091.36	
75	48	64.2	3081.6	2304	4121.64	
76	53	58	3074	2809	3364	
77	44	64	2816	1936	4096	
78	47	66.5	3125.5	2209	4422.25	
79	57	70.3	4007.1	3249	4942.09	
80	55	70.8	3894	3025	5012.64	

82	43	59.9	2575.7	1849	3588.01	
83	45	51.5	2317.5	2025	2652.25	
85	68	82.7	5623.6	4624	6839.29	
87	66	65.8	4342.8	4356	4329.64	
90	42	65.7	2759.4	1764	4316.49	
91	48	68.7	3297.6	2304	4719.69	
92	49	67.4	3302.6	2401	4542.76	
93	44	59	2596	1936	3481	
94	56	77.9	4362.4	3136	6068.41	
95	51	65.1	3320.1	2601	4238.01	
96	52	60.6	3151.2	2704	3672.36	
97	65	55.2	3588	4225	3047.04	
102	54	46.3	2500.2	2916	2143.69	
Σ	4066	5062.5	255833.5	207898	321020.53	

$\Delta X=29$
 $\Delta Y=37$

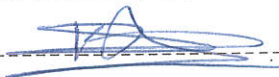
$r_{xy} =$
0.4082877

ADDIS ABABA
LIBRARY
P.O. BOX 11
ADDIS ABABA

DECLARATION

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

Name Kessa Amare

Signature 

Date of submissions March 27, 2007