THE PRACTICES AND CHALLENGES OF TEST CONSTRUCTION AND ADMINISTRATION IN SELECTED SECONDARY SCHOOLS OF KOLFE KERANIO SUB-CITY IN ADDIS ABABA

By

Ashenafi Tilahun Tesheberu

June 6/2014

Addis Ababa

THE PRACTICES AND CHALLENGES OF TEST CONSTRUCTION AND ADMINISTRATION IN SELECTED SECONDARY SCHOOLS OF KOLFE KERANIO SUB-CITY IN ADDIS ABABA

BY

ASHENAFI TILAHUN TESHEBERU

A THESIS SUBMITTED TO
COLLEGE OF EDUCATION AND BEHAVIORAL SCIENCES
DEPARTMENT OF CURRICULUM AND TEACHERS PROFESSIONAL DEVELOPMENT STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR DEGREE OF MASTER OF ARTS IN CURRICULUM AND INSTRUCTION

ADDIS ABABA UNIVERSITY

ADDIS ABABA, ETHIOPIA

JUNE 6/2014

Addis Ababa University

College of Education and Behavioral Sciences

Department of Curriculum and Teachers

Professional Development Studies

This is to certify that the thesis prepared by Ashenafi Tilahun Tesheberu, entitled: *The Practices and Challenges of Test Construction and Administration in Selected Secondary Schools of Kolfe Keranyo Sub city in Addis Ababa* and Submitted in partial fulfillment of the requirements of the degree of Degree of Masters of Arts (Curriculum and Instruction) complies with the regulation of the University and meets the accepted standards with respect to originality and quality.

Signed by the Examining Committee

_____	_____	_____
External Examiner	Date	Signature
_____	_____	_____
Internal Examiner	Date	Signature
_____	_____	_____
Advisor	Date	Signature
_____	_____	_____
Chairman, Dept, Graduate committee	Date	Signature

ACKNOWLEDGEMENTS

I am greatly indebted to my research advisor, Dr. Abdulaziz Hussein, for his unreserved help, suggesting the problem for guidance and encouragement throughout the course of this study.

I greatly appreciate receiving help form my dear wife W/ro Birtukan Amare and my children for their unforgettable contributions throughout the course and to complete this work, and special help of W/t Terowork Damene is unforgettable.

I wish to thank Ato Girmaye Alemu, Ato Daniel Tesfaye, and W/t Firehiwot Dagne for their material and technical support to complete this thesis.

My thanks are also due to my research participants: teachers, directors and students of Repi, Atlas and Beteseb secondary school, for their co-operation in giving reliable responses on time.

Table of Contents

Pages

List of Tables Pages

ABSTRACT

THE PRACTICES AND CHALLENGES OF TEST CONSTRUCTION AND ADMINISTRATION IN SELECTED SECONDARY SCHOOLS OF KOLFE KERANIO SUB-CITY IN ADDIS ABABA

Ashenafi Tilahun
Addis Ababa University

The purpose of the study was to assess the practices and challenges of test construction and administration in selected secondary schools of Kolfe Keranyo Sub city in Addis Ababa City Administration. Descriptive survey research design was employed. Primary data was collected from one government and two private secondary schools. The participants were 85 teachers and 80 first to third ranking students which were selected from the above mentioned schools using random and purposive sampling techniques respectively. To collect primary data, questionnaire with close ended and open ended questions and interview guide were used. The major findings of this study indicated that most of the teachers lack the awareness of test construction guidelines, and lack the skill to apply it practically and even had confusion about it. Teacher- made tests were not effectively prepared as observed in the review. It was also found that the controlling mechanism of proper construction of test is very weak. Finally, in relation to the assessment, the possible recommendations were made.

CHAPTER ONE: INTRODUCTION

This chapter deals background of the study, statement of the problem, objectives of the study, significant of the study, delimitation and limitations of the study and organization of the study.

1.1. Background of the Study

The historical development of educational measurement dates back to 1950s up to the present. For instance, the rules for conducting written exams, establishment of examination board, practical mental tests, identification of factors of intelligence, use of objective classroom tests, scholastic aptitude tests, the development of test scoring machine, taxonomy of educational objectives are some of the reforms up to 1960s (Smith D. 2005).

In the formal school system, assessments in general and formative tests in particular are intended to provide feedback that can be used both by teachers and by learners. Teachers can use the results to gauge learning, monitor performance, and guide day-to-day instruction. Students can use the results to assist them in identifying their own strengths and weaknesses and focusing their studying.

In school situation, all assessments should be designed to be of high quality to measure the intended learning outcomes, provide useful and accurate information and meet technical and psychometric standards. When assessments are used to make decisions like promotion or retention, graduation, admissions and the like, they must meet accepted standards to ensure that they are reliable, valid and fair to all students who take them (ETS, 2010).

In relation to this, (Sarita, 2005) noted that good tests avoid all undue pressure, promotes self confidence and thinking skills and therefore, stimulate academic progression. Test can be used as guide to pupils' needs, for placement of pupils, as part of learning experience, as a means of diagnosis, and as a means of pupil feedback.

In the Ethiopian context, the importance of assessment and evaluation is highly emphasized in the blue-print of the school improvement program of 2007. According to this guideline of the MoE (2007), assessment techniques and related activities should help the improvement of the lesson presentation and students' achievement results. In implementation of the techniques, considering age and individual capacity of students, teacher must use the various types of instruments or assessment procedures related to their lesson contents. Such instruments include assignments, projects, observations, interviews, tests, etc. As indicated by Borich (2011), a test must assess specific ability or comprehension of content developed during the teaching-learning of the subject. The primary focus in deciding on the type of a test is to consider the knowledge, skill, and/or behavior that the test developer would like to elicit and then to consider the best and most cost-effective way to elicit it.

Equally important in using classrooms tests are how will the assessment be administered to test takers? Where will it be administered? When will it be administered and how often? Who will administer it? There are numerous options for how the test may be delivered to examinees and how they respond to it. Choosing among these options requires consideration of practical constraints.

This study, therefore, was designed to assess the practices and challenges of using classroom tests in selected secondary schools of Kolfe Keranio Sub-city in Addis Ababa City Administration.

1.2. Statement of the Problem

Educators like Borich G. (2011) and Sarita K. (2001) emphasize the importance of having a clear articulation of what to assess using tests: that is, the knowledge, skill, and/or behavior the stakeholders would like to measure. This helps the test developer to determine how to measure the intended learning outcomes. This means, in order to develop a test that meets appropriate standards, one should have clear understanding of the intended leaning outcomes and the associated subject matter in a detailed and sufficiently precise manner to support the development of test items. As noted by Sarita K. (2001) test development is less challenging when the construct is the objectives and the specific subject-matters are concrete and discrete.

In relation to this, educators also give emphasis to the type of test one has to develop based on the type of knowledge and skills to be measured. For instance, if the objective is to measure teamwork skills, it may be necessary to observe the test takers actually performing their teamwork skills. However, it may be sufficient to ask questions to know how students collaborate with others and effectively work as a team. Whatever test may be set should be guided by and based on test construction principles. These principles, among other things, focus on the learning targets and the match between the type of test and the learning targets (Plessis, 2003).

Besides, the other concern, as mentioned earlier is, related to test administration which includes issues related to resources or manpower allocation, testing time, nature of test takers, testing place and other questions related to the how of testing and other constraints.

When seen against the above concerns on test construction and administration, the student researcher has his own informal

observations with regard to the quality of test construction and test administration in general and in the current secondary schools in particular. Some of these include inability to follow the guidelines in item construction and administration, unclear purpose as to why to use classroom tests, imbalance between the test and time allotted for tests, extreme extraneous factors like ambiguities, grammatical errors, unclear instructions, etc. On top of this, there are complaints from students and even some teachers about scheduling and grading. In relation to this, the student researcher could locate those studies conducted on the practices and challenges of test construction and administration in selected secondary schools of Kolfe Keranio Sub city of Addis Ababa. Hence, the informal observations and the complaints from the school level actors triggered the researcher to pick this issue that worth studying.

With this in mind, study was guided by the following research questions.

1. To what extent are teachers awareness of the principles of test construction and administration in selected secondary schools of Kolfe Keranio Sub city in Addis Ababa City Administration
2. To what extent do teachers apply the principles of test construction in setting classroom test items?
3. To what extent do teachers apply the principles of test administration in selected secondary schools of Kolfe Keranio Sub city?
4. What are the major problems encountered by the teachers in constructing test in selected secondary schools of Kolfe Keranio Sub city?
5. What are the major challenges encountered in test administration in selected secondary schools of Kolfe Keranio Sub city?

1.3. Objectives of the Study

1.3.1. General Objectives

The major objective of the study is to assess the extent to which teachers apply principles of test construction and administration in selected secondary schools of Kolfe Keranio Sub city in Addis Ababa City Administration.

1.3.2 Specific Objectives of the Study

- To assess whether or not teachers in selected secondary schools of Kolfe Keranio Sub city are familiar with the principles of test construction and administration.
- To assess whether or not teachers apply the principles of test construction and administration in selected secondary schools of Kolfe Keranio Sub city of Addis Ababa City Administration
- To investigate the challenges encountered in test construction and administration in selected secondary schools of Kolfe Keranio Sub city in Addis Ababa City Administration.

1.4. Significance of the Study

In the modern world, where updating and upgrading teachers is important for the success of teaching leaning activities, studies conducted on assessment in general and on classroom test construction and administration in particular may have their own enlightening nature. Hence, this study may have the following significances to school level actors and other education officials.

- The study helps teachers, departments, directors of the secondary schools to create awareness on the existing strengths and weaknesses in implementing the principles of test construction and test administration.

- The study may help wereda education officers and supervisors to plan for proper implementation of the principles of test construction and administration.
- The work could also be a source of information for those interested to do further study on related topics.

1.5. Delimitation of the study

As mentioned above, the study deals with the practices and challenges of Test Construction and Test Administration in selected secondary schools of Kolfe Keranio Sub city in Addis Ababa City Administration. The study focuses on the investigation of the teachers' awareness of the principles, the actual practice and the challenges in test constructions and administration. In so doing the researchers delimited his study only to classroom tests. Because of resource constraints, economic conditions the study also focused on selected general (9 – 10) secondary schools and took only one government and two private secondary schools (Grade 9-10). For the number of teachers are very few in the department.

1.6. Limitation of the study

In conducting this study, the data gathered was constrained due to the reluctance of some secondary school teachers to fill and complete the questionnaire. The results would have been more reliable if all the research participants had filled the questionnaire. Nevertheless, the researcher tried to get the responses of the majority of respondents by making repeated attempts to get the questionnaires back.

1.7. Definitions of Operational Terms

1.7.1. Practices of test construction and administration: The better activities of experience carried out in relation to construction or preparation of tests and test and administration and the weaknesses observed.

1.7.2. Challenges of test construction and Test Administration: the problems encountered in the preparation and administration of tests in general.

1.7.3. Awareness: Recognition of Something, knowing something

1.7.4. Test or examination: Is an assessment intended to measure a test taker's knowledge, skill aptitude, physical fitness, or classification in many other topics. A test may be administered orally, on paper, on a computer, or in a confined area that requires a test taker to physically perform a test of skills, which vary in table, rigor and requirements.

1.7.5. Test Administration: refers to how will the assessment be administered to test takers, where will it be administered, when will it be administered and how often, who will administer it and the like.

1.8. Organization of the study

The study contains five chapters. The first chapter deals with the problem and its approach, the second chapter deals with the review of related literature, the third chapter is about the research design and methodology, the forth chapter deals with the presentation, analysis and interpretation of data. Finally, summary, conclusion and recommendations forwarded on the basis of the analysis are in the fifth chapter. Finally, references and necessary annexes are attached at the end of the thesis.

CHAPTER TWO: REVIEW OF RELATED LITERATURE

This chapter attempts to review available related literatures and research findings on test construction principles and its challenges in measurement and evaluation as a whole and test administration practices and the challenges encountered.

2.1. Measurement

According to Brad Field, “measurement is a process of quantification, a process of assigning symbols bo the dimensions of phenomenon in order to characterize the status of the phenomenon as precisely as possible”, as quoted by Lal, J.P. (2007).

The essentials of the measurement of learning was explained by the same author, above, in the following way: identifying and defining the quality, attribute or the variable that is to be measured; determining the set of operations by which the attribute or variable may be made, manifest and perceivable; and establishing a set of procedure or definitions for translating observations into quantitative statements and degree, extent or amount.

As Kumar S. (2005) pointed out, in his “Improving Assessment, Evaluation and Remedial”, formative and summative assessments play a tremendous role in school assessment activities. Formative assessment is concerned with a view to help the learner and the teacher to overcome the existing problems about the learner and summative assessment aims at certifying and grading the attainment of the learner at the end of a given course. Tests for formative assessment are given at regular and frequent intervals during the course, but the tests for summative assessment are offered at the end of a course; at the end of a term, or a semester, or a year.

2.2. Characteristics of Good Assessment in Teaching and learning

J.P.Lal also classified the good assessment techniques related to self evaluation and reflection of the teacher. The components of good assessment, hence, include: command, planning and organization of the subject matter or content and activities; class control and discipline; realization of psychology of the learner; self-evaluation by both the teacher and the learner.

“Evaluation of pupils progress is a major aspect of the teacher’s job. A good picture of where the pupil is and how he is progressing is fundamental to effective teaching by the teacher and to effective learning by the public, (Thorndik, Hagen)”, as quoted by Lal, J.P. (2005).

As to the suggestion by J.P.Lal, the above writers, Thorndile and Hagen, have briefed the purposes of evaluation as: motivating students, diagnosing weakness, defining teaching objectives, differentiation of pupils and for vaions purposes and lastly for certification of pupils.

Evaluation also helps the teacher in a number of ways: gives adequate knowledge about the learner, for setting or formulating enroller objective, to organize appropriate learning activities, to identify whether the objectives are realized, and finally to improve classroom procedures and methods instruction or teaching Lal, J.P. (2005).

Evaluation also helps the students, too; to aware of the objectives, to increase motivation, to encourage study habits, and to increase ability and skills (Ibid, 2005)

2.3. The Teaching Objectives

Our teaching objectives are the changes we wish to produce in the child. The changes that must take place through education are represented in: the knowledge the children acquire, the skills and abilities children attain, the interest children develop and the attitudes children manifest Kumari S. (2001)

These objectives must involve points of information, the skills and attitudes to be developed and interests that could be created though the particular topic or subject taken up for work in the classroom. As pointed out additionally by the same author, a statement of class room objectives serves: as a basis for the schools of classroom procedures that should provide for suitable experiences to the children, as guide in seeking evidence to determine the extent to which the classroom work has accomplished what it set out to do (Ibid, p.59)

2.4. The Need for Educational Objectives

The kind of questions to be constructed depends on the educational objectives you want to include in the test. Since the 1950's researchers have developed several taxonomies (or hierarchical categorization) of cognitive, affective and psychomotor objectives to be addressed with in testing situations. Kumari S. (2005)

2.5. Adaptation of Bloom's Taxonomy of Educational Objectives for Test Development

The cognitive domain: This domain lists six categories which vary in difficulty with respect to cognitive abilities or level of understanding: Sarita K. (2005)

1. Knowledge: recall of specific facts.

   Knowledge questions can be identified by key verbs or action verbs such as define, identify, list and name.

2. Comprehension: Understanding the purpose or meaning of something.

   Comprehension questions can be identified with key verbs such as covert, explain and summarize.

3. Application: using information and ideas in novel situations.

   Application questions can be identified with key verbs such as compute, determine and solve.

4. Analysis: breaking down large pieces of information in order to examine the structure and interrelationships among its component parts.

   Analysis questions can be identified by key verbs such as analyze, differentiate, and relate.

5. Synthesis: Combining various elements or parts in to a structural whole.

   Syntheses questions can be identified by key verbs such as design, devise, formulate and plan.

6. Evaluation: Making judgment based upon reasoning. Evaluation of questions can be identified by key verbs such as compare, critique, evaluate, and judge.

2.6. How to plan a classroom Test

Before planning and setting a class room test, the following decisions must taken into account: How the test is to be used, when to test, what emphasis to give to various aspects of achievement, whether the test should emphasize complex achievement. To cut line the context to be covered by the test, what level and distribution of difficulties are appropriate for the questions included in the test, what means and format to use in presenting the test to students and finally to test important our comes of instruction Kumari S. (2005)

To plan a test, we prepare a two- way table, which is called *a test blue-print.* In the table, the names of the major categories of a taxonomy head

the table columns, while the row heading indicates the major topics of the subject matter to be tested Kumari S. (2005) explains.

In the body of the table, the “cells”, formed by a combination of a particular taxonomy category and a particular subject- matter topic, contain specific instructional objectives. Thus, the blue print serves as a double- entry classifying a scheme for specific objectives Kumari S. (2005)

After objectives are classified, the number of test items that will be used to test each objective is record in the table. Thus, the test blue- print serves as a plan which assumes that all important objectives are included and that they receive the proper emphasis on the test (ibid, 2005).

2.7. The blue- print or Table of Specification

According to Gary B. (2011), a blue print is a table that matches that best items to be written with the content areas and levels of behavioral complexity taught. He continued describing that the test blue print ensures that the test will sample learning across the range of content areas covered by the teacher and the cognitive and affective skills and processes we consider important.

He also suggests the steps in constructing a test blue print. The logical steps include: classifying each instrumental objective for the contents, recording the number of items to be constructed for each objective, totaling the items of each instructional objective totaling the number of items falling in to each behavior and completing the column and row percentages.

These are: Identifying, naming, describing, constructing, ordering and demonstrating.

2.8. Test preparation

Time and effort is needed of the teacher as well as the lest developer. For test takers too, time is very much need.

When a test developer constructs a test, the amount of time and effort need is dependent up on:

> *"the significance of the test, the proficiency of the test taker, the format of the lest, class size, deadline of the test and experience of the test developer".... PTI. 2006*

The test construction has been greatly aided by several ways

In some countries (USA for example), book publishers often provide teaching packages that include test banks, up to 4,000 sample test item questions that have been peer- reviewed and time tested, to university institution who adopt their published books for their courses (so that the instructor who chooses to use this test bank would only have to select a fixed number of test questions for this test bank to construct a test (adopted from Wikipedia cited above)

Some Principles of Test Constructions

As pointed out by Kumari S. (2005) P. 98, principles for writing True-False include making item either definitely true or definitely false, avoiding verbal clues (specific determiner) that give away the answer, asking important ideas, knowledge or understanding (rather than trivia, general knowledge, or common sense), keeping the word length of true statements about the same as that of false statements, avoiding copying sentences directly from textbook and other written materials and voiding presenting items in a repetitive or easily learned pattern.

Principles for Constructing Matching Types, Kumari S. (2005.) p. 98-99,

Within a single matching exercise make the premises (in column A) and responses homogeneous (commonly those in column B), Write directions that explain completely the intended basis for matching, Cheek to see that all the responses function as plausible options to each premise, keep the least of premises and responses within a single matching exercise relatively short. Avoid creating “perfect matching” in which each response matches only one premise, Place the larger phrases (sentences) in the premise list and the shorter premises, words, or symbols in the response list, If at all possible, arranges the responses in a logical, meaningful order, use numbers to identify the premises and letters to identify the responses, Avoid using incomplete sentences for premises, keep all the premises and responses belonging to a single matching exercise on the same page.

Principles of constructing multiple choice items

With emphasis on how to formulate the stem of the Item Avoid “clueing” and “linking” items (i.e, having the correct answer to are item be clued or linked to the correctness of the answer to a previous item, If possible, write as a direction question, if an incomplete sentence is used be sure it imposes a direct question, the alternatives come at the end (rather than in the middle) of the sentence, control the wording so that vocabulary and sentence structure are at a relatively low and of an option, in items testing definitions, place the word or stem in the stem and use definitions or descriptions as alternatives, avoid extraneous, superfluous and non functioning words and phrases that are mere “window dressing”, avoid (or use sparingly) negatively worded items, avoid phrasing the stem so that the personal opinion of the examinee is non-

technical level, avoid text book wording and “text bookish” or stereotyped phraseology.

Hints for Improving to quality of the Alternatives in multiple- choice item Kumari S. (2005) p. 100

1. In general strive to create three to five functional alternatives.
2. All alternatives’ should be homogeneous and appropriate.
3. Put repeated words and phrases the stem.
4. Use consistent and correct punctuation in relation to the question.
5. Arrange alternatives in a list format rather than in alternatives.
6. Arrange alternatives in a logical or meaningful order.
7. All distracters should be grammatically correct with the stem.

To Avoid:

1. Avoid overlapping alternatives
2. Avoid making the alternatives a private to the stem collection of true- false items.
3. Avoid using “not given”, “none of the above”, etc as an alternative in best- answer type of items. (use only with correct answer variety)
4. Avoid using “all of the above”, limit to the stem. Its use to the correct- answer variety.
5. Avoid using verbal clues in the tandem?
6. Avoid using technical terms, unknown words, or names, and “silly” terms or names as distracters.
7. Avoid making it harder to eliminate a respect to the same distracter than to choose the keyed alternative.

Principles for Writing Short answer or Completion Items

The Following suggestions are formulated by Sarita Kumari, P. 97-98m 2005.

1. Word each tem in specific terms with clear meanings so that the intended answer is the only are possible, and so that the answer is a single word, brief phrase or number.
2. Word each item so that the blank or answer space is toward the end of the sentence.
3. Avoid copying statement verbatim from texts or classroom materials.
4. Omit important rather than trivial words.
5. Avoid “butchered” or “mutilated” sentences, use only one or two blanks in a completion sentence.
6. Keep the blanks of equal length and arrange the items so the answers are placed in a column at the right or left of the sentences.
7. State the precision, numerical units, or degree of specificity expected of the answer.
8. Word the items to avoid irrelevant clues or specific determiners.

Principles for constructing Essay- Type Tests. Kumari S. (2005) p. 94-95

1. Define behavior the examinee is expected to exhibit or describe the process to be exhibited before beginning to write the essay question.
2. Ask questions that require the examinee to determine the ability to use essential knowledge and to do so in situations that are new or novel for the examinee, rather than simplify recalling information from a textbook or a classroom.
3. Ask questions that are relatively specific of focused, and which require relatively brief responses.
4. If a test includes several essay questions, be sure that they cover the appropriate range of topics and complexity of behavior called for in the test blue-print, but the sure that the complexity of the

questions are within the educational maturity level of the examinees.

5. Require all the examinees to answer the same questions: don't give optional questions.
6. Word questions so that all examinees interpret the task the way you intend.
7. Word questions so that all examinees know the limits of the tasks, their purposes, and can answer them in the time allotted.
8. Word questions so that experts can agree on the correctness of an examinee's response.
9. Word questions calling for examinee opinion on controversial matters so that they ask the examinee to give evidence to support the opinion and evaluate the examinee's response in terms of the evidence presented rather than the opinion to position taken.
10. Word questions so that the examinee can judge the approximate length of the answer desired and know the point value of weight each will be given.

Effective System of Item Construction

Prepare new exams each time you teach a course, make up test items throughout the term, Ask student to submit test questions, Call items from colleagues exams, Consider your tests cumulative, Prepare clear instructions, Include a few words of advice and encouragement on the exam, Put some easy items first, Challenge your best students, give some though questions to the layout of the test (spacing and attractive looks.)

2.9. Test Administration

The major purposes test administration according to an article in Wikipedia for preparing exam program, to reduce measurement errors, to increase fair, valid and reliable assessment as well as to in test security and consistency and lastly to maintain the integrity of score for all examinees. (Wikipedia)

The Importance of Test Administration

One of the major issues as the above mentioned paper pointed out is *consistency*, which means increasing the fairness of the tests and exam programs, making the scores comparable, creating comfortable sites for the examinees, good lighting, ventilation and handicap accessibility, avoidance of interruption and excessive noises.

The second one is test security, which implies prevention of cheating, practicing test items and contents from being exposed to future test takers, registration procedures (eligibility), examinee's identification, restricting invalid materials, and protection of integrity of test items

Source: www.proflesling.com/pdf/test, Professional Testing Inc. PTI 2006 Wikipedia

The challenges of Test Administration:

The challenges and current problems in test administration in Ethiopia as indicated in an article produced in the Wikipedia are mainly the following:

Large class size, poor resources and facilities, insufficient expansion of staffing, forged documentation, mark and grade inflation, cheating on tests, lack of academic honesty are some of them.

2.10. Allaying students' anxieties about tests:

Make first exams relatively easy, give more than one examination, avoid "pop" gives (un announced), give students advice on how to study, encourage students to study in group, encourage extra office hours before a test, schedule review sessions before major exams. Kumari S. (2005)

General Strategies of Test Development

Spend adequate amounts of time developing your tests, match your tests to the content you are teaching, try to make your tests valid, reliable, and balanced, content validity (the content of the test representing skill etc., reliability (accuracy and consistency), use a variety of testing methods, write questions that test skill other than recall. (Ibid, p. 148-149, 2005

Item Requirements/ Criteria

To be an effective instrument of learning, items on questions should meet the following requirements.

- The task that an item specifies should, in the process of the learner- response to it, demand and reflect only those specific aspects of skills or bits of learning that are being tested
- It should specify precisely:
  - What the learner is to do
  - The conditions under which it is to be done and,
  - To what level/ standard it is to be accomplished.
  - The medium (linguistic, graphic, semantic, used to present the task specification should be such that there may not be any gap in its communication to the prospective i.e, the learner should

be able to follow the medium without any misunderstanding (of dubious understanding) (Ibid, p. 42)

The Requirements or Criteria of Evaluating Tests

1. The ultimate goal of any evaluation should be to collect relevant, valid, reliable and economical information for decisions to be appropriately made. Relevance, validity, reliability and the economical aspect are currently the most expected requirements or criteria for any evaluation test.
2. The relevance of data collected implies that the subject under evaluation precisely and specially corresponds to the objectives targeted by the evaluation. For example, the relevance of examination tests at the end of a given curriculum makes it necessary to differentiate between examinations meant to evaluate learner's qualification for promotion to an upper class or to move in to the job market and active life.
3. The validity data collected implies that the evaluation has actually been focused on the subject initially targeted for evaluation. For instance, for the sake of validity, learner's written and oral skills cannot be evaluated with the same lasts.
4. The reliability of data collected implies that they are not determined by the free will and choice of the individual who collected them. For example, the double grading of examination papers is meant to consolidate and further ascertain reliability. Kum.S (2005) P. 100-101.

2.11. Item Banking in Testing and Assessment

Availability and quick access to good quality items is usually expected by both teacher and test developers. A large collection of good items will help teachers to concentrate more on their teaching without having to spend much time on item construction (Torsten Husen, 2003)

There is no agreement on how an item bank is defined. A restrictive definition by Choppin, cited by Torgten reads, “collection of test items organized and catalogued to take in to account the content of each test item and also its measurement characteristics (difficulty, reliability, validity etc)”

Based on the extent of computer involvement in its operation, as Torsten explained, item banking could be classified as fully manual item banking, and manual item banking.

The use of tem banking is associated with the need for making test construction easier, faster and more efficient explained the author. The other advantages of calibrated item banking include: decentralization policy, reduction of cost and time spent, item large, quality of testing program, designing private assessment instruments, designing best possible test, facilitating a criterion referenced interpretation (Ibid, 2003).

Choppin (1981), as pointed out by Torsten, also identified specific advantages of item banking for the development and operation of a system of national examinations. These advantages are categorized as: economy, flexibility, consistency and security in a brief generalization.

CHAPTER THREE: RESEARCH DESIGN AND METHODOLOGY

3.1. Design

The purpose of this study was to assess the practices and challenges of test construction and test administration in selected government and private secondary schools of Kolfe Keranio sub city in Addis Ababa City Administration.

To achieve this purpose of the study, analytics survey research design was employed, on the consideration that it would help the researcher to explore the existing condition as they exist now. To examine the practices and by the problems the researcher used also qualitative research approach to obtain the necessary data from the pertinent data sources.

3.2. Participants

The sources of data for the study were secondary school teachers of Repi, Atlas and Beteseb secondary schools. Repi is a government school while Atlas and Beteseb are private schools. Two private schools were included because the number of teachers in these schools was relatively less, as compared to the government schools. Hence, in order to secure data from relatively less large number of teachers, the researcher included two private schools and one government school in this study.

In addition to this, grade 9 and 10 ranking (1^{st} - 3^{rd}) students of the above schools were the research participants with the assumption that they forward some ideas on the practices, challenges and students complaints on classroom tests. The other source of data were teacher-made tests or exam question papers of classroom teasts of different subjects randomly collected for document review.

Population, Sample and Sampling Techniques

The population of this study was teachers and students in the above mentioned selected general secondary schools of Kolfe Karniyo Sub-city of Addis Ababa City Administration. Using stratified random sampling, from three woredas 112 teachers were selected to fill the questionnaire while ranking students were purposely included in the study from the same schools.

The questionnaire was distributed to 76 teachers of Repi secondary school, 16 teachers of Atlas secondary school and 20 teachers of Beteseb secondary school. Out of 112 teachers, 85 of them had properly, filled and returned the questionnaire; while the remaining 27 copies of the questionnaire were not either returned from teachers or were incomplete.

Ranking students (1st to 3rd) were also the research participants purposely selected from the above mentioned schools. Therefore, 80 first to third ranking students from the three secondary schools filled the questionnaire prepared in Amharic.

3.3. Instruments

Based on the research basic questions and the review of related literature, the instruments or the tools of data collection for the study were developed and later checked and approved by the advisor. The important tools employed were the open-ended and closed ended questionnaire using the rating scales or Likert scale. As noted by Sarita Kumar (2005) this tool is less expensive and used to reach large number respondents. With this assumption, the researcher used questionnaires for both teachers and students. Besides, interview was conducted with school directors and document review was also employed to gather the relevant data.

3.3.1. Questionnaire

Questionnaire was used in order to gather data from teachers of government and private secondary schools teaching in grade 9 and 10. Specifically, teachers' questionnaire which has 54 items was used to assess whether the teachers aware of the guidelines and principles of test constructions and administration using five rating scales: strongly agree (5), agree (4) undecided (3), disagree (2) and strongly disagree (1). Additional questions were added for teachers to allow them further clarify or express their views or opinion in written form on important issues not mentioned or asked in the closed ended items.

On the other hand, students' questionnaire was prepared for ranking (1st to 3rd) students of the above mentioned schools. The number of items for their questionnaire were 21 (Amharic Version) to indicate their responses by five rating scale like that of teachers. The purpose of the questioners is to collect data related to practices and challenges of test construction and administration.

3.3.2. Interview

Interview schedule which bears 11 questions were prepared for directors and vice directors of the selected 3 secondary schools. The purpose of the interview is to gather data on the challenges of test administration.

3.4. Procedures

Pilot test was conducted using experienced teachers and colleagues other non sample school. The feedback was organized, analyzed and later employed to improve the tools constructed as much as possible.

After preparing and duplicating the questionnaire and the interview tools, explanation was given on the purpose of the study and on how to fill information correctly in time before the respondents started filling out

the questionnaires. The researcher himself distributed and collected back the questionnaire from the majority of the respondents who were cooperative and willing to fill the questionnaire on time. In all cases, proper format and ethical considerations were taken into account. The purposes of the tools, and how to give relevant information were clearly described in the general directions of the tools.

3.5. Method of Data Analysis

As pointed earlier in the research design and methodology, both quantitative and qualitative data were employed to address issues raised in the basic questions. The data gathered by means of questionnaires from teachers and students was tabulated properly and analyzed and interpreted using analytic statistics, like the mean and percentage. The data gathered through interview and document review were used in the elaboration and interpretation of the qualitative data.

CHAPTER FOUR: PRESENTATION AND ANALYSIS OF DATA

In this section, the main emphasis is the presentation and analysis of the data gathered from the selected teachers and ranking students of government and private secondary schools (grade 9 and 10), using the tools like questionnaire, interview and document review to support the items in the questionnaire.

As mentioned earlier, out of 112 teachers, 85 (75.9%) teachers filled the questionnaire properly and returned on time. On the remaining, 20 teachers were not volunteers to fill and return the questionnaire while 7 copies of the questionnaire were rejected because of not incompleteness. Similarly, out of 100 copies of questionnaire distributed to students, 80 (80%) of them filled the questionnaire properly; 6 of them didn't return the questionnaire and 10 of them were not interested to fill the questionnaire and 8 copies were rejected due to similar problem. The data gathered from interviews and document analysis was used as supplementary to the analysis of the quantitative data.

4.1. Analysis of Teacher's Responses

4.1.1. Background of Teacher Respondents

The data gathered for this study was obtained from teachers of one government and two private secondary schools, teaching grade 9 and 10 students. The following table shows the characteristics of these respondents in terms of sex, age, service year and educational qualification.

Table 1. Teacher respondents by sex

Sex	Frequency	percent
Male	64	75.3
Female	21	24.7
Total	85	100.0

As seen from this demographic data, the number of male teachers involved in the study was 64 which is 75.3% and that of female teachers in all the three secondary schools was 21 which is 24.7%. This data shows clearly that the participation of female teachers was very low compared to their male counterparts.

Table 2. Teacher respondents by Age

Age Interval in years	Frequency	Percent
21-30 years	67	78.8
31-40 years	7	8.2
41-50	4	4.7
Above 50	5	5.9
Total	83	97.6
Missing	2	2.4
Total	85	100.0

As displayed in the above table, concerning the age of the respondents, 67 (78.8%) of them were in the age of range of 21-30, 7 (8.2%) in the age category of 31-40, 4 (4.7%) were in the age interval of 41-50 and 5 (5.9%) were above the age 50. This implies that the majority of the teacher respondents were in their active age.

Table 3. Teacher respondents' experience of teaching in years

Age Interval in years	Frequency	Percent
Less than 5	54	63.5
5-10	18	21.2
11-15	1	1.2
16-20	1	1.2
21-25	3	3.5
26-30	2	2.4
Above 30	6	7.1
Total	85	100.0

As indicated in table 3 above, 54 (63.5%) of teachers had taught for less than 5 years. Furthermore, 18 (21.2%) of them had taught for 5-10 years. Only one teacher taught for 11-15 years still another one teacher taught for 16-20 years. Those who taught for 21 years and above altogether were 11 (12.9%) of the total teacher respondents. This clearly indicates that relatively majority of the teacher respondents teaching in the selected school were young and have lesser experience (less than 5 years) in the teaching profession. This might have its own impact on the know-how and skills of test construction and administration.

Qualification wise except one teacher who has diploma, all the teachers respondents were first degree holders, either in natural science stream or social science stream.

Table 4. Teacher respondents in terms of training they undergone

No.	Courses/specific topics	No.	%*
1.	General pedagogy	71	83.5
2.	Various methodology course including measurement and evaluation	58	68.2
3.	Training only on active learning	61	71.8
4.	Training only on lesson planning	52	61.2
5.	Continuous assessment	55	64.7
6.	No training at all	14	16.5

**percentage for each row is calculated out of the total as each respondent can have different training opportunities.*

As can be seen from the above table, 71 (83.5) of the teacher respondents indicated that they had taken general pedagogical courses. Fifty eight (68.2%) of them had also taken various methodology course including measurement and evaluation and 61 (71.8%) of the teachers were trained on active learning. Still 52 (61.2%) of the teacher respondents pointed that they got training on lesson planning and 55 (64.7%) of the respondents got trained on continuous assessment. Only 14(16.5%) of the teacher respondents reported that they didn't get any chance to be trained. This means majority of the teachers respondents had some kind of training which can either directly or indirectly impacts their know-how and skills of test construction and administration.

4.1.2. Background of Student Respondents

Table 5. Student respondents by sex

Responses	Frequency	Percent
Male	49	61.3
Female	27	33.8
Total	76	95.0
Missing	1	5.0
Total	80	100.0

As indicated in the table above, 49 (61.3%) of the student respondents were males while females were 27 (33.8%) of the total students participated in the study. This means, majority of the student respondents were males which was the reflection of student population in the schools under discussion.

Table 6. Student respondents by school

School	Frequency	Percent
Repi secondary	59	3.8
Atlas secondary	12	15.0
Beteseb secondary	9	11.3
Total	80	100.00

As mentioned earlier, Repi is a government secondary school while Atlas and Beteseb are private secondary schools. The number of student selected was proportional to the total number of student population in each school. Hence, 59 (73.8%) of the ranking students were from Repi secondary school while the remaining 21 (26.3%) were from Atlas and Beteseb secondary schools.

4.1.3 Analysis of Data Related to Construction of Test Items

In order to assess the perceptions and practices of the respondents, concerning test construction, items related to different kinds of test items were administered for rating by respondents using the five rating scale

(strongly agree to strongly disagree) as described previously. The responses are presented and analyzed using percentage and mean of the scores in the tables below the mean indicates the category where the response of the items lie.

Table 7: Construction of True- False Item

No	Item Descriptions	Rating Scales /Responses/										Mean ($\bar{X}$)
		SD		DA		U		A		SA		
		N	%	N	%	N	%	N	%	N	%	
1.1.	Teachers don't usually specified allotted time	28	32.9	25	29.4	5	5.9	14	16.5	13	15.3	2.52
1.2	Teachers don't usually give clear instruction	21	24.7	37	43.5	3	3.5	14	16.5	10	11.8	2.47
1.3	Teachers don't balance the number of true and false items	9	10.6	24	28.2	27	31.8	20	23.5	6	5.9	2.91
1.4	The statement questions are equal in length	12	14.1	33	38.8	20	23.5	15	17.6	5	5.9	2.62
1.5	Teachers usually use negative statements	2	2.4	11	12.9	11	12.9	51	60.0	9	10.6	3.64
1.6	Teachers commonly use double negative statements	22	25.9	22	25.9	14	16.5	22	25.9	5	5.9	2.6

Key: SD= Strongly disagree DA= Disagree U= Undecided A= Agree SA= Strongly Agree

Under section I of True-false item construction, as shown above, respondents were asked to rate Item 1.1 (table 7) which reads "Time allotted is not specified", for classroom tests using the five rating scales. Accordingly 32.9% and 29.4% of the respondents revealed their strong disagreement and disagreement respectively. Only 16.5% and 15.3% of the teachers indicated their strong agreement and agreement respectively, while the remaining respondents (5.9%) couldn't decide.

The weighted mean of 2.52%, however, shows that respondents couldn't decide or couldn't show their agreement or disagreement. This implies that teachers don't have clear views when appropriate time for the classroom tests is allocated. Item 1.2 (table 7) stated "clear instruction is not usually given", was rated as "strongly disagree" and "disagree" by 2.47% and 43.5% of respondents respectively. Those who showed their agreement (strongly agree and agree), with the statement accounted for

28.3% of the respondents, while only 3.5% rated “undecided”. The weighted mean for the item is 2.47, showing that the weighted mean falls in the category of “disagree”, which implies that teachers prepare clear instructions in the construction of classroom tests.

In responding to item 1.3, in the same table, 38.8% of the teacher respondents in sum rated “strongly disagree” and “disagree”, while 29.4% them rated “agree” and “strongly agree” respectively where as 31.8% of the respondents responded “Undecided”. The mean score for this item is 2.91 which indicates that it falls in the category of “disagree”, reflecting that teachers don’t consider the number of true and false statement in a classroom test construction though in principle they are required to have balanced number of true-false statements.

In rating item 1.4 (table 7) which is read “The statement questions are equal in length”, 14.1% and 38.8% of the respondents indicated “strongly disagree” and “disagree” in which 52.9% of them oppose the case. On the other hand 23.5% of the teachers couldn’t decide on the item. But, 23.5% of the respondents had agreed revealing “agree” (17.6%) and “strongly agree (5.9%) with the average score of 2.62 which implies that most of the teachers (52.9%) had no idea about the length of the statement of the item. On the contrary, the guide lines of test construction; as proposed by many educators, the statement questions of the true-false items must be almost equal in length to avoid clues.

In item 1.5, (table 7) teachers were requested whether negative statement in the question are also used. Concerning this, the respondents rated 2.4% and 12.9% for “strongly disagree” and “disagree” respectively, in which 15.3% didn’t agree. 12.9% rated “undecided”. The majority of the teachers, 71.6%, had agreed, and the mean score, 3.64, shows that negative statements are also employed in the main questions of the item.

As many related review literatures suggest, the use of negatives in the true-false item construction must be avoided.

Item 1.6 (table 7), which states the use of double negatives in the true-false item. The item was rated “strongly disagree” by 22 (25.9%), “disagree,” by 22 (25.9%), “disagree,” by 22 (25.9%), and 14 (16.5%) and “strongly agree”, ‘agree” by 22 (25.9% and 5(5.9%) respectively. As shown here, more than 50% of the teachers disagreed about the use of double negatives. The weighted mean calculated, 2.6, shows that teachers believed in using items without double negatives. Supporting this idea, avoiding double negatives is one of the guide lines of test construction.

Table 7: Continued

No	Item Descriptions	Rating Scales /Responses/										Mean ($\bar{X}$)
		SD		DA		U		A		SA		
		N	%	N	%	N	%	N	%	N	%	
1.7	Teachers sometimes give clues or hints in the statements	9	10.6	18	21.2	18	21.2	31	36.5	9	10.6	3.15
1.8	General or common sense questions are asked	4	4.7	10	11.8	11	12.9	44	51.8	15	17.6	2.33
1.9	Teachers mostly ask the learned concepts	1	1.2	4	4.7	9	10.6	47	55.3	24	28.2	4.04
1.10	Questions are sometimes ambiguous (not clear)	22	25.9	28	32.9	4	4.7	25	29.4	6	7.1	2.58
1.11	Teachers take questions directly from the textbooks	9	10.6	20	23.5	18	21.2	28	32.9	10	11.8	3.15
1.12	Absolutes like “Always”, “never”, are often used in the questions.	9	10.6	22	25.9	17	20.0	31	36.5	6	7.1	3.03
1.13	The question statements are based on the domains of educational objectives.	6	7.1	11	12.9	13	15.3	23	27.1	32	37.6	3.75
1.14	Teachers consider the level of difficulty of the questions against the level of students	8	9.4	20	23.5	13	15.3	27	31.8	17	20.0	3.29
1.15	Long and complete statements are usually used.	14	16.5	33	38.8	16	18.8	18	21.2	4	4.7	2.58

Key: SD= Strongly disagree DA= Disagree U= Undecided A= Agree SA= Strongly Agree

In responding to Item 1.7 (table 7), whether clues or hints appear in the question statement, 9 (10.6%) of the teachers rated" strongly disagree and 18 (21.2%) of them rated "disagree", while 18 (21.2%) of the respondents rated. "Undecided". The other respondents 31 (36.5% and 9 (10.5%) pointed out "agree" and "strongly agree" respectively. In this item, the mean score, 3.15 showed that most of teachers couldn't decide whether clues in the test items appear or not. The familiar guide lines for test construction suggests that clues in any item must be avoided.

To respond to item 1.8 (table 7), teachers were asked the statement which reads, "General or commonsense questions are asked". The majority of the respondents, 69.4%, of them agreed, but only 15.5% of the respondents disagreed, while 12.9% of them rated "undecided". When we look in to the mean score which is 2.33, it lies in the category "disagree", meaning the respondents didn't accept the above statement the true- false item construction.

Under item 1.9 (table 7), the statement reads as "The learned concepts are mostly used". For this statement, 5.9% the teachers didn't agree while 83.5% of them strongly agreed. 10.6% of the respondents didn't decide to agree or not. Almost all respondents responded positively, 71 (83.5%), clarifying that the common concepts are mostly asked in the classroom as it is also, demonstrated by the mean average, 4.04, contrary to the premise, simple and the learned concepts only shouldn't be set in the test construction in general.

Item 1.10 (table 7) refers to the idea that questions are seen to be ambiguous. This statement was rated by the participants. Among the participants, 25.9% and 32.9% rated, "strongly disagree" and "disagree" respectively, while 29.4% and 7.1% rated the statement as "agree" and "strongly agree". The remaining teachers, 4(4.7%), pointed out as "undecided". The data interpreted by the mean score, 2.56, explains that

most of the teacher participants as respondents were not able to identify ambiguous questions in the construction of test items in the classroom. As indicated in the guide lines of test constructions, pointed out in the review of the related literatures, the items prepared must not be ambiguous.

Considering item 1.11 (in table 7), which is read as “questions are directly taken from the text”, was asked to assess whether teachers are reasonable to identify if questions are set, directly, from texts. As seen from the table, this item was rated “strongly disagree” and “disagree” by 10.6% and 23.5% respectively by the teachers, while 21.2%, 32.9% and 11.8% of the respondents rated, “undecided”, “agree” and “strongly agree” respectively. To generalize this data, we obtained the score of the mean as 3.15, which falls in the category of the rating scale, “undecided”, which had an interpretation of being unable to evaluate the question papers how they are prepared. As is already known, questions must not be taken directly from texts for it makes or misguides the learners to be so bookish: suggested by many scholars.

Item 1.12 (table 7), “Always”, “never”, are often used in the questions”, is rated by the respondents as, “strongly disagree” and “disagree” respectively by 36.5%. 43.6% of the teachers, but, responded “strongly agree” and “agree” respectively, while 20.0% of them pointed out that they couldn’t decide either to agree or to disagree. The weighted mean of the above item, 3.03, is, therefore, in the category of “undecided”, which implies that teachers don’t have clear ideas whether those absolute terms, “always”, and “never” are used in the true-false item construction.

Item 1.13 (table 7), states whether the respondents use the domains of educational objectives in their true-false test constructions. The item was rated positively by 64.7% of the teachers and opposed by 20% of them. Only 15.3% of the respondents rated “undecided”. To sum up this data,

the mean score, 3.75 lies in the category “agree”, which indicates that almost the majority of the teachers agreed that the domains of educational objectives: cognitive, affective and psychomotor domains are employed while constructing the true-false item. This idea is definitely supported by many educational theories.

Item 1.14 reads “Teachers consider the level of difficulty of the questions against the level of students”. This statement was answered by 32.9% of the respondents by ticking “strongly disagree” and “disagree”. 51.8% of the teachers responded “agree” and “strongly agree”, while 15.3% of them rated “undecided”. The average score of the responses is 3.29, which falls in the category of “undecided” in the rating scale. The weighted mean implies that teachers couldn’t decide whether the true-false item is prepared considering the level of understanding of the learners.

Item 1.15 (table 7), which is the last item in the table related to true-false test construction, stated “Long and complete statements are used” is rated as “strongly disagree” and “disagree” by 55.3% of the respondents and 25.9% them had ticked under the category “agree” and “strongly agree” respectively. The remaining respondent 18.8% responded “undecided”. The mean score (2.58) falls in the category of “undecided” indicating that the respondents are not sure of the type of statements used in the items. Overall, with the exception of focus on the learned materials and basing items on the domains of educational objectives and use of negative statements which were rated as ‘agree’, all the remaining items were rated “undecided”.

Table 8: Completion Type of Item Constriction

No	Item Descriptions	Rating Scales /Responses/										Mean ($\bar{X}$)
		SD		DA		U		A		SA		
		N	%	N	%	N	%	N	%	N	%	
2.1.	Teachers prepare items in accordance with the domains of educational objectives	2	2.4	6	7.1	4	4.7	36	44.7	35	41.2	4.05
2.2	Teachers usually ask common key terms	2	2.4	11	12.9	11	12.9	42	49.4	19	22.4	3.76
2.3	Teachers usually give one or two blanks are at the end	5	5.9	9	10.6	12	14.1	40	47.1	19	22.4	3.69
2.4	Clear instruction is not formulated usually	27	31.8	29	34.1	6	7.1	13	15.3	10	11.8	2.41
2.5	Terms are usually asked from text directly	7	8.2	15	17.6	15	17.6	31	36.5	17	20	3.42
2.6	Trivial issues are usually asked.	9	10.6	18	21.2	16	18.8	38	44.7	4	4.7	3.11
2.7	The sentences for completion are sometimes difficult ones.	4	4.7	19	22.4	22	25.9	32	37.6	8	4.0	3.24
2.8	Time allotted and point value (mark) is not usually specified.	21	24.7	28	32.9	9	10.6	20	23.5	7	8.2	2.57
2.9	Blanks are indicated at the end	6	7.1	13	15.3	17	20	38	44.7	11	12.9	3.41

Key: SD= Strongly disagree DA= Disagree U= Undecided A= Agree SA= Strongly Agree

Under section II of completion type of item construction, the respondents were asked to rate item 2.1 of table 5, as shown above, which is about the use of domains of educational objectives while preparing tests. Accordantly the teachers who marked “strongly disagree” and “disagree” accounts only 9.5%. The ones who marked “agree” and “strongly agree” are 85.9%, which is the majority of the respondents. Among the respondents, only 4.7% didn’t decide. The mean score for this data is 4.05, which falls under the rating scale “agree”, implying that teachers construct completion types of items related to the domains of educational objectives.

Item 2.2 (table 8), which reads “usually common key terms are used”, was rated “strongly disagree” and “disagree” by 2.4% and 12.9% of the teachers respectively, while “undecided,” “agree” and “strongly agree was rated by 12.9%, 49.4% and 22.4% respectively. This means that the majority of the secondary school teachers, 71.8%, feel that common key terms are usually used in the construction of completion type of items. The mean score, calculated as 3.76, also reveals the phenomena falling under the rating, “agree”.

The teachers who were asked to respond item 2.3 in the same table which is read as “one or two bleaks are indicated at the end”, was rated opposed by 16.5% and positively by 69.5% of them. On the other side, 14.1% of the teachers marked “undecided”. The mean score for this data is 3.69, falling under the rate of agreement, “agree”. The implication of this mean is that, at the end of the completion, item, one or two blanks of the same length is acceptable.

In item 2.4 (table 8), the respondents were requested to reflect whether clear instruction is formulated. For this item 31.8% and 34.1% of the teachers rated “strongly disagree” and “disagree” respectively. 14.1% of them ticked “undecided”, while 15.3% and 11.8% of the respondents marked “agree” and “strongly agree”. Respectively most of the respondents 65.9% didn’t agree that clear instruction is not given. The mean score of the data is 2.41 which shows the disagreement, which implies that most of the teachers believe that clear instruction is prepared in the construction of class room tests supporting the guide lines of test construction, of completion types.

Regarding item 2.5 (table 8), 25.8% of the teachers disagreed in their rating and 17.6% didn’t decide whether terms are usually asked from texts directly, but56.5% of the teachers did agree that terms are mosly taken from texts. To generalize, the mean score found is 3.42 which lies

in the category of the rating scale “undecided”. The implication is that teachers were reserved to give an idea whether terms are asked from texts directly in completion type of items.

Item 2.6 (table 8) which reads “Terms of small importance (trivial) are asked”, was rated “strongly disagree” by 10.6% and “disagree” by 21.2% of the respondents respectively, while 18.8% of them rated “undecided”. The majority of the teachers had a positive idea towards the item in which 49.4% of them were the supports, where 44.7% and 4.7% rated “agree” and “strongly agree”. The mean average (3.11) lies in the category of “undecided”, meaning that teachers did not agree or disagree if terms of small value are usually asked in the completion type of items. The guide lines of test construction opposes the tendency of using terms of small importance in any types of item.

As shown in the table above, respondents were asked to rate item 2.7 (table 8), which reads “The sentences for completion are sometimes difficult ones”. Accordingly 4.7% and 22.4% of the respondents revealed their strong disagreement and disagreements respectively. Only 37.6% and 4.0% of the teachers indicated their strong agreement and agreement respectively, while 25.9% of the respondents couldn’t decide. The weighted mean, 3.25, however shows that respondents couldn’t’ decided or couldn’t show their views whether the statements of completion types of tests are difficult or not at all.

Item 2.8 (table 8), which reads “Time allotted and point value (mark) is not usually specified”, is rated “strongly disagree” and “disagree” by 24.7% and 32.9% of the respondents respectively, while 10.6% of them rated “undecided”. Furthermore, “agree” and “strongly agree” was rated by 23.5% and 8.2% of the teachers respectively. From this data, it could be observed that the majority of the respondents, which is 57.6% didn’t agree on the premise of this item and the mean score indicates 2.57

which lies in the category of the rating scale "undecided", implying that teachers have no clear idea whether time and mark is specified in the item constructed.

Item 2.9 (table 8), which is read as "Blanks are indicated at the end" is rated by 7.1% and 15.3% of the respondents as "strongly disagree" and "disagree" respectively. The category "Undecided" was rated by only 20.0% of the respondents. On the other hand "agree" and "strongly agree" was rated by 44.7% and 12.9% of the respondents respectively, showing that the majority of them (57.6%) supported that, the item. Any how the mean score reads 3.41 lying in the category "undecided", which implies that the respondents didn't agree or disagree on whether blanks are indicated at the end of completion types of the items.

One can conclude that most of the teachers responded "undecided" for the items 1.1 to 2.9, that is, they couldn't agree or disagree on the statements.

Table 9: Multiple- Choice Item construction

No	Item Descriptions	Rating Scales /Responses/										Mean ($\bar{X}$)
		SD		DA		U		A		SA		
		N	%	N	%	N	%	N	%	N	%	
3.1.	The statement questions are based on the domains educational objectives	1	2.4	6	7.1	4	4.7	42	49.4	31	36.5	4.10
3.2	The stem of the item is formulated with blank at the end	6	7.1	22	25.9	23	27.1	28	32.9	5	5.9	3.04
3.3	Clear in instructions are not given very often.	32	37.6	32	37.6	6	7.1	9	10.6	6	7.1	2.11
3.4	Negatives are used in the statements.	2	2.4	13	15.3	13	15.3	45	52.9	12	14.1	3.16
3.5	Answer choices are randomly arranged	9	10.6	17	20.0	9	10.6	33	38.8	17	20.0	3.37
3.6	There appears only one or another alternative answer.	16	18.8	19	22.4	17	20.0	27	31.8	6	7.1	2.85

Key: SD= Strongly disagree DA= Disagree U= Undecided A= Agree SA= Strongly Agree

Concerning item 3.1 (table 9) of section III, the multiple choice item construction, the respondents were requested to rate this item which reads “the statement questions are based on the domains of educational objective” was rated “strongly disagree” and “disagree” by 2.4% and 7.1% of the respondents respectively, while 4.7 % of them rated “undecided”. The item indicated above get acceptance by 49.4% and 36.5% which is the majority of the respondents accounted for 85.9%, where the mean registered for this item is 4.10 implying the agreement of the majority of the respondents. Hence, in constructing the multiple choice type of item, the respondents agreed in using the domains of educational objectives, following the guide line of test construction.

Item 3.2 (table 9) which reads “The stem of the item is formulated with blank at the end” was opposed by 32% of the respondents, 27.1% of them marked “undecided” and 38.8% of the respondents had a positive idea about the item above. The mean of the score is 3.04 which falls in the category of “undecided”, which implies that the respondents couldn’t agree or disagree about the stem of the multiple choice item.

Item 3.3. (table 9) which reads “clear instructions are not given very often, reacted by 37.6% and 37.6% as “strongly disagree” and “disagree” respectively. Only 7.1 of them marked “undecided” Those who show their agreement (strongly agree and agree) with the statement account for 17.7% of the respondents. The majority of the teachers, 75.2% defended that clear instructions are indicated in the class room tests. The weighted mean for the item is 2.11, showing that the mean falls in the gatasury of “disagree”, which implies that teachers prepare clear instructions in the construction of multiple choice type of item.

In responding to item 3.4 (table 9), 27.7% of the respondents, in sum, rated “strongly disagree” and “disagree” respectively while 15.3% of the respondents rated “undecided”. Besides, 67.0% of the respondents rated

“agree” and “strongly agree” respectively, ticked by the majority of the teachers. The mean score for the item, is, hence, 3.61 which falls in the category of “agree”, implying that teachers supported the idea that negatives are used in the question statements of multiple choice test construction.

Item 3.5 (table 9), which reads “answer choices are randomly arranged”, is rated “strongly disagree” and “disagree” by 10.6% and 30.6% of the respondents, and only 10.6% them rated “undecided”, while 38.8% and 20.0% of the respondents rated “agree” and “strongly agree”, which is revealed by 58.8% of the respondents. The mean score being 3.37 describes a reservation of decision, which implies that teacher have no idea how the answers for the classroom test (Multiple choice) are arranged. The guide for test construction suggests that answers for multiple choice item must be set randomly.

Item 3.6 (table 9), which was stated as “There appears only one or another alternative answer”, was responded “strongly disagree” and “disagree” by 18.8% and 22.4% respondents respectively, only 20.0% of the respondents rated “undecided”, while 31.8% and 7.1% of them responded “agree” and “strongly agree” respectively. The mean of the item is 2.85, in which the respondents couldn’t decide to agree or disagree on the condition of the alternative answers in the classroom test construction of multiple choice type of items.

Table 9: Continued

No	Item Descriptions	Rating Scales /Responses/										Mean ($\overline{X}$)
		SD		DA		U		A		SA		
		N	%	N	%	N	%	N	%	N	%	
3.7	Tricky and ambiguous (unclear) options are also used.	21	24.7	22	25.9	13	15.3	24	28.2	5	5.9	2.64
3.8	There are grammatical clues in the question statements	10	11.8	21	24.7	22	25.9	27	31.8	5	5.9	2.95
3.9	The use of “none of the above”, “none”, all of the above is very common.	11	12.9	25	29.4	8	9.4	25	29.4	16	18.8	3.11
3.10	The stem of the question has sometimes no verb.	15	17.6	28	32.9	20	23.5	19	22.4	3	3.5	2.61
3.11	The use of difficult vocabularies, symbols, etc. are seen in the question.	16	18.8	25	29.4	15	17.6	25	29.4	4	4.9	2.71
3.12	There are long questions or statements instead of brief ones.	9	10.6	19	22.4	21	24.7	30	35.3	6	7.1	3.05

Key: SD= Strongly disagree DA= Disagree U= Undecided A= Agree SA= Strongly Agree

In responding to item 3.7 (table 9), 24.7% of the teachers rated “strongly disagree” and 25.9% of them rated “disagree” in favor of the statement, while 15.3% them rated “undecided”. Furthermore, 28.2% and 5.9% of the teachers responded “agree” and “strongly agree” respectively. The mean score is 2.64, which means that the respondents didn’t agree or disagree whether tricky and ambiguous options are used.

In rating item 3.8 (table 9), which was to assess the grammatical clues in the question statements of multiple choice item, 11.8% of the teachers responded “strongly disagree” and 24.7% of them responded “disagree”, while the rest, 25.9%, 31.8% and 5.9% of the respondents, rated “undecided”, “agree” and “strongly disagree” respectively with the average score of 2.95 in which the respondents still didn’t decide to agree or not about the use of grammatical clues in the questions.

In item 3.9 (table 9), which reads “The use of “none of the above”, “none”, “all of the above”, is very common”, was rated “strongly disagree” by 12.9% of the teachers and “disagree” by 29.4% of them, which “undecided”, “agree”, and “strongly agree” were rated by 9.4%, 29.4% and 18.8% of the respondents respectively. Although 48.2% of the teachers agreed, the mean score, 3.11 interprets the disagreement or agreement was not observed.

In item 3.10 (table 9), the secondary school teachers, were also requested to recognize their belief about the stem of the question if it has no verb. In response to this 17.6%, 29.4% and 23.5% of the teachers rated “strongly disagree”, “disagree”, and “undecided” respectively, while 22.4% and 3.5% of them rated “agree” and “strongly agree”, respectively with the mean score of 2.61. This data reveals that the respondents in general didn’t decide to suggest their agreement or opposition.

The teachers were also asked to rate about the use of difficult vocabularies, symbols, etc seen in the questions. In responding to the item 3.11 (table 9) 18.8%, 29.4% and 17.6% of the teacher respondents rated “strongly disagree”, “disagree” and “undecided” respectively, while 29.4% and 4.9% of them responded “agree” and “strongly agree” respectively. For this item, the average score is 2.71 which implies that the teachers couldn’t decide on the item to agree or disagree.

Moreover, teachers’ view about the appearance of long sentences instead of brief ones in the multiple choice item constructions was assessed in item 3.12 (table 9). Accordingly, likewise, 10.6%, 22.4% and 24.7% of the teachers responded “strongly disagree”, “disagree”, and “undecided” respectively. On the other hand, as seen from the data, 35.3% and 7.1 of them rated “agree” and “strongly agree” respectively. From the data observed, the mean average is 3.05 which fall in the category of

"undecided", reflecting that the respondents were not in a position to suggest their agreement or disagreement about the item indicated above.

Table 9: Continued

No	Item Descriptions	Rating Scales /Responses/										Mean ($\bar{X}$)
		SD		DA		U		A		SA		
		N	%	N	%	N	%	N	%	N	%	
3.13	Very often the statement are not related to the level of difficulty of the students	13	15.3	32	37.6	8	9.4	22	25.9	10	11.8	2.81
3.14	Time allotted and point value (mark) is specified	7	8.2	11	12.9	9	10.6	38	44.7	20	23.5	3.62
3.15	Distracters are plausible or reasonable	9	10.6	8	9.4	22	25.9	32	37.6	14	16.5	3.40
3.16	The correct answer should appear in random order	5	5.9	12	14.1	16	18.8	32	37.6	20	23.5	3.58
3.17	Multiple choice items are used where other item are more appropriate	7	8.2	12	14.1	19	22.4	38	44.7	9	10.6	3.35
3.18	The stem of the item should be clear and meaningful.	4	4.7	4	4.7	10	11.8	42	49.4	25	28.4	3.14

Key: SD= Strongly disagree DA= Disagree U= Undecided A= Agree SA= Strongly Agree

In a similar way, item 3.13 (table 9), was ticked by 15.3%, 37.6% and 9.4% of the respondents as "strongly disagree", "disagree" and "undecided", respectively, and further, 25.9% and 11.8% of them responded "agree" and "strongly agree" respectively. The weighted mean of the item is 2.81. which implies that the respondents couldn't decide whether the question statements are related to the level of difficulty of the learners or not pointing out their agreement or disagreement.

Item 3.14 (table 9), which reads "Time allotted and point value (mark) is specified" was rated by 8.2%, 12.9% and 10.6% of the teachers as "strongly disagree", "disagree", and "undecided" respectively, while 44.7% and 23.5% of the respondents answered "agree" and "strongly agree" respectively, in which the majority of the teachers agreed about the

specification of “time” and “mark” on the items, with the mean score, of 3.61 which lies in the scale category of “agree”, implying the positive response of most of the teachers.

In responding to item 3.15 (table 9), which was stated “distracters are plausible or reasonable”, was marked “strongly disagree”, “disagree”, and “undecided” by 10.6%, 9.4% and 25.9% of the respondents respectively. 54.1% of them had a positive idea with the mean 3.40 which means no decision is made by the teachers, didn’t agree or disagree.

In responses to the item 3.16 (table 9), which reads “the correct answer should appear in random order”, was rated by 5.9%, 14.1% and 18.8% of the teachers respectively as “strongly disagree”, “disagree”, and “undecided”, while 37.6% and 23.5% of the teachers responded positively as ‘agree” and “strongly agree” respectively, where 61.1% of them agree. The mean score is 3.58 which imply that the teacher had a positive attitude on the preparation of answers for the item in random order to avoid clues.

Considering item 3.17 (table 9), which is rated as “Multiple choice items are used where other item formats are used where other item formats are more appropriate” was responded by the teachers in the rating scale as, “strongly disagree”, “disagree” and “undecided”, respectively accounted to 8.2%, 14.1% and 22.4% of them, while 44.7% and 10.6% of the teachers rated “agree” and “strongly agree” respectively which scores the mean of 3.35 falling in to the category “undecided”, meaning that the respondents couldn’t specify to agree or disagree on the item.

In responses to item 3.18 (table 9), which is the last item of section III of the multiple choice item construction, reading “The stem of the item should be clear and meaningful”, was rated by the respondents as “strongly disagree”, “disagree” and “undecided” which account to 4.7%, 4.7% and 11.8% respectively. On the other hand, 49.8% and 28.4 of

them rated “agree” and “strongly agree” respectively where the majority of the teachers (77.8%) agreed. The mean of this item is 3.94 clarifying that the stem of the item must be clear and meaningful was an idea supported by most of the teachers as indicated.

Concluding the responses of the items 3.1. to 3.18, most of the teachers didn’t decide to respond agree or disagree on the statement of the items.

Table 10: Matching Type of item construction

No	Item Descriptions	Rating Scales /Responses/										Mean ($\bar{X}$)
		SD		DA		U		A		SA		
		N	%	N	%	N	%	N	%	N	%	
4.1.	Clear instruction is not usually given	25	29.4	30	35.3	6	7.1	18	21.2	6	7.1	2.41
4.2	Point value or mark is not usually indicated	10	11.8	28	32.9	15	17.6	24	28.2	8	9.4	2.90
4.3	Each list of column A and column B is kept short	3	3.5	17	20.0	9	10.6	43	50.6	13	15.3	3.54
4.4	Homogeneity of the content of the question is not considered	10	11.8	27	31.8	11	12.9	28	32.9	9	10.6	2.98
4.5	The items are prepared in sequential order	6	7.1	20	23.5	17	20.0	26	30.6	15	17.6	3.24
4.6	In both columns equal no of premises and responses sometimes appear	7	8.2	14	16.5	19	22.4	36	42.4	8	9.4	3.28

Key: SD= Strongly disagree DA= Disagree U= Undecided A= Agree SA= Strongly Agree

Regarding item 4.1 (table 10) of section IV, matching type of item construction, the respondents were requested to rate this item which reads “clear instruction is not usually given”, and was, hence, responded “strongly disagree” and “disagree”, by 29.4% and 35.3% of the respondents respectively, while 21.2% and 7.1% of the teachers rated “agree”, and “strongly agree” respectively with 7.1% of them again responded “undecided”, the mean score being 2.41 in which the respondents’ disagreement is the correct implication.

Item 4.2 (table 10), which reads “point value or mark is not usually indicated” was rated by 11.8% and 32.9% as “strongly disagree” and “disagree” respectively. Only 17.6% of the respondents rated “undecided”. Those who showed their agreement (agree and strongly agree) with the statement account for 37.6% of the respondents. Most of the teachers, 44.7%, disagreed that the mark is indicated on the question statement. The mean score, 2.90, identified that the respondents had a tendency not to give a clear decision, to agree or not agree.

In responding to item 4.3 (table 10), reading “Each list of column A and column B is kept short,” 23.5% of the respondents responded “strongly disagree” and “disagree” respectively, while only 10.6% of them replied “undecided”. On the positive side, 65.9% of the respondents rated “agree” and strongly agree” respectively by the majority of the teachers. The mean score for this item is, hence, 3.54, which signifies that there is an agreement among the respondents concerning the item.

Item 4.4 (table 10), which reads “Homogeneity of the content of the question is not considered”, was responded “strongly disagree” and “disagree” by 11.8% and 31.8% of the respondents respectively. Only 12.9% of them replied “undecided”, while 32.9% and 10.6% of the respondents rated “agree” and “strongly agree” respectively. The mean showed, which are 2.98, that the respondents didn’t decide to rate, their agreement or disagreement, where 43.6% disagreed and 43.5% of the respondents agreed, being an equal number.

In responding to item 4.5 (table 10) 7.1% and 23.5% of the respondents rated “strongly disagree” and “disagree” respectively in relation to the statement, while 20.0% of them rated “undecided”. Besides, 30.6% and 17.6% rated “agree” and “strongly agree”, respectively. The average mean score is 3.24 indicating that the teachers couldn’t decide whether the

statements of the question in the matching type are prepared in sequential order on not at all.

In rating item 4.6 (table 10), which was to assess the equal number of premises and responses in the matching type of item, 24.7% of the respondents. Rated “strongly disagree” and “disagree” respectively, on the other hand 51.8% of the teachers responded “agree” and “strongly agree” respectively. While only 22.4 of them rated “undecided”. Even though the majority of the teachers agreed on the statement, the mean score, 3.28, indicate, there is still a problem of rating to agree or not to agree, a neutral side.

Table 10: Continued

4.7	All items may also be edited in different pages	12	14.1	23	27.1	24	28.2	18	21.2	8	9.4	2.84
4.8	Premises and responses are kept short	6	7.1	12	14.1	24	28.2	33	38.8	10	11.8	3.34
4.9	The items are also prepared for small unit of the subject matter.	3	3.5	26	30.6	11	12.9	33	38.8	12	14.1	3.29
4.10	Long premises in column A and short responses in column B is the actual format.	14	16.5	29	34.1	13	15.3	24	28.2	5	5.9	2.72
4.11	For the premise part (col. A), complete sentences are offered?	18	21.2	18	21.2	13	15.3	29	34.1	7	8.2	2.87
4.12	The stem or premise part is usually 6 to 10, while the response part (column B) is 2 or 3 more.	18	21.2	24	28.2	11	12.9	23	27.1	9	10.6	2.76

Key: SD= Strongly disagree DA= Disagree U= Undecided A= Agree SA= Strongly Agree

Item 4.7 (table 10), was asked to identify if the part of the items are also printed in different pages, are also printed in different pages, for the matching item. In this case, summing up, 41.2% of the teachers replied with opposition and 30.6% with rating an agreement and 28.2% of them rated category of “undecided”. Anyway, the mean for this item is 2.84

which falls under the category “undecided”, meaning that the teachers did not decide to agree or disagree.

Similarity, item 4.8 (table 10) was asked to rate if the premises and responses of the matching item are kept short. Therefore, 50.6% the teachers agreed and 21.2% of them didn’t agree. On the other side 12.9% of the teachers didn’t decide. The mean score in this case observed was 3.34 which falls on the ‘undecided” category. There was no decision made for agreement or the vice versa.

Item 4.9 (table 10), which reads “the items are also prepared for small unit of the subject” was rated by 3.5% and 30.6% as “strongly disagree” and “disagree” respectively, while 12.9% of them rated “undecided”, 38.8% and 14.1% of the teachers, but, responded “agree” and “strongly agree” with mean score 3.29 which fall in the category of “undecided”, which implies that the respondents couldn’t decide on the preparation of the matching item for small units of a subject matters.

In responding to item 4.10 (table 10), the respondents were asked to assess about the long premises and short responses in column A and B respectively. Hence, 16.5% and 34.1% of the teachers rated “strongly disagree” and “disagree” respectively and 15.3% of them rated “undecided” while 28.2% and 5.9% of the respondents rated “agree” and “strongly agree” respectively in which the mean scored is 2.72, which lies in the category of “undecided”, which shows the respondents didn’t decide yet concerning the premises and responses of the matching item.

Item 4.11 (table 10), which says “For the premise part, complete sentences are given”. Was asked to assess whether complete statements are offered in column A. as seen in the table, the item was rated “strongly disagree” and “disagree” by 21.2% of the respondents for both and “undecided” is marked by 15.3 of the teachers, while 34.1% and 8.2% of them rated “agree” and “strongly agree”. The mean of the responses is

2.87. this means that the teachers did not decide whether complete sentences are organized under column A or not.

Item 4.12 (table 10), which is the last item of section IV of the matching type of item, was asked to get information about the number of premise and response part in both columns. In rating this item, as indicated in the table, 49.4% of the teachers ticked “strongly disagree” and “disagree” respectively while 12.9% them rated “undecided”. Further, as seen in the data, 37.7% of the respondents, as a sum, ticked also “strongly agree” and “agree” respectively with the mean 2.76. Accordingly, the respondents didn’t agree or disagree as to the interpretation of the mean score, which opposes the guide line of the matching type of item.

To conclude the responses of the items 4.1 to 4.12, most of the respondents couldn’t agree or disagree.

4.2. Analysis of students’ Responses

Table 11: Students’ Responses on the teacher made classroom test

No	Item Descriptions	Rating Scales /Responses/										Mean ($\bar{X}$)
		SD		DA		U		A		SA		
		N	%	N	%	N	%	N	%	N	%	
5.1.	Orientation is not given before exam	20	25.0	25	31.3	9	11.3	17	21.3	9	11.3	2.63
5.2	Exam or test instructions are not clear	34	42.0	25	31.3	4	5.0	11	13.8	6	7.5	2.13
5.3	Test have usually editorial problems	21	26.3	29	36.3	6	7.5	20	25.0	4	5.0	2.46
5.4	Some amount of exam time usually is spent to correct errors during the exam	16	20.0	17	21.3	11	13.8	22	27.5	14	17.5	3.01
5.5	The exam doesn’t correspond with the level of students	20	25.0	19	23.8	12	15.0	14	17.5	15	18.8	2.81

Key: SD= Strongly disagree DA= Disagree U= Undecided A= Agree SA= Strongly Agree

Concerning time 5.1 (table 11) of section V based on the students questionnaires, the students were asked to rate the item which reads “orientation is not given before exam”. And rated “strongly disagree”, and “disagree” “undivided” “agree” and “strongly agree” by 25% 31.3%, 11.3%, 21.3% and 11.3% of them respectively. The mean sore is 2.65 which falls under the category “undecided”, meaning that the respondents didn’t agree or disagree whether orientation is given.

Item 5.2 (table 11) which as stated as “exam or test instructions are not clear”, was responded as “strongly disagree” and “disagree” by 42.0% and 31.3% of the students respectively, while only 5.0% of them rated “undecided”. 13.8% and 7.5% of the ranking students ticked “agree” and “strongly agree”. 73.8% of the respondents opposed the item which means they support that clear instructions are given in the test. The mean score 2.13 reflects also that there is disagreement on the premise of the item.

Item 5.3 (table 11) is about the problem of editing tests or exams. For this, item “strongly disagree” and “disagree” was ticked by 26.3% and 36.3% of the students respective on the other side 7.5 % of them responded “undecided”, which 25.0% and 5.0% the students rated “agree” and “strongly agree” respectively, with the mea average 2.46 which implies that there is disagreement to support the opposite sense, Almost 60.6% of them opposed it, with the idea that there is no problem of edition.

Responding to item 5.4 (table 11), the students were asked to comment on the time lost in correcting exam faults during examination. They rated it as “strongly disagree” and “disagree” by 20.0% and 21.3% of their participation respectively. 13.8% of the students rated “undecided”, while 27.5% and 17.5% of them ticked “agree” and “strongly agree” respectively. Almost an equal number of respondents, 41.5% and 45%

respectively disagreed and agreed which leads the mean average 3.01, falling under “undecided”.

Item 5.5 (table 11), which read “The exam doesn’t cide with the level of student” was marked “strongly disagree” and “disagree” by 25.0% and 23.0% of the students. 15.0% of them marked “undecided”. 17.5% and 18.8% of the respondents ticked “agree” and “strongly agree” respectively. The mean score of this item is 2.81 which falls on the category “undecided”, meaning the students couldn’t agree or disagree whether the exam prepared matehes the individual level of students background.

Table 11: Continued

No	Item Descriptions	Rating Scales /Responses/										Mean ($\bar{X}$)
		SD		DA		U		A		SA		
		N	%	N	%	N	%	N	%	N	%	
5.6	Time given doesn’t’ correspond to the difficulty of the exam.	15	18.8	15	18.8	9	11.3	21	26.3	20	25.0	3.20
5.7	Time allotted is not usually printed on the exam paper	29	36.3	25	31.3	6	7.5	15	18.8	5	6.3	2.28
5.8	Teachers don’t make revision before exams	15	18.8	20	25.0	13	16.3	14	17.5	18	22.5	3.00
5.9	Choices are sometimes be vague and correct answers sometimes more than one choice	7	8.8	19	23.8	17	21.3	25	31.3	12	15.0	3.20
5.10	True-false questions are very long and very few	21	26.3	28	35.0	14	17.5	11	13.80	5	6.3	2.37

Key: SD= Strongly disagree DA= Disagree U= Undecided A= Agree SA= Strongly Agree

Item 5.6 (table 11) was requested to rate whether the time allotted corresponds to the level of difficulty of the exam. Therefore, the students, 18.8%, and 18.8%, of them in the same way, rated “strongly disagree”

and “disagree” respectively. 11.3% of them responded “undecided”, while 51.3% of the students agreed, and strongly agreed, when we see the average, 3.20, meaning they couldn’t give a decision to agree or disagree.

Item 5.7 (table 11), which is stated as “Time allotted is not printed on the exam paper usually” was rated by 36.3% and 31.3% of the respondents as “strongly disagree” and “disagree” respectively. 7.5% of them rated it as “undecided” and 18.8%, 6.3% responded “agree” and “strongly agree” respectively. As deduced from the data 67.6% of the respondents didn’t agree on the subject. Still the mean demonstrates this tendency which is 2.28. The time allotment on the question paper not given any decision.

Considering item 5.8 (table 11) which reads “Revision is not carried out before the exam”, Almost an equal number of the respondents disagreed and agreed. Obviously 43.8% and 40.0% of the respondents opposed it and had a positive idea respectively. Accordingly, 3.00, the mean, anyhow leads to the fact that the students were not able to decide whether revision before examination is given in their schools.

Item 5.9 (table 11), is asked if the choices given in the questions is vague and is more than one, Regarding this item, it was responded by 8.8% and 23.8% of the respondents as “strongly disagree” and “disagree” respectively. 21.3% of them marked “undecided”, and also, 31.3% and 15.0 of them rated “agreed” and strongly agreed”, respectively, which accounts to 46.3% of the majority of the respondents. The mean is 3.20, where no decision was made by the students to agree or not.

Item 5.10 (table 11), which reads “True-False, questions are very long and very few “, was responded by 26.3% and 35.0 as “strongly disagree” and “disagree”, respectively while 17.5% of the respondents rated “undecided”. Besides, 13.8% and 6.3% rated “agree” and “strongly agree” respectively. The mean score of this data is 2.37 which falls under “disagree”, in which 61.3% of the students disagreed according to the

data. The respondents disagreed about the length of true-false questions and their numbers.

Table 11: Continued

No	Item Descriptions	Rating Scales /Responses/										Mean ($\bar{X}$)
		SD		DA		U		A		SA		
		N	%	N	%	N	%	N	%	N	%	
5.11	Matching type of exams are sometimes vague or not clear	18	22.5	26	32.5	8	10.0	21	26.3	5	6.3	2.60
5.12	Copying is not avoided from student's attitude	10	12.5	4	5.0	6	7.5	18	22.5	42	52.5	3.97
5.13	Point value or mark on the question paper is not marked	2	25.0	18	22.5	10	12.5	20	25.0	12	15.0	2.82
5.14	There is lack of control in the exam hall.	14	17.5	17	21.3	7	8.8	11	13.8	17	21.3	3.00
5.15	After exam, is corrected, the question papers are not given back in time	15	18.8	29	36.3	11	13.80	12	15	13	16.3	2.73

Key: SD= Strongly disagree DA= Disagree U= Undecided A= Agree SA= Strongly Agree

Item 5.11 (table 11), was requested to asses if the matching type of items are some, times vague or not clear. Regarding the item, 22.5% and 32.5% of the respondents ticked "strongly disagreed" and "disagreed", respectively, while 26.3% and 6.2% of them ticked "agree" and "strongly agree" respectively. On the other hand, 10.8% of the students marked "undecided". Most of the respondents, 55.0%, disagreed. The weighted mean is 2.60 which falls on the category "undecided", implying that the respondents could not decide whether the matching type of items are sometimes vague.

In responding to item 5.12 (table 11), which is read as "Copying is not avoided from student' attitude", was rated by 22.5% and 5.4% of the

respondents as “strongly disagree”, and “disagree: respectively, while 7.5% of them responded “undecided”. On the other side, 22.5% and 52.5% of them responded “agree” and “strongly agree” respectively. From the data observed, 75% of the respondents, the majority of them replied a positive side with the mean 3.97, which implies that students believed on the fact that copying, an academic cheating was not avoided from the students’ attitude.

In rating item 5.13 (table 11) the students was asked to rate the statement “Point value or mark on the question paper is not marked”. The students who ticked strongly disagree” and “disagree” accounted for 25% and 22.5% respectively, and who marked “undecided” were 12.5%. Looking into the positive side, 25.0% and 15% of them responded “agree’ and “strongly agree” respectively on the category” undecided” which had an interpretation that the respondents couldn’t decide to agree or not to agree.

Item 5.14 (table 11), which reads “There is lack of control in the exam hall”, was rated “strongly disagree” and “disagree”, by 17.5% and 21.3% of the respondents respectively, and 8.8% of them rated “undecided”, while 13.8% and 21.3% of the respondents rated “agree” and “strongly agree” respectively. The mean is 3.00 which falls under the rating scale “undecided”, meaning the respondents didn’t agree or disagree.

Item 5.15 (table 11) which is read as “After exam the question papers are not given back to students in time” is rated “strongly disagree” and “disagree” by 18.8% and 36.3% of the respondents respectively, while 13.8% of the teachers responded “undecided”. The teachers responded positively by rating “agree” and “strongly agree” which accounts to 15.0% and 16.3% of the respondents with 2.73 mean score which is grouped in the category of rating termed as “undecided”, which implies that the respondents couldn’t decide to agree nor to disagree.

Table 11: Continued

No	Item Descriptions	Rating Scales /Responses/										Mean ($\bar{X}$)
		SD		DA		U		A		SA		
		N	%	N	%	N	%	N	%	N	%	
5.16	Exam solutions are told earliest	8	10	18	22.5	11	13.8	24	30	19	23.8	3.35
5.17	Final results, 60% and 100% are not always told earlier by teaches	20	25.0	23	28.8	10	12.5	18	22.5	9	11.3	2.66
5.18	Portions are not covered in time	10	12.5	16	20	8	10	23	28.8	23	28.8	3.41
5.19	Short answer questions are not mostly answered by most of the students	8	38	5	6.3	7	8.8	25	31.3	40	50.0	4.17
5.20	Students of good achievement are not motivated and prized	12	15.0	15	18.8	11	13.8	18	22.5	24	30.0	3.33
5.21	Not announcing exam schedules on time.	26	32.5	20	25.0	7	8.8	16	20.0	11	13.8	2.57

Key: SD= Strongly disagree DA= Disagree U= Undecided A= Agree SA= Strongly Agree

In responding to item 5.16 (table 11), the students were asked to rate the statement “Exam solutions are told earlier”. The respondents, rated it as “strongly disagree” and “disagree” which accounts to 10.0% and 12.5% of their participation respectively. Only 13.8% of them ticked “undecided”. Comparatively, most of the respondents, 53.8%, of them expressed their agreement. The mean average of this item is 3.35 which still lies on the category of the rating scale of “undecided”, which means that the students didn’t decide to agree or not to agree.

Item 5.17 (table 11) which reads “final results, 60% and 100%, are not always told earlier by teachers” was answered using the rating scale by most of the students, 54.5%, as “strongly disagree”, and “disagree”, while

33.8% of them rated as “agree” and “strongly agree”. Only 12.5% the students rated “undecided”. The mean is 2.66, which falls in the catagry of “undecided”, which implies that the respondents couldn’t agree or disagree whether the students’ final results is announced to students earlier.

Concerning item 5.18 (table 11), which emphasizes on portion coverage In time, is responded as “strongly disagree” and “disagree” by 32.5% of the respondents, and as “undecided” by 10.0%, “agree” and “strongly agree” by most of the respondents (57.6%). The mean score of this data is 3.41 which still falls in the category of “undecided”, in which the implication is that the respondents couldn’t agree or disagree about the portion coverage in time.

Item 5.19 (table 11) which reads “short answer questions are not mostly answered by most of students” is responded by the majority of the respondents, with agreement to the statement, accounting to 81.3% of the students. 10.1% of the respondents opposed it which 8.8% of the students answered as “undecided”. The mean of the data scores 4.17, which falls in the category of “agree” implying that students agreed on the difficulty of short answer items.

Item 5.20 (table 11) is requested to asses if students with good achievement are motivated and given prize for encouragement. This item, hence, was responded by the majority of the students, 55.5% as “agree” and “strongly agree”, but 33.8% of them disagreed, while 13.8% of the respondents rated “undecided”. Anyhow, the mean score, 3.33 falls in the “undecided” category, which implies that the respondents were not interested to give an idea of agreement or disagreement.

Item 5.21 (table11), which is the last item of section V of students questionnaire, asks the respondents about the announcement of exam schedule or program in time. Based on this premise, 57.5% of the

respondents disagreed, while 33.8% them had a positive idea about it, on the other hand 8.8% of them kept neutral. The mean score of 2.76 for the above item falls in the category of “undecided”, implying that the respondents didn’t indicate their agreement or disagreement.

4.3. Teacher made Test Assessment

In order to review teacher made classroom tests, this year first semester mid test papers, first semester final test papers, and second semester final test papers of at least three different subjects were collected randomly from Repi, Atlas and Beteseb secondary school.

In reviewing and analyzing the exam papers, the researcher used organization (sequential order), layout of the question paper, attractiveness and clarity of printing, clarity of the language, teachers use of table of specification in setting the exam and teachers application of test construction principles were used as assessment criteria.

Based on the above criteria, the exam papers reviewed show that most of the True-False items have incomplete and unclear instructions; the number of items are mostly less than or equal to five; the statements are very long and the number of True and False statements are not equal, not balanced.

With regard to multiple choice item, the review indicated that instruction is not clear; the destructors or choices are very long; teachers overuse alternatives like “none”, “all of the above”, “A except D”, A and C etc., even in one question itself; the items sometimes have more than one answer.

Similarly, for matching Items, a number of weaknesses were observed from the sample papers. These include inability to use homogenous materials, using very long expressions and putting mostly equal number

of premises and responses on the premise and responses. In relation to completion item, the common problems were putting the blank on the left of the questions, using very long statements which are not clear, and taking statements directly from textbooks. Besides, from the review of the exam papers a number of mistakes were observed. These include printing errors, grammatical errors, absence of proof reading, spelling errors and inability to indicate the weight of each type of items,

4.4. The challenges of Test Construction and Administration

In response to the open-ended and interview questions that asks the challenges encountered teachers, directors and students noted that there was no opportunity of training regarding the concept of test construction and test administration. The role departments is very weak. The majority of students are liable to copy. The relationship between students and the school is not strong. Orientation about exams, the feedback on exam results and revision before the test is not usually, practiced in the schools. Many students fail in science subjects. Very few teachers use the table of specification in setting tests and lastly no discussion is made about the guide lines of test construction in their departments. There are problems concerning the ethics of invigilation among teachers and the practicality of giving continuous assessment.

CHAPTER FIVE: SUMMARY CONCLUSION AND RECOMMENDATION

5.1. Summary

The main purpose of this study was to assess the practices and challenges of test construction and test administration in selected secondary schools (government and private) of Kolfe Keranyo sub city in Addis Ababa, and to investigate how teachers use the test construction standard guidelines in the 4 types of items: true- false, item completion type of item, multiple choice item and matching type of item.

Descriptive survey method of research was used in the study. The main instruments of data collection were questionnaires. Other than these, document assessment was also employed for teacher- made test analysis. The participants in the study were teachers of selected secondary schools teaching grade 9 and 10, ranking students (1st to 3rd) of selected secondary schools and directors and vice directors of secondary schools. They were chosen employing simple random sampling techniques.

The data gathered through questionnaire of teachers and students were analyzed and presented clearly in the tables using frequency, percent, mean and standard deviation.

The data collected from directors' interview, teachers' additional written comments and students' additional written comments were used in the analysis whenever necessary.

From the data analysis made so far, employing the quantitative and qualitative data gathered, the following major findings were drown:

There were good practices investigated in the secondary schools concerning the four (4) types of item constructions cover the following.

Some teachers employ the concept of the domains of educational objectives. The units or lessons are covered in time. Answers to some type of questions are prepared randomly to avoid clues. Time is allotted for each types of item. The logical and sequential order of questions are considered. Clarity of the stem of the main question and proof reading before duplication.

For the specific item types, the number of true and false statements to be equal is not decided. There is disagreement opposed to the rules. (Specified in the review of related literature). The same is true for the length of true and false items.

The use of “none”, “all of the above”, A and B, except etc., as option in the multiple choice is common, it must be avoided Whether the stem of the question must have a verb or not was under hesitation. The characteristics or the nature of the distracters was not recognized by the respondents (teachers).

In the matching type of items also, the premises in column A and the responses in column B must be short, and homogeneity of the equations in both columns must be considered, unequal number of premises and responses, must be considered, unequal number of premises and responses, long sentences in column A, and short terms in column B were the actual formats observed.

For completion type of item, blanks must be kept at the end, with the same length. These guidelines were not practically seen in the teats papers or in the responses.

5.2. Conclusion

Concerning true-false item, the teacher respondents had agreed upon the fact that teachers use the classifications of the domains of educational objectives. On the other hand they believed that most of questions asked are the common key terms learnt followed by negatives very often. In multiple choice type of item to time allotted, point value (mark) for every part is specified in test papers. These premises are points of agreement among teachers. In addition to these, the respondents agreed on the idea that the stem of the question of multiple choice type of item must be clear.

About the matching type of item, the concluding remark by the teachers is that each column (A & B) must be kept short is agreed by most of the respondents. On all the remaining parts of the items, the teacher respondents didn't decide at all. This reflects that the teachers are not familiar with the basic gridlines of test construction. Students had strongly suggested that copying in the schools must be avoided gradually be changing the students' attitude.

As observed in the review of some teacher- made question papers incomplete and vague instructions are seen. Some options and destructors are seen. Some options and clues are ambiguous. Negatives and verbal dues are appeared. Most of the questions seem as they are directly taken from learned materials. Most of the teachers were unfamiliar with the use of domains of educational objectives, problem of selecting action verbs for their objectives. For each type of test, the point value or the mark the question carries was not indicated and very long statement question and distracters are used.

5.3. Recommendation

1. Controlling exam- committee must be organized at wereda and school level
2. Training opportunity for both government and private secondary schools must be given.
3. Standard guide for test construction and test administration must be prepared in the schools.
4. Sample test and exam items as well as item bank system besides teacher made tests must be demonstrated in the schools.
5. Strengthening the role of the departments and the school directors in the schools.
6. Programmed open discussion with students regarding the current problems of the teaching and learning process.
7. Employing only the well trained teachers
8. Organization of the school facilities (class rooms desks etc)

References

Allan, G. (2012), Elementary Statistics: A Step by Step Approach, New York: Mc Grawltill.

Borich, Gary D. (2011), Effective Teaching, Boston: Pearson Education, Inc. 7th Edition.

D. Smith, (2005), Educational Measurement and Evaluation...

D.E. Abdullahi and S.A Dnasanya, 2010, Challenges facing the Administrations of Educational Assessment, Measures at the school level in Nigeria, Journal of Applied Sciences, (Http://scialert,net/fultext?doi=jas.2010)

David Hopkins, (2002), Improving the Quality of Education: Books of Staff Development, London: David Fulton Publishers.

Dennis E Hinkle, William Wiersma X others, Applied Statistics for Behavorial Sciences, Boston.

Educational Testing Center (ETC) (2010). Test Administration Procedure, NJ: Princeton

Falchikov Nancy, (2005), Improving Assessment Through Student Involvement, New York: Routledge Falmer.

http,//library-mp.b-berlin.

http://en-wikipedia.org/wiki/Education-in-Ethiopia.

John A. Walthon, (2007), Encyclopedia of Modern Techniques of Educational Testing.

John Mac Beath, (etal), (2001), Improving school Effectiveness, Philadelphia: Open University Press.

Jony Parsons and etal, (2005), How to do your Dissertation in Geography.

Judy W. Eby. (1997), Reflective planning, Teaching and Evaluation, New Jersy: Merril, Upper Saddle River.

Lal J.P (2006), Educational measurement and Evaluation, New Delhr; Anmol Publications, Pvt, Ltd .

Lecil R. Reynolds and Ronald B. Livingston and Victor Willson, Measurement and Assessment in Education, Boston, N. York: Person Publication

Meherens, William A., (1984), Measuremetn and Evaluation in Education, New York, Rinehart.

Noll, Scannel, Craig (1979), Introduction to Educational Measurement, Boston: Houshton, Miffin.

Perrot, Elizabeth, (1982), Effective Teachign: A practical Goide to Improving your Teaching, New York: Longman London.

Plessis, Joy Du et al (2003). Assessment: A Practical Guide for Teachers. New York: Mc Graw Hill.

Rani, T. Swarupa, (2004), Educational Measurement and Evaluation, New Delhi, Discovery Publication House.

Ranjit Kumar, Research Methodology: A step by Step Approach for Beginners, New Delhi: Sage Productions.

Ravi Parkash, (2005), Vol. 1, Encyclopedia of Educational Research: Historical Research in Education, Delhi: Common Wealth.

Ronal Cohn, (2010), Psychological Testing and Assessment, New York: Mc Graw Hill.

Ronald Jay Cohn and etal, (2005), psychological Testing and Assessment: An Introduction to Test and measurement, 6th, edition, New York: Mc Graw. Hily, International Edition.

Sarita, K. (2005), Improving Assessment, Evaluation and Remedial, New Delhi, Isha Books.

Siaran Surgue, and etal, (2002), Developing Teachers and Teaching Practice: International Research Perspectives, London: Routledge Falmer.

Thornodike, Robert, (1974), Measurement and Evaluation Ladd in psychology and Education, New York: Willey publisher.

Wisker... (2007), Research for Postgraduate students.

WWW.proflesling.com/pdf/lest,professionlaTestingInc,PTI2006.

Appendix A

Addis Ababa University

College of Education and Behavioral Sciences

School of Graduate Studies

Department of Curriculum and Teachers Professional Development Studies

Questionnaire to be filled by Teachers:

General direction: Dear respondent! The main purpose of this questionnaire is to collect relevant data on practices and challenges in constructing or preparing teacher-made tests and to suggest the possible recommendations based on the findings of the study.
Therefore, your sincere cooperation in answering each question is very important according to the direction given under the sections I to V.

Writing your name is Not Required

The individual respondent's data will be kept so confidential and is merely for academic purpose.

Thank you in advance for your cooperation.

Section I.

Direction: Please give the required information by marking "√" in the box indicated according to the question. Give your opinion in the blank spaces where necessary.

1. Name of the school _____ wereda _____ sub city _____
2. Sex: male ☐ Female ☐
3. Age: ☐ 21-30 years ☐ 31-40 years
   ☐ 41-50 years ☐ Above 50 years

4. Service in years

   ☐ Less than 5 years ☐ 5-10 years ☐ 11-15 years

   ☐ 16-20 years ☐ 21-25 years ☐ 26-30 years

   ☐ Above 30 years

5. Qualification ☐ BA/BSC ☐ MA/MSC

   Any other, please specify _____

6. Job responsibility: ☐ Teacher ☐ Director

   If other, please specify _____

7. Field of training/ specialization _____

   Major subject _____

8. Training opportunity after graduation (deployment) indicate by marking "√" more than one response is possible

   8.1. Pedagogical science courses

   ☐ Measurement and evaluation

   ☐ Lesson planning

   ☐ Methodology

   ☐ Active learning

   ☐ Continuous assessment

   ☐ Other related

   8.2. If no training is given on the above, please specify the reasons. _____

   _____

   _____

9. Did you take any professional courses (pedagogy, measurement and evaluation etc during your study at a university?

   Yes ☐ No ☐

Section II: On True False Test Item (Exam) Construction or Preparation

Direction: Mark 5, 4, 3, 2, or 1 according to the information given at the right columns in relation to the situation existing currently about True-False item construction of any subject in you school.

Position of Agreements:

5= Strongly Agree 2= Disagree

4= Agree 1= Strongly disagree

3= Undecided

No	True-False Item construction or preparation	Position of Agreement				
		5	4	3	2	1
2.1	Time allotted and point value (mark) is not specified					
2.2	Clear instruction is not items are equal					
2.3	Number of true and false items are equal					
2.4	The statement questions are equal in length					
2.5	Negative statements are also used					
2.6	Double negative statements are also used					
2.7	Verbal clues or hints sometimes appear in the statements					
2.8	General or commonsense questions are asked					
2.9	Questions are taken directly from the text					
2.10	The learned concepts are mostly asked					
2.11	The questions are sometimes ambiguous (not clear)					
2.12	“Always”, “never” are often used in question statements					
2.13	The question statements are based on domains of educational objectives					
2.14	The statements are related to the level of difficulty of the learner					
2.15	Long and complex statements are usually used					

Section III. About Multiple Choice Item Construction or Preparation

Direction: Mark 5,4,3,2, or 1 according the information given at the right columns in relation to the situation existing currently about multiple choice item construction or preparation of the question papers of any subject in your school.

Position of Agreements:

5= Strongly Agree 2= Disagree

4= Agree 1= Strongly disagree

3= Undecided

No	Construction or preparation of multiple choice items	Position of Agreement				
		5	4	3	2	1
3.1	The statement questions are based on domains of education objectives					
3.2	The stem of the item is formulated with blank at the end					
3.3	Clear instructions are not given very often					
3.4	Negatives are used in the statements					
3.5	Answer choices are randomly arranged					
3.6	There appears only one or another alternative answer					
3.7	Tricky and ambiguous (unclear) options are also used					
3.8	There are grammatical clues (hints for the answer) in the question statement					
3.9	The use of “none the above”, “none” “all of the above” in the test/ question paper is very common					
3.10	The stem of the question has sometimes no verb					
3.11	The use of difficult vocabularies, symbols, etc are seen in the question paper					

No	Construction or preparation of multiple choice items	Position of Agreement				
		5	4	3	2	1
3.12	There are long questions or sentences instead of brief ones.					
3.13	Very often the statement questions are not related to the level of difficulty of the students					
3.14	Time allotted and point value (mark) is specified					
3.15	Destructors are plausible or reasonable					
3.16	The correct answer should appear in random order (in exam alternative position)					
3.17	Multiple choice items are used where other item formats are more appropriate					
3.18	The stem of the tem should be clear and meaningful					

Section IV. About completion or Fill the Blank Type of Item Construction

Direction: Mark 5,4,3,2 or 1 according to the information given at the right column in relation to the situation existing currently about completion type of item construction.

Position of Agreements:

5= Strongly Agree 2= Disagree

4= Agree 1= Strongly disagree

3= Undecided

No	Construction or preparation of completion item	Position of Agreement				
		5	4	3	2	1
4.1	Prepared in accordance to the domain of educational objectives					
4.2	Usually common key terms are asked					
4.3	One or two blanks are indicated at the end					
4.4	Clear instruction is not formulated usually					
4.5	Terms are usually asked from text directly					
4.6	Terms of small importance or value (trivial) are asked					
4.7	The sentences for completion are sometimes difficult ones					
4.8	Time allotted and point value (mark) is not usually specified					
4.9	Blanks are indicated at the end					

Section V. About Matching Type of Item Construction

Direction: Mark 5,4,3,2 or 1 according to the information given at the left column in relation to the situation existing currently. About matching type of item construction of the question papers of any subject in your school.

Position of Agreements:

5= Strongly Agree 2= Disagree

4= Agree 1= Strongly disagree

3= Undecided

No	Matching type of item construction	Position of Agreement				
5.1	Clear instruction is not usually given	5	4	3	2	1
5.2	Point value or mark is not usually indicated					
5.3	Each list of column A and B is kept short (max no 10-12)					
5.4	Homogeneity of the content of the question is not considered					
5.5	The items are prepared in sequential order					
5.6	In both columns equal no of premises and responses sometimes appear					
5.7	All items may also be edited in different pages					
5.8	Premises and responses are kept shout premise in column B is the actual format					
5.9	The items are also prepared for small units of the subject matter					
5.10	Long premises in column A and short premise in column B is the actual format					
5.11	For the premise part (column A) complete sentences are indicated					
5.12	The stem or premise part is usually 6 to 10 in no while the response part (column B) is 2 or 3 more					

1.13. Concerning exam preparation or test construction, what are the current problems and challenges in the secondary school? _____

_____

_____

1.14. What majors must be taken to slove such problem in 5.13 above? _____

_____

1.15. What must be done for the future ahead concerning preparation of examinations _____

_____

1.16. If you have additional comment, please specify _____

_____

_____

Appendix B

Addis Ababa University

College of Education and Behavioral Sciences

School of Graduate Studies

Department of curriculum and Teachers professional Development studies

Questionnaire to be filled by Ranking Students

General Directions

Dear respondent: The main purpose of this questionnaire is to collect relevant data on practices and challenges in construction or preparation and implementation of teacher- made tests and to suggest the possible recommendations based n the analysis and findings of the study.

The randomly selected students of the selected secondary schools are the active participants of the research. Therefore, your sincere cooperation in answering each question seriously is very important. You don't need to write your name.

Your data will be kept so confidential and it is merely for academic purpose only.

Thank you in advance for your co-operation.

Instruction: Mark 5,4,3,2 or 1 according to the information given at the right columns, by marking "√" in the column of the space give your additional information/ opinion in the blank spaces at the end the question depending on the questions, number 22.23 and 25.

Key for Responses: 5 Means = strongly agree 2 Means Disagree

4. Means = Agree 1 Means strongly disagree

3. Means = undecided

Age ☐ Sex ☐ Grade ☐ Rank (1st, 2nd, 3rd) ☐

School Name .....

No	Different Views on Test construction	Rating Scale				
		5	4	3	2	1
1	Orientation is not given before the exam					
2	Exam or test instructions are not clear					
3	Problem of edition/printing exams or test					
4	Time lost in correcting exam faults during the exam					
5	The exam/test doesn't coincide with the level of students					
6	Time given doesn't correspond with the difficulty of the exam.					
7	Time allotted is not printed on the exam paper usually.					
8	Revision is not carried out before the exam.					
9	Choices will sometimes be vague and more than one					
10	True-false questions are very long and very few					
11	Matching type of exams are sometimes vague					
12	Copying is not avoided among students					
13	Point value or mark on the question paper is not indicated					
14	There is lack of control in the exam class					
15	After exam, the question papers are not given back to students in time					
16	Exam solutions are not told earlier.					
17	Final results 60% and 100% are not told earlier by teachers					
18	Potions are not covered in time					
19	Short answer questions are not always answered by most of students					
20	Hard working students of good achievement are not motivated and priced					
21	Exam program is not announced in time.					
22	Exam schedule is not informed earlier by the school.					

23. Other than points discussed in the questionnaire what problems can you suggest during or about examinations _____

_____

_____

24. What must be improved for the future to improve test construction problem _____

_____

_____

25. Any additional comments? (if you have) please specify: _____

_____

_____

Appendix C

□□□ □□□ □□□□□□

□□□□ □□□ □□□□□ □□□□

□□□□ □□□□ □□□□ □□□□□

□□□□ □□□□ □□□□- □□□□ □□□□ □□□□ □□□□□ □□□ □□□□ □□□□ □□□□□□ □□□□□□ □□□□□ □□□ □□□□□ □□□□ □□□□□ □□□□□□ □□□ □□□□□□ □□□□ □□□□□ □□□ □□□□□ □□□ □□□ (MA) □□ □□□□□□□□ □□□□□ □□□□ □□□□□ □□□□□□ 2□ □/□/□□ □□□□ □□□□□ □□□ □□□□ □□□□ □□□□□ □□□ □□□□□ □□□□□ □□□□□□ □□□□□□ □□□□□ □□□□□ □□□□□ □□□□□ □□ □□□ □□□□□ □□□□□□□□ □□□□□ □□□□□□□□

□□ □□□□ □□□□□ □□□□□□□□□□□

□□□□□ □□□□/ □□□□□ □□□□□-

□□□□ □□□□ □□ □□□□□□ □□□□□ □□□ □□□□□ □□□□□□ □□□□□ □□□□□□□□ □□□□□□ □□□□ □□□ □□□ □5□ □□□□□ □□□□ □□ □□ □□□□□□ □□□□ □□□□□ □□□□□ □□□□□ □□□□□ □□ "√" □□□□ □□□□□ □□□ □□□ □□□□□ □□□□□□ □□□□ □□ □□ □□□□□□□□

□□□□ □□□□□□□□ 5= □□□□□ □□□ □□□□□□ □□□ □□

4= □□□□□□ □□□ □□

3= □□□□□□□ □□□□ □□

2= □□□□□□ □□□□ □□

1= □□□□ □□□□□□ □□□□ □□

□□□□□ □□□ …… □□ …… □□□ ……

□□□ …….. (1□□ 2□□ 3□)

□□/□□ □□ ……

□.□	□□□□□ □□ □□□□□ □□□□ □□□□ □□□□□	□□□□□□ □□□				
		5	4	3	2	1

1	□□□□ □□□ □□□□□□□ /□□□/ □□□□□					
2	□□□□ □□□□□ □□□ □□□□□/ □□□ □□□□□					
3	□□□□ □□□□ □□□□ □□□ □□□					
4	□□□□□ □□□□ □□□□ □□□ □□□□ □□□□□□ □□□□ □□□ □□□					
5	□□□□ □□□□□□ □□ □□□□□□□ /□□□□□ □□□ □□□□ □□□					
6	□□□□□ □□□□□ □□ □□□□□□□□ □□□□□ □□□□					
7	□□□ □□□□ □□□□ □□□□□□ □□ □□□□□□□ □□□□ □□ □□□ □□□□					
8	□□□□ □□□ □□□□ □□□ □□□□□□					
9	□□□□□ □□□□□ □□□□□□ □□□ □□□ □□□ □□□ □□□ □□□□□					
10	□□□□□ □□□ □□□ □□□□ □□□□ □□□□ □□□ □□□□					
11	□□□□□ □□□ □□□□□□ □□□□□ □□□					
12	□□□ □□□□□□ □□□ □□□□□□					
13	□□□□□ □□□ /□□/ □□□□□□ □□□□□□					
14	□□□□ □□□□ □□□ □□□□□ □□□					
15	□□□□ □□□ □□□□ □□□□ □□□□ □□□□ □□□□□□					
16	□□□□ □□□ □□□□ □□□□□ □□□□□□					
17	□□□□□ □□□ □ 60% □□ □100 □□□□□ □□□□ □□□□□□					
18	□□/□□□□ □□□□ □□□□□□ □□□□□□□					
19	□□□ □□□ /short answer type/ □□□□ □□ □□□□□□ □□□ □□□□□□					
20	□□ □□□ □□□□ □□□□□ □□□□□□□□/ □□□□□□					
21	□□□□ □□ □□□□ □□□□□□					

22. □□□ □□□□□ □□ □□□□ □□□□□□ □□□ □□□□ □□ □□ □□□?

.....

.....

.....

.....

.....

.....
.....

2.3. □□□□ □□□□ □□□□ □□□ □□□□ □□□□ □□?

.....
.....
.....
.....
.....
.....
.....

24. □□□□ □□□□□□ □□ □□□ □□□□?

.....
.....
.....
.....
.....
.....
.....

Appendix D

Addis Ababa University
College of education and Behavioral Sciences
School of Graduate Studies
Department of Curriculum and Teachers Professional Development Studies

Interview Schedule for School Directors

Direction: The propose of this interview is to collect relevant data on practices and challenges in construing or preparing teacher-made tests and to suggest the possible recommendations based on the findings of the study.
Therefore, your co-operation in taking part in this interview is a paramount importance for the study. Your responses will be kept confidential and is merely for academic purpose.

Please fill free to answer all brief questions, suggestions and opinions correctly.

Thank you in advance for your co-operation

1. How can you follow/ control whether a test or exam is constructed or prepared properly?
2. Do you have awareness on how to construct test? Have you been trained in such courses?
3. Similarly, do teachers know how to construct test in general?
4. What training opportunity is given to teachers on pedagogical courses for instance?
5. What is test administration as a concept and what practices are observed in you school?
6. What do students say about examination/ test and its programs? Do they complain?
7. What do teachers say as a feedback about assessment and its outcomes or results as a whole?

8. What is the role of subject departments and experienced teachers in this regard?
9. What problems, at last do you face concerning assessment and the role of the school?
10. What would be the remedy ahead in such cases?
11. Any additional comments other than the pointed out above, if you have?

Appendix E

□/□□□□□ **1999** □.□□.□□

□□□□□□ □□ □□□□ □□□□

□□□ □□□□ □.□.□.□. □□□□

□□□□□□ □□ 8 “□□□□□ □□□□□ □□□□ □□□□□□ □□□□□ □□□□ □□□ □□□ □□□□ □□□”

“□□□□□□ □□□□ □□□-□□□ /Feedback/ □□□□□□ □□□□□ □□□□ □□”.□□ 8

□□□□□ □□□□□ □ 1.2.12 “□□□□□ □□□ □□□□ □□□ □□□□□□□ □□□□ □□□□□□ □□□□□ □□□ □□ □□□” … □□ 8□1.2.14

1.1.1.10 □□ 7 “□□□□□□ □□□□□□ □□□□ □□□□□ /□□□□□□/ □□□□ □□□□ □□□□ □□□□□ □□□□ □□□□ □□□□□ /Procedures/ □□”

1.2.10. “□□□□□ □□□□□□□ □□□□ □□□ □□□ □□□ □□□□□□ □□/□□□ □□ □□□□□ □□□□ □□□□□ □□□□□ □□□□ □□□□□□” … □□ 8

1.1.17. “□□□□□ □□□□□□ □□□□ □□□ (learning content) □□/□□ □□□□ □□□□□ □□□□ □□□ □□□ □□□□□ □□□□□ □□□□□ □□□□□□□□□□□ □□□□□□□□ □□□□□”… □□ 7

Appendix F

Code of Fair Testing Practices in Education

A. Developing and Selecting Appropriate Tests

Test Developers	Test users
Test developers should provide the information and supporting evidence that test users need to select appropriate tests.	Test users should select tests that meet the intended purpose and that are appropriate for the intended test takers.
A-1. Proved evidence of what the test measures, the recommended users, the intended test takers, and the strengths and limitations of the test, including the level of precision of the test scores.	A-1. Define the purpose for testing, the content and skills to be tested, and the intended test takers. Select and use the most appropriate test based on a thorough review of available information.
A-2. Describe how the content and skills to be tested were selected and how the tests were developed.	A-2. Review and select tests based on the appropriateness of test content, skills tested, and content coverage for the intended purpose of testing.
A-3. Communicate information about a test's characteristics at a level of detail appropriate to the intended test users.	A-3. Review materials provided by test developers and select tests for which clear, accurate, and complete information is provided.
A-4. Provide guidance on the levels of skills, knowledge, and training necessary for appropriate review, selection, and administration of tests.	A-4. Select tests through a process that includes persons with appropriate knowledge, skills, and training.
A-5. Provide evidence that the technical quality, including reliability and validity, of the test meets its intended purposes.	A-5.Evaluate evidence of the technical quality of the test provided by the test developer and any independent reviewers.
A-6. Provide to qualified test users	A-6. Evaluate representative

representative samples of test questions or practice tests, direction, answer sheets, manuals, and score reports.	samples of test questions or practice tests, directions, answer sheets, manuals, and score reports before selecting a test.
A-7. Avoid potentially offensive content or language when developing test questions and related materials.	A-7. Evaluate procedures and materials used by test developers, as well as the resulting test, to ensure that potentially offensive content or language is avoided.
A-8. Make appropriately modified forms of tests or administration procedures available for test takers with disabilities who need special accommodations.	A-8 Select tests with appropriately modified forms or administration procedures for test takers with disabilities who need special accommodations.
A-9. Obtain and provide evidence on the performance of test takers of diverse subgroups, making significant efforts to obtain sample sizes that are adequate for subgroup analyses. Evaluate the evidence to ensure that differences in performance are related to the skills being assessed.	A-9. Evaluate the available evidence on the performance of test takers of diverse subgroups. Determine to the extent feasible which performance differences may have been caused by factors unrelated to the skills being assessed.

Source: Code of Fair Testing Practices in Education, 2004 Washington, DC: Joint Committee on Testing Practices.

B. Administering and Scoring Tests

Test Developers	Test Users
Test developers should explain how to administer and score tests correctly and fairly.	Test users should administer and score tests correctly and fairly.
B-1. Provide clear descriptions of detailed procedures for administering tests in a standardized manner.	B-1. Follow established procedures for administering tests in a standardized manner.
B-2. provide guidelines on reasonable procedures for assessing persons with disabilities who need special accommodations or those with diverse linguistic backgrounds.	B-2. Provide and document appropriate procedures for test takers with disabilities who need special accommodations or those with diverse linguistic backgrounds. Some accommodations may be required by law or regulation.
B-3. Provide information to test takers or test users on test question formats and procedures for answering test questions, including information on the use of any needed materials and equipment.	B-3. Provide test takers with an opportunity to become familiar with test question formats and any materials or equipment that may be used during testing.
B-4. Establish and implement procedures to ensure the security of testing materials during all phases of test development, administration, scoring, and reporting.	B-4. protect the security of test materials, including respecting copyrights and eliminating opportunities for test takers to obtainscores by fraudulent means.
B-5. provide procedures, materials and guidelines for scoring the tests, and for monitoring the accuracy of the scoring process. If scoring the test is the responsibility of the test developer, provide adequate training for scorers.	B-5. If test scoring is the responsibility of the test user, provide adequate training to scorers and ensure and monitor the accuracy of the scoring process.
B-6. Correct errors that affect the interpretation of the scores and communicate the corrected results promptly.	B-6. Correct errors that affect the interpretation of the scores and acommunicate the corrected results promptly.
B-7. Develop and implement procedures for ensuring the confidentiality of scores.	B-7. Develop and implement procedures for ensuring the confidentiality of scores.

Source: Code of Fair Testing Practices in Education, 2004 Washington, DC: Joint Committee on Testing Practices.

C. Reporting and Interpreting Test Results

Test Developers	Test Users
Test developers should report test results accurately and provide information to help test users interpret test results correctly.	Test users should report and interpret test results accurately and clearly.
C-1. provide information to support recommended interpretations of the results, including the nature of content, norms or comparison groups, and other technical evidence. Advise test users of the benefits and limitations of test results and their interpretation. Warn against assigning greater precision that is warranted.	C-1. Interpret the meaning of the test results, taking into account the nature of the content, norms or comparison groups, other technical evidence, and benefits and limitations of test results.
C-2. Provide guidance regarding the interpretations of results for tests administered with modifications. Inform test users of potential problems in interpreting test results when tests or test administration procedures are modified.	C-2.Interpret test results from modified test or test administration procedures in view of the impact those modifications may have had on test results.
C-3. Specify appropriate uses of test results and warn test users of potential problems in interpreting test results when tests or test administration procedures are modified.	C-3. Avoid using tests for purposes other than those recommended by the test developer unless there is evidence to support the intendeduse or interpretation.
C-4. When test developers set standards, provide the rationale, procedures, and evidence for setting performance standards or passing scores. Avoid using stigmatizing labels.	C-4. Review the procedures for setting performance standards or passing scores. Avoid using stigmatizing labels.
C-5. Encourage test users to base decisions about test takers on multiple sources of appropriate information, not on a single test score.	C-5. Avoid using a single test score as the sole determinant of decisions about test takers. Interpret test scores in conjunction with other information about individuals.

C-6. Provide information to enable test users to accurately interpret and report test results for groups of test takers, including information about who were and who were not included in the different groups being compared, and information about factors that might influence the interpretation of results.	C-6. State the intended interpretation and use of test results for groups of test takers. Avoid grouping test results for purposes not specifically recommended by the test developer unless evidence is obtained to support the intended use. Report proceduresthat were followed in determining who were and who were not included in the groups being compared and describe factors that might influence the interpretation of results.
C-7. Provide test results in a timely fashion and in a manner that is understood by the test taker.	C-7. Communicate test results in a timely fashion and in a manner that is understood by the test taker.
C-8. Provide guidance to test users about how to monitor the extent to which the test is fulfilling its intended purposes.	C-8. Develop and implement procedures for monitoring test use, including consistency with the intended purposes of the test.

Source: Code of Fair Testing Practices in Education, 2004 Washington, DC: Joint Committee on Testing Practices.

D. Informing Test Takers

Under some circumstances, test developers have direct communication with the test takers and/or control of the tests, testing process, and test results. In other circumstances the test users have these responsibilities.

Test developers or test users should inform test takers about the nature of the test, test taker rights and responsibilities, the appropriate use of scores, and procedures for resolving challenges to scores.

D-1.inform test takers in advance of the test administration about the coverage of the test, the types of question formats, the directions, and appropriate test-taking strategies. Make such information available to all test takers.

D-2. when a test is optional, provide test takers or their parents/ guardians with information to help them judge whether a test should be taken-including indications of any consequences that may result from not taking the test (e.g., not being eligible to compete for a particular scholarship) and whether there is an available alternative to the test.

D-3. Provide test takers or their parents/guardians with information about rights test takers may have to obtain copies of tests and completed answer sheets, to retake tests, to have tests rescored, or to have scores declared invalid.

D-4. Provide test takers or their parents/guardians with information about responsibilities test takers have, such as being aware of the intended purpose and uses of the test, performing at capacity, following direction, and not disclosing test items or interfering with other test takers.

D-5. Inform test takers or their parents/guardians how log scores will be kept on file and indicate to whom, under what circumstances, and in

what manner test scores and related information will or will not be released, protect test scores from unauthorized release and access.

D-6. Describe procedures for investigating and resolving circumstances that might result in canceling or withholding scores, such as failure to adhere to specified testing procedures.

D-7. Describe procedures that test takers, parents/guardians, and other interested parties may use to obtain more information about the test, register complaints, and have problems resolved.

Source: Code of Fair Testing Practices in Education, 2004 Washington, DC: Joint Committee on Testing Practices.

Appendix G

Test Blue Print for a unit on Subtraction without Borrowing

To show the correspondence between instructional objectives and content areas:

Number	Content Outline	Knowledge	Comprehension	Application	Total	Percent
1	The student will discriminate the subtraction sign from the addition sign	1			1	4%
2	The student will discriminate addition problems from subtraction problems	2			2	8%
3	The student will discriminate correctly solved subtraction problems from incorrectly solved subtraction problems		4		4	16%
4	The student will solve correctly single-digit subtraction problems			6	6	24%
5	The student will solve correctly subtraction problems with double-digit numerators and single digit denominators			6	6	24%
6	The student will solve correctly doule-digit subtraction problems			6	6	24%
	Total	3	4	18	25	24%
	Percent	12%	16%	72%	-	100%

Source: Borich G., 2011, Effective Teaching, Boston: Pearson Education Inc. 7th Edition P. 395

Declaration

This thesis is my original form and has not presented for a degree in any other university and that all sources of my information used for this thesis have been fully acknowledged.

- Name _____
- Signature _____
- Date _____

This thesis has been submitted for examination with my approval as a university research advisor.

Name _____

- Signature _____
- Date _____