ADDIS ABABA UNIVERSITY

COLLEGE OF EDUCATION AND BEHAVIORAL STUDIES

SCHOOL OF PSYCHOLOGY

Validity of Grade 10 Mathematics Model Exam:

In Some Selected Secondary Schools in Oromia Regional State

By

Fikremariam Regassa

April, 2014

Validity of Grade 10 Mathematics Model Exam

(In Some Selected Secondary Schools in Oromia Regional State)

By

Fikremariam Regassa

A Thesis Submitted to the School of Psychology in Partial Fulfillment of the Requirements for the Degree of Master of Arts in Measurement and Evaluation

April, 2014

ADDIS ABABA UNIVERSITY

COLLEGE OF EDUCATION AND BEHAVIORAL STUDIES

SCHOOL OF PSYCHOLOGY

Validity of Grade 10 Mathematics Model Exam

(In Some Selected Secondary Schools in Oromia Regional State)

By

Fikremariam Regassa

Approval of Board of Examiners

______________________________	______________________________
Advisor	**Signature**
______________________________	______________________________
Examiner	**Signature**
______________________________	______________________________
Examiner	**Signature**

ACKNOWLEDGEMENT

This study would not have been possible without the help of my advisor. Dr Dame Abera, thank you for being my advisor. I have always looked up to you, and you never fail to make me a better student of measurement. I have learned that a comment or question from you is always the right one at the right time; and, it is always an improvement. Special thanks go to the NEAEA for providing and facilitating the opportunity to attend my graduate study. Special thanks are due to, Ato Araya G/Egziabeher, Director General of the NEAEA, for his fatherly advice on the direction of my future career is also appreciated. Many thanks are given to Ato Kefelegn Tsige Director of Exam Development and Administration Directorate at the NEAEA, for his comments and suggestions on my research and for his quick response to all questions related to my research. I would like to thank my colleagues in the Exam Development Team at NEAEA. It was great to have them around. Thank you all!!! Especially for your willingness and help during your busy, hard and tiresome working days, when I was in the middle of this work and the final stage of my study. Thanks to the participant of the study who gave their time towards the work. I have learned considerably from the information provided during interviews conducted with school principals and officials at the education bureau of Bishoftu town. I would like to thank my mother, brothers and niece for their endless spiritual support and words of encouragement. Knowing, as I always have, that all my parents, especially my niece Gelila Samuel, would love me through my success or failure made me willing to take this risk and to see it through. I consider Zewide W/Amanuel, Solomon Teferi and Gashaw Tadesse as “Big Brothers”, always ready to spend time with me on queries of any sorts, ranging from research-related matters to personal and family-related concerns. Last but not least my special thank to Hiwot Bekele. You have supported me in every way and unconditionally, Thank. You Very Much!!!

Table of Contents

List of Acronyms

EGSECE –	Ethiopian General Secondary Education Certificate Examination
EGSEC Math Exam –	Ethiopian General Secondary Education Certificate Mathematics Examination
EUEE –	Ethiopian University Entrance Examination
EUE Math Exam –	Ethiopian University Entrance Mathematics Examination
MoE –	Ministry of Education
NEAEA –	National Educational Assessment and Examinations Agency
OECD -	Organization for Economic Co-operation and Development

List of Tables

Abstract

The main purpose of this study was to assess the validity of Grade 10 Math Model Exams by considering some selected secondary schools in Oromia Regional State. Thus, evidences on the content validity and criterion related validity of these exams were gathered and analyzed. In the study, mixed-methods research design was employed. The necessary data were collected from the selected secondary schools by using interview, questionnaire, check-list and document analysis. Questionnaire was administered to 31 teachers currently teach Grade 10 mathematics. Moreover, booklets of Grade 10 Math Model Exams were gathered from the schools and evaluated for appropriateness, relevance and representativeness of the items that were sampled by these exams. By using systematic sampling scores of 230 students on Grade 10 and 12 Math Model Exams, EGSECE and EUEE were also collected and analyzed. The criterion related validity evidences of Grade 10 Math Model Exams were analyzed through correlation and multiple regression analysis accordingly. The findings indicated that the Grade 10 Math Model Exams had moderate criterion related validity but suffered from weak content validity. The main reason for this weakness is that the purpose of the exam was not clearly documented and hence, this situation leads to unsystematic preparation of the exams which results in unintended practices in the education system. Therefore, the study pointed out some directions for possible solutions and suggested for further research in the field.

CHAPTER 1

INTRODUCTION

1.1. Background of the Study

Measuring students' academic achievement using exams that have been developed at school and regional/national levels remain the common practice of education systems in many countries. For instance, member countries of the Organization for Economic Co-operation and Development (OECD) implemented different kinds of examination and assessment systems. On her paper Looney (2011) presented that, OECD countries like Denmark, Finland, Ireland, Italy, South Korea and Sweden use national exams in two or more subjects areas, whereas countries like U.S.A. and Turkey do not use national school leaving examinations (2011, p. 49-51).

On the other hand, in the education systems of African countries the situation is seen in virtue of two historic periods, i.e., during the colonial period and after independence. The web sites of "National Examination Council of Tanzania" (www.matokeo.necta.go.tz) and "The Kenya National Examinations Council" (www.kcse-online.info/knec.ac.ke.) indicated that during the colonial period almost in all education systems of the colonized countries, students' academic achievements were measured by using External Exams from the colonial countries. Moreover, these countries (Tanzania and Kenya) use nationally set and moderated examinations having different names but serving similar purposes. Each exam is developed based on the respective national curriculum of the corresponding educational level in the county.

However after independence some countries used a combination both External and National Exams. For example, as indicated on the "South African Department of Education" site (www.education.gov.za) in South Africa students who are continuing into senior secondary school of grades 10-12 sit for the nationally set and moderated Matriculation Examinations, or an approved alternative such as the Independent Examinations Board (IEB) test series, to obtain the National Senior Certificate (NSC) at the end of grade 12. IEB is a South African independent assessment agency which offers examinations for various client schools and the most prominent in setting the examinations for the school-leaving NSC.

In the Ethiopian context, the policy document, i.e., “Ethiopian Education and Training Policy” (ETP, 1994), under section 3.3 (No 3.3.2) explained that “National examinations will be conducted at grade eight and ten to certify completion of primary and general secondary education respectively” (1994, p.18). Thus, students are required to take nationally (centrally) prepared examinations after they have completed the first cycle of General Secondary Education (or Grade 10) so as to be certified. As the result of this, the exams that are given after the completion of the first cycle of secondary education are known as the Ethiopian General Secondary Education Certificate Examinations (EGSECE).

With regard to students, who completed the second cycle of secondary education (or Grade 12) and who want to continue with their advanced education or training in the Higher Institutions in the fields of specialties of their choice, this policy document (ETP, 1994) under section 3.3 (No 3.3.6) sates that “After the second cycle of secondary education, students will be required to sit for examinations of relevant institutions for admission.” (1994, p. 19). This implies that the students should fulfill the admission requirement of the institution. However, through personal communication with officials at the National Educational Assessment and Examinations Agency (NEAEA) and by consulting lecture notes (or class handout) on “Standardized Achievement Tests” the researcher has found that, the mandate to prepare the common entrance exams (in the form of national exams) for all government institutions was given to the Institute of Educational Research (IER) at the Addis Ababa University (AAU). The handout indicated that, on 24 August 2002 all Higher Education Institutions (HEIs) gave the mandate to the Addis Ababa University (AAU) to develop the entrance examination for HEIs. The AAU in turn gave the responsibility to the IER based on the institute's rich experience in institutional, national and international testing services. It was also decided that the administration and scoring be the responsibilities of the "National Organization for Examinations (NOE)" currently named after National Educational Assessment and Examinations Agency (NEAEA). Based on the decisions of the representatives of HEIs and MOE, the preparation and administration of the entrance examination became effective as of the 1995 E.C. (2003/04) academic year (Desalegn, 2012).

In addition to this, in the policy document (ETP, 1994) under section 3.3 (No 3.3.3), it is evidently explained that that, "In order to- get promoted from one level to the next, students will be required to have a minimum of fifty percent achievement." This implies that, (a) in order to measure students' achievement level and to make the "Pass or Fail" decision the tool that should be used is an "Achievement Test" and (b) the students' achievement scores are interpreted based on criteria referenced interpretation (i.e., a minimum of 50%) which is feasible with the help of Criteria Referenced Tests (CRTs). This is because, Criteria Referenced Tests (CRTs) are measures designed to estimate mastery of an identified unit of a curriculum (Salkind, 2007).

Actually both the nationally prepared exams (EGSECE and EUEE) are "High stakes" by nature, which means, they have either positive or negative impacts on each and every stake-holder in the education system. Amrein and Berliner (2002) defined "High-stakes tests" as;

> *...tests from which results are used to make significant educational decisions about schools, teachers, administrators, and students. (2002: 1)*

Similarly, Kubiszyn & Borich (2004) stated that:

> *'High-stakes' testing refers to the use of tests and assessments alone to make decisions that are of prominent educational, financial, or social impact. Examples include whether (a) a student may be promoted to the next grade ... or graduate from high school; (b) a school principal, or teacher receives a financial reward or other incentive, such as a school being identified as 'exemplary' or 'low performing'... (2004: 9)*

Through preliminary investigation the researcher was able to find that in the secondary schools of Ethiopia, after the completion of Grade 10 and Grade 12 and one or two weeks before the administration of the national exams (EGSECE and EUEE), additional exams that are prepared regionally or at school level are given to students of the respective grade levels. Each of these exams goes by the name "Model Exams" assuming that they model the corresponding nationally prepared exams. This practice begun a long period of time ago and is still going on. The information gathered in advance to the main study indicated that stereotypically the trend of developing Model Exams for students of Grade 10 and Grade 12 at the school or regional levels is to familiarize the students with the type and nature of exams developed at the national level. In other words these Model Exams are considered as "Practice Tests" which give students an idea of what to expect on the national exams and will increase their confidence in taking these exam.

Preparing students for National or Standardized Achievement Tests is not a unique practice to Ethiopian schools, for instance "The Iowa Tests of Basic Skills Practice Tests" are available for grades K–12. These short practice tests are published by the Iowa Test Publisher and give students an idea of what to expect on test day and will increase their confidence in taking the test. In doing so, a few practice questions are given in each of the test categories in the Compete Battery and answers are provided. Students actually practice while they do not emotionally influenced by their answers. Moreover Test Administrators are given script similar to the actual test. These practice tests are untimed.

However since test are just tools, (a) tools can be appropriately used, unintentionally misused, and intentionally abused, (b) like other tools, tests can be well designed or poorly designed, and (c) both poorly designed tools and well-designed tools in the hands of ill-trained or inexperienced users can be dangerous, Kubiszyn & Borich (2004). This implies that unless carefully handled, tests may result in unintended consequences. For instance, replicating or modeling the national exams may enforces classroom teachers to “Teach to the Test” or “Coaching”, that is, teachers teach students only those things they know will be tested in the national exams, spending hours memorizing facts, drilling students on test taking strategies. In relation to this Amrein and Berliner (2002) stated that:

> *When teachers teach to the test, the exams define the curriculum. As teachers become familiar with high-stakes testing programs they analyze the intellectual activities required on tests, see the test questions themselves and use what they learn to give their students an extra edge on the test. By doing so, teachers teach students how to respond to practice test items even if students have never learned the underlying concepts. (2002: 40)*

From the information obtained during the preliminary investigation (prior to the study) the researcher was able to know that, in the Ethiopian context, there is a long staying belief about these Model Exams. According to this belief (which persists to date), the Model Exams are considered as “Practice Test”. However after correcting and scoring these Model Exams, the results are reported as the students’ achievement scores at the end the academic year. This means academic decision is made based on the scores on these Model Exams.

Therefore, the validity of these Model Exams based on which academic decisions are made, need to be assessed so as to investigate whether they fulfill the job they are employed to do (as “Practice Test” and/ or as an “Achievement Test”) in the education system.

As a matter of fact Cureton (cited in Shepard, 1997) said that:

> *The essential question of test validity is how well a test does the job it is employed to do. The same test may be used for several different purposes, and its validity may be high for one, moderate for another and low for a third. (*1997*: 5)*

However, the role of these Model Exams in the education system and their validity were not thoroughly researched. There is also lack of empirical analysis in this regard. Some research studies conducted in the area of high school students’ performance considered one aspect of validity, that is ‘predictive validity’, where secondary school students’ performance used as a predictor variable to predict the students performance either at the preparatory level or at a higher education level (e.g., Kebede, 1991; Legesse, 2006; Surafel, 2004; Wosen, 2007).

Therefore, this study analyzes the overall process, (i.e., starting from item writing and test assembly stage to the decision-making stage) of Grade 10 Mathematics Model Exam prepared at the school level in the Oromia Regional State with respect to the educational testing or/and measurement principles.

Furthermore, in the field of education for every educational testing program there is a purpose. To fulfill this purpose, a test score has a clearly stated interpretation and an intended use. Hence, in our context the Regional or Sub-city Education Bureaus or the schools that are taking the responsibility for the testing program of Mode Exams are supposed to create a logical argument and assemble validity evidences which support that argument. However, in some instances, the assembled validity evidences work against the argument and hence lessen validity.

When this happens the Education Bureaus or the schools should seek and take remedies to reverse the gravity of these negative kinds of evidences. Therefore, by collecting and analyzing available validity evidences of Grade 10 Math Model Exams from some selected schools in Oromia Regional State, this study tries to create an argument which is supported by these evidences. As the result of this the study attempts to identify the drawbacks of these Model Exams from the selected schools and contribute its own solution that leads to remedies. The study may initiate further studies on the validity of Model Exams of other subject areas so as to enhance quality of assessment techniques, which is of course a timely issue in the education system.

1.2 Statement of the Problem

The perception, that has become fixed through being widely held, about Model Exams is to familiarize students with the nature and type of the national examinations that would be developed at the national level as a result of which students will be able to score passing mark on these centrally prepared exams accordingly. From personal communication with education bureaus officials, teachers, students and parents the researcher has a chance to know more about these Model Exams. Hence it is found out that in some cases, decision is made at the school level based on the scores that the students obtained on the model exams. For instance, in some schools of Addis Ababa not only the passing mark that the students obtained in the national exams but scoring a passing mark on the Model Exams is also mandatory in order to continue to the next level in that school, whereas in some other schools this is not the case. On the other hand, in regions like Oromia, the model exams are prepared at regional level, whereas in Addis Ababa these exams are prepared by a team of teachers at Sub-city level, in some schools they are prepared by department of the respective subjects or even by individual subject teachers. Moreover some schools combine the student's classroom performance results with the results of the Model Exams so as to report the students' achievement scores at the completion of the respective grade levels at the end of the academic year and to make their own decision accordingly.

But why all these discrepancies exist? Is there no clearly stated official document that describes the intended purpose/s of these Model Exams that are prepared at the school or regional levels for Grade 10 and Grade 12? Do these Model Exams really serve as "Practice Test" by modeling the National Exams or "Achievement Test" to measure what the students achieve as the result of their schooling? To what extent the validity of these exams are maintained that ensure the uniformity of the job they are employed to do across schools with respect to educational measurement? Due to this and other reasons, can we say that the Grade 10 Mathematics Model Exams really model (replicate) in every aspect those exams that are centrally (nationally) prepared?

Therefore, if these problems persist, remain unstudied and corrective measures shall not be taken accordingly, then any exam that needs to replicate or model such kind of high stakes national exam may results in unintended outcome, such as schools may facilitate conditions and enforce teachers to teach students only those things they know will be tested in the national exams. Shortly teachers will "teach to the test", spend more hours on methods of teaching that enable students memorizing facts, give several tests and drilling students on test taking strategies which may results in "over use of testing". Moreover, beyond the studies on "Predictive Validity" of test scores, (Wosen Moga et al, 2007) there is no practical study that has been undertaken to investigate the validity of instruments used to measure high school students' achievement, such as "Model Exams" since one of its assumed purpose is to measure high school students' achievement.

Urbina (2004, p. 151) indicated that, validity – which is, by far, the most fundamental issue regarding test scores and their uses –depends entirely on the validity evidence we can bring to support any inference, such as deciding whether a student "Pass or Fail" or predicting students' future performance that is to be made on the basis of test results. Therefore, making such inference on the basis of test results is part of the story unless it is supported by establishing a logical argument and assembles validity evidences that the test/ instrument can produce and this is the gap that this study attempts to fill. Thus this study differs from past works in that it focuses to reduce the undesirable effect of the study of predictive validity alone.

Moreover this study assesses the relevance of the Model Exams in the teaching-learning process and their implications in achieving quality education in the country by evaluating the Utility of Testing, which refers to the benefits they bring to decision making. As "Utility" is dependent on the extent to which the use of tests can increase the rate of accuracy of the inferences and decisions we wish to make—over and above what it would be, if we use other available tools. In addition to this the Utility of Testing that also refers to individual and social consequences of test use, which is also called *consequential validity* that can be either positive or negative, have to be assessed in terms of their value implications. Eventually, this study is designed to be carried out in assessing what implication these model exams will have in the Ethiopian education system. In relation to the aforementioned problem this research will try to suggest its own solution by answering the following research questions.

1.3 Research Questions

This research will attempt to answer the following questions

- What is the main purpose or function of preparing and administering Grade 10 Mathematics Model Exam?
- Are the developers of Grade 10 Mathematics Model Exam, at the school levels in Oromia, well aware of the principles of test development and prepare this Model Exam based on these principles?
- To what extent the content validity Grade 10 Math Model Exam is maintained?
- To what extent the criterion related validity of Grade 10 Math Model Exam is maintained and supported by logically related validity evidences?
- Do the reported scores on Grade 10 Math Model Exam in Oromia enable the test users to make accurate decision?

1.4 Objectives of the Study

1.4.1 General Objective

The overall purpose of this study was to assess the extent to which validity of Grade 10 Math Model Exam (prepared and administered in some selected secondary schools in Oromia) is maintained by gathering and analysing the relevant validity evidences with respect measurement principles.

1.4.2 Specific Objective

Specifically, this study intends to:

- Investigate any available official document that describes the main purpose or function of preparing and administering Grade 10 Mathematics Model Exam

- Explore critically the overall processes of preparation (i.e., starting from item writing and test assembly stage to the decision-making stages) of Grade 10 Math Model Exam prepared and administered in the selected school with respect to measurement principles.
- Examine the extent to which the content coverage and the behavioral dimensions are sampled by the items in the Grade 10 Math Model Exams during their development processes.
- Find out whether the scores on Grades 10 Mathematics Model Exams (considering them as “Achievement Tests”) and the scoring process enable the test users to make accurate decision.
- Establish a logical argument, based on evidences, that implies the extent at which the criterion related validity of Grade 10 Math Model Exam is maintained.
- Shade light on and to be a springboard for further study on other subject areas on related issue.

1.5 Significance of the Study

Since the goal of this study is to analyze the validity evidences of Grade 10 Math Model Exams, it contributes some ideas in facilitating conditions that put the “Model Exams” in a favourable position in the education system. As the result of this, the responsible bodies in the testing program of these “Model Exams” will be able to formulate a kind of rules/regulations that clearly indicate the purpose of “Model Exams” and will work to improve upon the isolated variables that have been found to have direct relationships to the relevance of model exams, their implementation and usages. This by its turn contributes a great deal to the educational measurement and assessment orderliness in the education system of the country.

Moreover, the study may provide some evidence that are helpful to understand whether the model exams met their intended goal, i.e., either they really serve as a “Practice Test” by modeling the EGSECE and enables students to obtain a passing score on these centrally prepared exams or serve as “Achievement Tests” to measure students’ academic achievement.

Therefore, this study brings one of the multifaceted problems to the surface by presenting an issue (i.e., the validity of Grade 10 Math Model Exam prepared and administered at the secondary schools level in the Oromia Regional state) that can contribute to future and broader research in the field.

1.6 Scope of the Study

This study is about the validity of Grade 10 Mathematics Model Exam prepared at the selected schools level in the Oromia Regional State. The study aimed to assess and validate this Model Exam with respect to the corresponding centrally prepared exams in relation to measurement principles. Therefore, its findings pay particular attention to indicate the extent at which the validity of this Model Exam is maintained.

Furthermore, this study attempts to answer the questions which are determined by the type of evidence needed as well as the logical relationships – inductive and deductive – that have to be established to address the issues of (a) what we are measuring with the Math Model Exams and (b) what inferences we can draw from the scores of this Model Exams.

Thus the variables to be studied are 1) the *relevance* of the content sampled by the Grade 10 Math Model Exam developed at the selected school in the region of Oromia to the specified domain; 2) the *representativeness* of the contents sampled by this exam with regard to the specifications about the domain that it is designed to be covered. Both deal with *Construct* *representation that is concerned primarily with identifying differences in the test's tasks.* The other variables (which are quantitative) are based on the validity evidences needed to address the significance that the scores on the Grade 10 Math Model Exam may have in matters that go beyond those scores or in realms that lie outside the direct purview of the Model Exam. In other words, exam scores must be shown to correlate with the various criteria used in making decisions and predictions. Hence this study will take the scores on both Mathematics Model Exams of Grade 10 and Grade 12 and that on EGSEC Math Exam as predictor variables to predict the scores on EUE Mathematics Exams which is the criterion variable.

1.7 Limitation of the Study

The study is limited to the validity of Grade 10 Mathematics Model Exams. Although at the outset this study was designed to examine the reliability of these Model Exams that were developed at the school level, lack of knowledge of reliability measurement techniques on the part of teachers and school administrations make the study on the reliability aspect of the Math Model Exam difficult. Hence the absence of reliability is the limitation of the study. Moreover, the study does not include all or majority of the schools in the Region due to time and budget constraints to do so.

1.8 Operational Definitions

Achievement Tests – Tests that are designed to measure students' learning in the context of an educational setting and include a range of samples of behavior or items by which students' knowledge or skill is assessed based on their responses to these items.

Practice Tests –Tests containing items which give students some ideas of what to expect on a standardized national exams and will increase their confidence in taking these exam.

CHAPTER 2
LITERATURE REVIEW

2.1 Educational Tests or Examinations

"Examination" as defined in *The Cambridge International Dictionary of English* "is a ***test*** of student's knowledge or skill in a particular subject area which results in a qualification if the student is successful"(*bolding and italicization of the word "test" is added*). This definition asserts that "Examination" in short is a "Test" and hence in this study these two terms will be used interchangeably. Moreover, Anastasi (as quoted in Domino & Domino, 2006) defined "Test" as it is an "objective" and "standardized" measure of a sample of behavior. A broader definition, by adding the word "*Psychological"* in front of the term "*Test",* is given by Urbina (2004) which is stated as:

> *A psychological test is a systematic procedure for obtaining samples of behavior, relevant to cognitive or affective functioning, and for scoring and evaluating those samples according to standards*. (*2004: 1*).

Furthermore when describing the exact nature of psychological tests, which include educational tests as well, Anastasi & Urbina (2003) added that "psychological tests are tools" which means instruments that help us in gathering information about the test takers or students.

So when we talk of educational tests with measurement perspective, the first two questions that come to our mind are: (*i*) what is the intended purpose of the test? (*ii*) What are the main factors that affect the quality of the test? As a matter of fact the answers to these questions, directly or indirectly, depend up on the quality of the building blocks or units known as "*Test Items*" that make up the test. Regarding this and some important matters related to test items, Urbina (2004) said that:

> *Test items are the units that make up a test and the means through which samples of test takers' behavior are gathered. It follows that the overall quality of a test depends primarily on the quality of the items that make it up, although the number of items in a test, and their sequencing or position within the test, are also matters of fundamental importance. Just as tests are evaluated with regard to the extent to which they meet their intended purposes, individual items must be evaluated based on the extent to which they meet the purposes of the test as a whole. (2004:214)*

As to the essential characteristics of test items, their decisive role and important function, as well as the basic methods of scoring test takers' response elicited by these items, Haladyna (2004) pointed out that:

> *A test item is the basic unit of observation in any test. A test item usually contains a statement that elicits a test taker response. That response is scorable; usually* 1 *for a correct response and* 0 *for an incorrect response, or the response might be placed on a rating scale from low to high. ...the more effort we put into building better test items, the better the test is likely to be. Toward that end, one can design test items to represent many different types of content and cognitive behaviors. Each test item is believed to represent a single type of content and a single type of cognitive behavior. For a test item to measure multiple content and cognitive behaviors goes well beyond the ability of a test item and our ability to understand the meaning of an item response. A total score on a test represents some aggregate of performance across all test items for a specific ability or domain of knowledge. ...the test item is intended to measure some aspect of human ability generally related to school learning or training. (2004: 3-4)*

Therefore this explanation mainly concerned with items in *Achievement Tests*. It also implies that, items of Achievement Tests are designed to measure students' learning in the context of an educational setting and include a range of samples of behavior or items by which students' knowledge or skill is assessed based on their responses to these items. Then by employing systematic correction and scoring methods, students will be judged whether they acquired the necessary knowledge or skill or ability and illegible for the next higher level in the educational ladder.

Moreover, educational measurement experts give more emphasis to three elements. These are: **(1)** ***objectivity***: that is, at least theoretically, most aspects of a test, such as how the test is scored and how the score is interpreted, are not a function of the subjective decision of a particular examiner but are based on objective criteria; **(2)** ***standardization***: that is, no matter who administers, scores and interprets the test, there is uniformity of procedure; and **(3)** ***a sample of behavior***: a test *is not* a psychological X-ray, nor does it necessarily reveal hidden conflicts and forbidden wishes; it is a sample of a person's behavior, hopefully a representative sample from which we can draw some inferences and hypotheses. More generally educational literatures, let us inspect closely and thoroughly into the four main crucial steps in the educational testing processes. These are, *(i)* item writing and test assembly, *(ii)* administering the test, *(iii)* scoring, interpreting and reporting results, and *(iv)* use of test results in decision making, Domino & Domino, et al. (2006). Therefore, when we consider the Ethiopian context in light of the above explanations and as it is evidently explained in the ETP (1994), both EGSECE and UEE are Achievement Tests.

2.2 Testing in Education

Scores that derive their validity from a direct and demonstrable connection between test content and the specifications used in developing the test abound at all levels of education and training in which the outcomes of instruction can be unambiguously defined. As do EGSECE and EUEE, almost all teacher-designed Model Exams and classroom tests are suppose to fit into this category. The primary purpose of these instruments is to gauge educational achievement–

that is, what students have learned through their schooling. Scores from these tests can answer mostly direct questions such as “How much of the specified domain has the learner mastered?” or “What degree of mastery or proficiency has the test taker attained in the skill in question?” Content or domain referenced instruments may be applied for a variety of decisions, including assigning grades in a course, conferring a degree or diploma after a program of study, certifying to take part or to practice in a given field of training as in the case of TVET in Ethiopian context, or even ascertaining readiness to undertake a more advanced level of training. Typically, the decisions based on these tests hinge on the levels of mastery displayed by test takers. These levels of mastery, in turn, may be gauged in terms of simple “pass-fail” determinations based on pre-established criteria as in criterion referenced tests, or percentage grades, percentile ranks in comparison to appropriate normative groups as in norm referenced test.

2.3 Validity

2.3.1 The Concept of Validity and Test Validation

In education many measurement specialists refer to “Validity” as a logical process we follow in testing where what we measure is defined, measures are created, and evidence is sought and evaluated pertaining to the validity of interpreting a test score and its subsequent use. This logical process applies equally to tests and the units that make up the tests; namely test items. In this regard Haladyna (2004) for instance, said that “In fact, a primary source of evidence in validating a test score interpretation or use comes from item development” *(2004:14)*.

Furthermore, Validity – which is, by far, the most fundamental issue regarding test scores and their uses –depends entirely on the evidence we can bring to bear to support any inference that is to be made on the basis of test results. According to the American Educational Research

Association (AERA) (as cited in Haladyna, 2004), in educational and psychological testing, the term *validity* refers to: "...the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests" (2004: 9).

From this definition, it can be deduced that; (a) tests can be evaluated only with respect to one or more specific testing purposes. (b) validation involves confirming the inferences derived from test scores, not confirming the test itself, and (c) evaluating inferences derived from test scores involves gathering and analyzing several different types of evidence. Thus based on the outcome of the analysis on validity evidences and the logical argument established, validity is best considered in terms of categories that specify degree, such as "*high validity*", "*moderate validity*", and "*low validity*" with reference to the specific purpose, appropriateness, the interpretation and use made of results of the test under consideration.

Hence as the main objective of this study is focusing on the validity of Grade 10 Math Model Exam that has been developed at the school level in the Oromia Regional State, attempt is made to investigate, gather and establish logical argument by considering; *content-related,* and *criterion-related evidences* as well as *consequences of the uses* and *interpritations of **results***.

Regarding the major consideration in test/ assessment validation; Miller, Linn & Gronlund (2009) said that:

> *The strongest case for validity can be made when evidence is obtained regarding all four considerations herein. That is, interpretations and uses of assessment results are likely to have greater validity when we have an understanding of (a) the assessment content and the specification from which it was derived, (b) the nature of the characteristic(s) being measured, (c) the relation of assessment results to other significant measures, and (d) the consequences of the uses and interpritations of results. (2009: 73-74)*

2.3.2 Sources of Validity Evidence

In general, the essence of judgments about the validity of test scores centers on the relationship between what the scores represent and the questions test users or decision makers need to answer with the use of tests. The questions we present determine the sources or type of evidence we need as well as the logical relationships–inductive and deductive–that have to be established to address the issues of (a) what we are measuring with tests and (b) what inferences we can draw from test scores. In this respect Urbina (2004) added that:

> *The types of evidence required for validating test score inferences, with the understanding that the proposed interpretation of a test score determines the theoretical framework for its validation. It is important to recognize at the outset that neither the aspects of validity nor the sources or types of evidence associated with them are mutually exclusive. Validation strategies should, in fact, incorporate as many sources of evidence as practicable or as appropriate to the purposes of a test. (2004:161)*

Although many other considerations are relevant to validity, this study focuses on those that are most useful in practical educational settings, i.e., *content, test scores-criterion relationship,* and *consequence* considerations.

2.3.3 Evidence of Validity from the Viewpoint of Students

The relevance and representativeness of test item is also pertinent with regard to an issue that is less substantive than score validity, but is nevertheless quite important. *Face validity* refers to the superficial appearance of what a test measures from the perspective of a test taker or any other naive observer. For instance, both the EGSECEs and UEEs discussed up to this point would have some face validity when used in the contexts we have been discussing. They would appear to test takers to be in harmony with the stated educational purposes of the assessment situations in which they are typically applied. Even though face validity is not necessarily an indication of validity from the psychometric perspective, it is nevertheless a desirable feature of tests because it promotes a close and harmonious relationship in which there is common understanding and acceptance of testing and test results on the part of test takers. If the content of a test appears to be inappropriate or irrelevant to test takers, their willingness to cooperate with the testing process is likely to be undermined. Thus, test developers need to consider the appearance of validity from the perspective of all parties–including students' parents and other nonprofessionals–and, whenever possible, incorporate test content that seems relevant and appropriate to the situations in which a test will be used.

2.3.4 Content Validity

The concept of Content Validity refers to the degree to which a test appropriately represents the content domain it is intended to measure. When a test is judged to have high content validity, its content is considered to be congruent with the testing purpose and with the current notions of the subject matter tested.

Therefore, content considerations are of special importance when we wish to describe how a student performs on a set of items that the test is supposed to represent. Having thus assured ourselves that we have a reasonably representative sample of behaviors, we would have good support for the desired interpretation of test scores or validity, in terms content considerations. In relation to this Miller, et al (2009) added that:

> *The essence of the content consideration in validation, then, is determining the adequacy of the sampling of the content that the assessment results are interpreted to represent. More formally,* ***the goal in the consideration of content validation is to determine the extent to which a set of assessment tasks provides a relevant and representative of the domain of tasks about which interpretations of assessment results are made.*** *(2009: 75)*

2.3.5 Test Items and Response Processes as Sources of Content Validity Evidence

For many testing purposes, such as determining whether students have mastered specific course material, evidence of content validity provides the most compelling argument that the test scores are appropriate for inferring conclusions about examinees' knowledge, skills, and abilities. For this reason, content validity evidence is critically important in supporting the use of a test for a particular purpose. Thus, content validation is an important activity for test developers, and research on improved methods for evaluating test content continues to this day.

As it is well known educational tests are instruments designed to sample behavior that can be linked more or less directly to the inferences we wish to make based on their scores. By and large, these instruments fall in the category of content or criterion-referenced tests. These tests are made up of items that either sample knowledge from a defined content domain or require test takers to demonstrate that they possess a given ability or competence in some skill. Validation procedures for tests of this type are in many ways the simplest and most agreed-upon aspect of test development because the evidence on which inferences are to be made can be defended on logical grounds as well as by demonstrable relationships between the content of the test and the construct that the test is designed to represent. Urbina (2004), for instance, stated that:

> *Evidence of test score validity that derives from the content of the test can be built into a new instrument at the outset, by the choice of items or tasks that are included in the test. The primary requirement for developing tests of this sort is a careful specification of the content domains, cognitive processes, skills, or types of performance to be sampled by the test and of their relative importance or weight. (2004: 162)*

In the Ethiopian context, for instance, such specifications for EGSECE are developed by using curricula or course syllabi, textbooks and other materials that outline, define, or prioritize the outcome objectives of the teaching-learning activities in terms of both content knowledge and performance capabilities.

Literatures on educational measurement indicated that content validity has at least four aspects: *domain definition*, *domain representation*, *domain relevance*, and *appropriateness* of the test construction process. The domain definition process or delimiting knowledge domains and determining the desired outcomes of instruction is within the purview of teachers, and other subject matter experts who design and develop curricula or write the textbooks in various disciplines/subject areas.

Regardless of the setting in which a content-referenced test is applied and once the specifications for the knowledge, skills, or processes to be gauged through the test have been set, content validation procedures involve the critical review and examination of test items from the next two perspectives. The first is the *relevance* of the items sampled by the test to the specified domain; the second one is the *representativeness* of the items sampled by the test with regard to the specifications about the domain that it is designed to cover. As a matter of fact the issue of relevance must rely on a consensus of subject matter experts. Test item writers and developers of tests must support claims of the content-related validity of test scores based on varies supporting documents for tests. When the primary basis of an instrument's validation evidence focuses on the specific content, skills, or cognitive processes it assesses, a description of the systematic procedures used to ensure the relevance and representativeness of test item/content to the specified domains is required. Under these circumstances, each test is believed to be a representative sample of the total domain of knowledge. In addition to test administration and scoring conditions to this area of concern Messick (as cited in Haladyna: 2004) noted that:

> *...the chief concerns are clear construct definition, test specifications that call for the sample of content desired, and attention to the test item formats and response conditions desired. (2004:189)*

Therefore, inadequate representativeness of content coverage can play in undermining both the validity and reliability of content-referenced test scores.

Thus, in the context of this study, any exam that needs to model either the EGSECEs or UEEs at least should provide similar, if not identical, validity evidences to that of the exams to be modeled.

2.3.6 Validity Evidence Based on Test Scores – Criterion Relationships

If the validation of exams were limited simply to describing test takers' ability or performance in terms of the frames of reference, i.e., *Criterion* or *Norm referenced* or to increasing our understanding of psychological constructs and their interrelationships, the sources of evidence already discussed might suffice. However, the valid interpretation of the scores of exams often involve as an inevitable part by applying whatever meaning is inherent in the scores to deal with inferences in a way that is based on practical rather than theoretical considerations is necessary for making decisions about students. When this is the case, validity evidence needs to address the significance test scores may have in matters that go beyond those scores or in the field of interest that lie outside the direct concerns of the test. In other words, test scores must be shown to correlate with the various variables or criteria used in making decisions and predictions. Regarding this Kubiszyn & Borich (2003) described that:

> *A second form of validity evidence is criterion-related validity evidence. In establishing criterion-related validity evidence, scores from a test are correlated*

> *with an external criterion. There are two types of criterion-related validity evidences: concurrent and predictive. (2003: 300)*

2.3.7 Some Essential Facts about Criteria

Merriam-Webster's Collegiate Dictionary (1995) defines *criterion* as "a standard on which a judgment or decision may be based" or "a characteristic mark or trait." Although the plural form, criteria, is often used as a singular, in the present context it is necessary to apply both forms of the word appropriately because they are both central to the purposes of this study. A criterion may also be defined, more loosely, as that which we *really* want to know. This last definition, though less formal than the dictionary's, highlights the contrast between what test scores tell us and the practical reasons why we use tests. In this regard Urbina (2004) said that:

> *For psychological tests that are used in making judgments or decisions about students, evidence of a relationship between test scores and criterion measures is an indispensable, but not necessarily sufficient, basis for evaluating validity. (2004: 182)*

Criterion measures are indexes of the criteria that tests are designed to assess or predict and that are gathered independently of the test in question. Since the nature of criteria depends on the questions one wishes to answer with the help of tests, it follows that validation procedures based partly or entirely on relationships between test scores and criterion measures must produce evidence of a link between the *predictors* (test scores) and the criteria. Criterion measures or estimates may be *naturally dichotomous* (e.g., graduating vs. dropping out) or *artificially dichotomized* (e.g., success vs. failure); *polytomous* (e.g., University Degree Program vs. College Diploma Program vs. other Occupational Training Programs); or *continuous* for instance, grade point average, scores on an achievement test, etc. (Urbina, 2004). Thus the nature of criteria depends on the decisions or predictions to be made with the help of test scores. The methods used to establish the relationships between test scores and criteria vary depending on the formal characteristics of both test scores and criterion measures. When criterion measures are continuous (e.g., scores on achievement tests, grades, etc.) the principal tools used to indicate the extent of the relationship between test scores and the criterion measure are correlation coefficients. However, if a certain value on a continuous criterion, such as a 2.0 grade-point-average, is used as a cutoff to determine a specific outcome, such as eligible to the preparatory schools, scores on the predictor or test score can also be evaluated in terms of whether they differentiate between those who meet or exceed the criterion cutoff and those who do not.

Therefore according to Urbina (2004) the following facts about criteria are now generally understood. These are:

1. In most validation studies, there are many possible indexes (both quantitative and qualitative) that can qualify as criterion measures, including scores from tests other than those undergoing validation. Therefore, careful attention must be given to the selection of criteria and criterion measures.
2. Some criterion measures are more reliable and valid than others. Thus, the reliability and validity of criterion measures need to be evaluated, just as the reliability and validity of test scores do.
3. Some criteria are more complex than others. As a result, there may or may not be a correlation among criterion measures, especially when criteria are multifaceted.
4. Some criteria can be gauged at the time of testing; others evolve over time. This implies that there may or may not be substantial correlations between criterion measures that are

available shortly after testing and more distant criteria that may be assessed only over a longer period of time.

5. Relationships between test scores and criterion measures may or may not generalize across groups, settings, or time periods. Therefore, criterion-related validity evidence needs to be demonstrated in a different way for populations that differ from the original validation samples in ways that may affect the relationship between test scores and criteria, as well as across various settings and times.
6. The strength or quality of validity evidence with regard to the assessment or prediction of a criterion is a function of the characteristics of both the test and the criterion measures employed. If the criterion measures are unreliable or arbitrary, indexes of test score validity will be weakened, regardless of the quality of the test used to assess or predict the criteria.

2.3.8 Criterion-Related Validity

The criterion-related decisions for which test scores may be of help can be classified into two basic types: (a) those that involve determining a person's current status and (b) those that involve predicting future performance or behavior. In a sense, this dichotomy is artificial because regardless of whether we need to know something about a person's current status or future performance, the only information test scores can convey derives from current behavior‒that is, from the way test takers perform at the time of testing. As the result of this, criterion-related validation procedures are frequently categorized as either *concurrent* or *predictive*, depending on the measures employed as well as their primary goals.

2.3.8.1 Concurrent Validity

At times we are interested in validity evidences in the relation of test performance to some other current measure of performance. In this case, we obtain both measures at approximately the same time and correlate the results. Hence this procedure for obtaining evidences of validity calls for *Concurrent Validation* study. In this regard Urbina (2004) stated that:

> *Concurrent validation evidence is gathered when indexes of the criteria that test scores are meant to assess are available at the time the validation studies are conducted. Strictly speaking, concurrent validation is appropriate for test scores that will be employed in determining a person's current status with regard to some classificatory scheme, such as levels of performance. (2004: 184)*

Furthermore, she added that:

> *As a result, concurrent validation is often used as a substitute for predictive validation, even for tests that will be used for estimating future performance, such as college admissions or pre-employment tests. In these cases, the test under development is administered to a group of people, such as college sophomores or employees, for whom criterion data are already available. (2004: 185)*

This is commonly done when a test is being considered as a replacement for a more time-consuming method of obtaining information. Moreover Kubsizyn & Borich (2003) confirm that:

> *Concurrent and predictive validity evidences require the correlation of a predictor or concurrent measure with a criterion measure. These types of validity evidence do yield numerical coefficients. By interpreting these coefficients, we can determine whether a test is useful to us as a predictor or as a substitute (concurrent) measure. In general, the higher the validity coefficient, the better the validity evidence for the test is. (2003: 305)*

They go on to say that; in establishing concurrent validity evidence, no time interval (or a very small time interval) is involved between administration of the new test and the criterion or established test. Thus the chances that the behavior or characteristics of examinees measured on both the new test and the criterion test has changed between testings are negligible. Often the distinction between the two kinds of validation procedures hinges on the way test users pose the questions they want to answer with the help of a test.

2.3.8.2 Predictive Validity

Predictive validation evidence, on the other hand, is relevant for test scores that are meant to be used to make decisions based on estimating future levels of performance or behavioral outcomes. Ideally, predictive validation procedures require collecting data on the predictor variable (test scores) and waiting for criterion data to become available so that the two sets of data can be correlated. Predictive validation evidence depends on how the question is posed and on the time horizon that is chosen.

Furthermore Kubsizyn & Borich (2003) added that:

> *"Wherever predictive validity evidence is being considered, it is necessary to be aware that the size of the resulting coefficient is dependent on both the reliability of the predictor and the criterion measure." (2003:307)*

CHAPTER 3

RESEARCH METHODS AND PROCEDURES

3.1 The Research design

In this study, mixed-methods research design was employed. To be more specific it employs the concurrent triangulation designs, because both qualitative and quantitative data are collected and analyzed almost at the same time and they are given equal priority but analyzed independently so as to integrate them at the interpretation stage in order to see the extent to which they converge. (Dame, 2013)

In the course of the study the data related to the purpose of the Model Exam, the consequences of the uses and interpretations of student's score on the exam, are obtained thorough in-depth interview with the respective officials at the sampled schools and education bureaus. Hence the study employed the methods of qualitative data analysis. Furthermore, to investigate the content validity of Grade 10 Math Model Exam the necessary data were collected by using two techniques, i.e., 1) through questionnaire (answered by Grade 10 mathematics teachers from the sampled schools) that deals with the Item writing and assembling, exam administration, scoring and decisions making processes, 2) from experts' evaluation (of the Math Model Exams booklets collected from the sampled schools) of the items with respect to *domain definition*, *domain representation*, *domain relevance*, and *appropriateness*. In his seminal treatise on validity, Messick (as cited in Domino & Domino, 2006) described that, "Tests are imperfect measures of constructs because they either leave out something that should be included . . . or else include something that should be left out, or both". Therefore these data help to identify whether these potential imperfections exist or not during the development of this Model Exam.

In addition to this the other quantitative aspect of the study is Criterion-related validity of Grade 10 Math Model Exams. Thus the corresponding data (scores on both Grade 10 and Grade 12 Math Model Exams, EGSEC Mathematics Exam and EUE Mathematics Exam) were collected by consulting documents and rosters from the respective institutions. Hence, for this quantitative aspect of the study, correlational design is employed to investigate the two types of Criterion-related validity, namely the *concurrent validity* and *predictive validity* of Grade 10 Math Model Exams.

3.2. Description of the study site

Out of the nine Regional States and two City Administrations in Ethiopia, the study area is selected from the Oromia Regional State. According to the Statistical Abstract (2009/2002E.C.), issued by Central Statistical Agency (CSA), the region covers an area of 284,537.83km^2. Oromia has 20 zones (i.e., 17 Zones and 3 Special Zones) and 275 Woredas. As described in the Statistical Abstract (2009/2002E.C.) the region has a population of 29,737,371 out of which 50.36 % are Females and 49.64% are Males. According to the data obtained from the National Educational Assessment and Examinations Agency (NEAEA), in Oromia Regional State there are 501 Government and 67 Private or Non-Government Secondary Schools and 179 Government and 24 Private or Non-Government Preparatory Schools. Moreover at the end of the 2003 E.C. academic year 211,311students from the region sat for EGSECE. From these students those who were eligible to pursue their second cycle of secondary education (preparatory education) 48,124 students were sat for EUEE at the end of the 2005E.C. academic year.

3.3 Participants of the Study

The participants who took part in this study form two groups. The first group comprises those who are representatives of the concerned bodies such as Grade 10 mathematics teachers, the school administrators, officials at Woreda or Regional education bureaus. Each of them provides valuable data for this study on validity, i.e., the inference that is made based on the scores of the Grade 10 Model Exam. In particular these mathematics teachers, who were included in the study, were responsible for the development, administration, correction and scoring processes of the Math Model Exams. Currently they are teaching Grade 10 Mathematics in the selected schools.

The second group of participant consists of subject matter experts who were invited to evaluate the booklets of Grade 10 Math Model Exams collected from the sampled schools. These experts were chosen based on their rich experiences in validation of national learning assessment instruments, long teaching experiences in HEIs (at University and Teacher Training Colleges) and Secondary Schools in the field of mathematics. These experts classified each and every item as it was indicated on the syllabi and/or in the exam specification as well as to rate the mean relevance, representativeness, and appropriateness ratings of all items with respect to all domain areas that are supposed to be tested.

3.4 Samples and Sampling techniques

This study was designed to assess the validity evidences of Grade 10 Math Model Exam developed and administered in the year 2003 E.C. in some schools of Oromia in relation to the EGSECE for the same year, the Grade 12 Math Model Exam and the Math EUEE both administered in the year 2005 E.C.

These two academic years were chosen for two main reasons. Firstly students who had taken Grade 10 Model Exam and EGSECE in the 2003 E.C. academic year (which enables to assess the concurrent validity of the Model Exam) also took Grade 12 Model Exam and EUEE after two years (in this case, in the year 2005 E.C). Therefore, according to this arrangement the time interval between the Grade 10 Math Model Exam and the EUEE is almost two years, which granted the maturation of the criterion variable. Therefore, this situation made suitable and fulfilled the assumption for the analysis of predictive validity of Grade 10 Math Model Exam accordingly.

The second reason to choose these two academic years was up-to-datedness. This means the respective data obtained from these two academic years are able to show the current situation of the issue under investigation.

From all Regional States in the country, Oromia was chosen for this study based on accessibility and the study sites were chosen by *Area sampling* method, because the total geographical area of the Oromia region happens to be big one. Therefore, out of the 20 Zones in the region, East Showa Zone was selected by convenient sampling technique due to budget and time constraints. Thus by using the administrative division of the total area of the Zone into a number of non-overlapping 13 Woredas, two neighbouring Woredas, i.e., Ada'a and Dukem Woredas in East Show zone were selected. These two Woredas were chosen for this study based on accessibility.

Following this, Bishoftu town which was the capital of the Ada'a Woreda, was selected purposively for the main study. In Bishoftu town there are 5 secondary schools and one preparatory school. Moreover from Dukem Woreda the capital of the Woreda, i.e., Dukem town was selected for the study. In this town there are one secondary and one preparatory school, where both are found in the town. For this study four secondary schools from the two towns were selected. Three of these schools (two Government and one private school) were chosen from Bishoftu town, while the remaining one government school was selected from the nearby town, i.e., Dukem. The number of schools selected for inclusion in the study was determined as a result of budgetary constraints, and based on the location of the schools and the accessibility of transportation to each school.

Moreover the location of the schools and accessibility of transportation were helpful especially for the distribution and timely collection of questionnaires as well as in making the field interviewing more efficient. However after distributing and collecting the questionnaires for the teachers, due to some unexpected situations the students' score on the 2003E.C academic year from Oda Nabi Secondary School in Dukem town could not be obtained. As the result of this the school was exempted from further analysis of criterion related validity of the Model Exam prepared in that school except the analysis of teachers' responses to the questionnaire.

Therefore, students' scores on Grade 10 and Grade 12 Math Model Exams from the selected secondary schools and the Preparatory school in Bishoftu town were collected within a minimum period of time. According to the data obtained from the town's education bureau, those students who were eligible to pursue their second cycle of secondary education, 758 students from the selected three secondary schools took EUEE in the year 2005 E.C. Hence the study employed probability sampling method and hence, used systematic sampling to select scores of a total of 230 students. This number of students being included in the sample was determined by considering the proportion of the number of students who took the EUEE in 2005 E.C from each school. In doing this, from the rosters of the preparatory school, the scores were systematically sampled just by starting from a random score and selecting every fourth score until the required 230 scores were obtained from all the sampled schools.

3.5 Data collection instruments

For collecting the necessary data, this study used Questionnaire, In-depth interview questions and a Checklist. The questionnaire had five parts where the first four parts contain 50 items. The first part, which was further divided in to 7 sub-groups, contains 34 items. All these items deal with the Test Item writing and Test Assembling process of the Grade 10 Math Model Exam. The second part, having 4 items, was dedicated to the administration process of the Model Exam. The third part which had 7 items deals with the scoring process of the Model Exam, while the fourth part having 5 items was dedicated to the academic decisions based on scores on the Grade 10 Math Model Exam. The last (fifth) part was designed to have "Personal Profile" of the teacher who filled the questionnaire, which helps to ensure that the study had managed to get responses from a good cross-section of mathematics teachers.

Furthermore, interview was conducted with officials of the different concerned bodies, such as School Principals and Officials at the Woreda and Regional Education Bureaus, by using a set of six major questions, in relation to Grades 10 Math Model Exams. These interview questions were prepared in such a way that they could elicit data on the Model Exams through several detailed questions under each of them. These detailed questions were about the role of Model Exams in our secondary schools and directed to the Test Item Writing, Test Assembling, Administering, Scoring, Interpreting and Reporting results, as well as Decision-making processes based on the scores on these Model Exams.

In addition to these, Booklets of Grade 10 Math Model Exams that were administered at the end of the 2003E.C academic year in the sampled school were critically reviewed. The reviewing process was done by the subject matter experts who were chosen to participate in the study. In doing so a table that contains the list of all specific objectives to be measured with the respective Bloom's Taxonomy of educational outcomes, guidelines for developing multiple choice items and a Check List are used. These documents were adopted from different sources and adapted so as to fit for the purpose of the study (see Appendices I – K) and presented to the experts.

3.6 Procedures

3.6.1 Instrument development procedure

In order to collect the necessary and relevant data for the study, a questionnaire and an In-depth interview questions were used.

The interview questions were intended to collect pertinent data from the school administrators and officials at the City or Woreda Education Bureaus on how they regard the Model Exams from their own point of view. Thus the interview questions were prepared in Amharic. This helps to convey the main idea and clarify the theme of the questions to the interviewees so that they can freely express what they fill or think about the issues under investigation. However the questionnaire, which was developed by the researcher, was prepared in English, because there are some technical terms which may not have equivalent and identical meaning in Amharic.

3.6.2 Instrument translation procedure

As mentioned above the interview questions were prepared in Amharic so as to conduct effective interview by presenting the main idea and clarify the theme of the questions to the interviewees without ambiguity. The interview sessions lasted from 30 min up to 45 min. One of the principals of the schools was willing to be recorded, but the others were not. After obtaining the feedback from the interviewees the backward translation, from Amharic to English, was made by an expert in English language. In doing so much care had been taken to transcribe the original ideas or responses as it was explained or stated by the interviewee.

3.6.3 Scoring procedure

Since the questionnaire was used to collect quantitative data, each item was developed so as to be rated with the help of Likert-type scale through which the participants indicated their degree of agreement or disagreement on a five-point range, i.e., the respondent may respond in any one of the following ways: (i) strongly agree, (ii) agree, (iii) undecided, (iv) disagree, (v) strongly disagree. The Likert scaling technique, thus, assigns a scale value to each of the five responses. Response indicating the least favourable degree of agreement was given the least score (i.e., 1) and the most favourable was given the highest score (i.e., 5). Normally these score–values were not printed on the questionnaire but are shown here just to indicate the scoring pattern. In this way the questionnaire yields a total score for each respondent, which would then measure the respondent's favorableness toward the given point of view.

Since the questionnaire consists of 50 statements the following score values would be revealing.

$50 \times 5 = 250$, Most favourable response possible,

$50 \times 3 = 150$, A neutral attitude, and

$50 \times 1 = 50$, Most unfavourable attitude.

The scores for any individual therefore, fell between 50 and 250. If the score happened to be above 150, it showed favourable opinion to the given point of view, a score below 150 meant unfavourable opinions and a score of exactly 150 was suggestive of a neutral attitude. Then the individual scores were added up to form an overall score for each respondent.

3.6.4 Reliability computation procedure

In connection with the questionnaire internal consistency or reliability was particularly important. It helps to answer whether the questionnaire was measuring a single idea and hence the items that make up the questionnaire were internally consistent. A number of procedures for estimating reliability exist but the study employed Cronbach’s alpha. Thus the questionnaire was found to be highly reliable (50 items; $\alpha = .794$).

3.6.5 Data collection procedure

The collected data were both primary and secondary. The sources of the primary data were the participants in the first group above, from which data were collected through questionnaire and interviews.

For the qualitative part of the study the primary data were obtained through an In-depth interview. The interview questions were designed so as to have content mapping (to open up the research territory and to identify the dimensions or issues that are relevant to the participant) and content mining to explore the detail which lies within each dimension, to access the meaning it holds for the interviewee, and to generate an in-depth understanding from the interviewee's point of view.

In addition to this, the other primary data were results obtained from the evaluation of Grade 10 Math Model Exams booklets prepared and administered for the year 2003E.C. in the sampled schools. The evaluation of these booklets was conducted by the subject matter experts. These experts had evaluated the items on the booklets for their representativeness, relevance, and appropriateness.

To facilitate the collection of these data additional supporting materials like, Guidelines For Developing Multiple Choice Items, List of all Specific Objectives to be measured and their respective classification according to Bloom's Taxonomy of the educational outcomes, were compiled by the researcher and given to the subject matter experts.

The secondary data (i.e., students' scores on Grade 10 and Grade 12 Math Model Exams and on the corresponding national exams) were collected from rosters in the schools records and from the NEAEA accordingly.

3.6.6 Data analysis procedure

In this study both qualitative and quantitative data analysis techniques were employed. A common procedure in the analysis of qualitative data is the identification of key themes, concepts or categories. Therefore, in the study thematic analysis was employed to analyze qualitative data so as to move from raw data to meaningful understanding through the process of generation/exploration of relevant themes. Many of these themes were discovered through inductive analysis; they were also identified through their agreement with the research question. Once the themes or concepts were inductively identified and deductively verified, then effort was made to look for interconnections between them. (O'Leary, 2004, p. 197 – 198)

In order to analyze the content validity of the Grade 10 Math Model Exams, the Booklets of these Exams collected from the sampled schools were given to the subject matter experts. The experts were asked to match each item to the content area, educational objectives (or learning outcomes) that are supposed to be measured as they were indicated in the curriculum. In addition to this the experts were also asked to rate the relevance or appropriateness of each test item in the Model Exams with respect to the corresponding content area or objective that it was intended to measure. The data gathered from this evaluation process were summarized using simple descriptive statistics, such as the proportion of experts who classified an item as it was listed in the curriculum or the mean relevance ratings for an item across each area to be tested. Using these techniques, this study tries to assess the content validity evidence of the Math Model Exams developed at the schools. As a matter of fact this had to be done based on the "Table of Specification", but these Math Model Exams were developed without this essential document.

Furthermore to estimate the criterion related validity of the Model Exams the correlation and multiple linear regression analyses were employed accordingly. These were summary of the relationship between criterion variable and the respective independent variable or set of predictor variables. The most important product of the multiple regression analysis was the explicit determination of the regression model. Thus, these methods were applied in two ways. The first one was to find out the strength of the correlation between scores on the school based Grade 10 Math Model Exam and the corresponding scores on the EGSEC Math Exam. The second one was performed by considering the scores on both Grade 10 and Grade 12 Math Model Exams together with the scores on EGSECE Mathematics Exam as predictor variables so as to predict the scores on the University Entrance Mathematics Exam which is the criterion variable accordingly.

Therefore to analyse these quantitative aspects, i.e., criterion related validity of the Math Model Exam, the study employed the correlation and multiple regression analyses to analyse the *concurrent validity* and *predictive validity* respectively. These analyses were viewed as a set of data analytic techniques that were used to understand the interrelationships among variables, i.e., the two Math Model Exam and the corresponding two national exams in the secondary schools. Moreover, the study had employed a hierarchical method in the multiple regression analysis so as to examine the effect of including a predictor variable in the preceding model. As the result of this during the multiple regression analysis the 2003 E.C Grade 10 Math Model Exam was included and analysed as a predictor variable (either individually or pair wise with the EGSEC Math Exam of 2003 E.C and the Grade 12 Math Model Exam of 2005 E.C). This is due the fact that the study was designed to investigate the validity of Grade 10 Math Model Exam.

CHAPTER 4

RESULTS

The primary objective of this study was to identify, gather and analyze the validity evidences of Grade 10 Math Model Exams prepared at the school level in Oromia Regional State and to assess the extent to which its validity was maintained in the testing program. Thus, to achieve this objective data obtained from, the Questionnaire, Interviews, Evaluation of items in the Booklets of the 2003 E.C. Grade 10 Math Model Exams as well as the related criterion related validity evidence were analysed. Therefore, the obtained results or findings are presented as follows.

4.1 Teachers Responses

Three days after the date on which the questionnaires were distributed, an attempt was made to collect them. Finally, 31 mathematics teachers returned fully completed questionnaires with their feedback.

Tables that are given below indicate the frequency of responses (the number of teachers who responded) to each item, on the 5-point rating scale, of the questionnaire. Each table is preceded by a description of the main finding that corresponds to the respective part of the questionnaire.

Majority of the teachers need the math model exam to be prepared at the regional level. Table 1, shown below, indicates the frequency of responses and proportion of respondent, for Part I, having questions about the preferable place where the Mathematics Model Exam should be prepared.

Table 1: *Responses to Items Number 1- 4 of Part I*

No	Item	SA	A	Und	D	SD	Total
1	I believe it is recommendable to prepare Mathematics Model Exam at the school level by capable and experienced individual teacher	**9** *29.0%*	**5** *16.1%*	**3** *9.7%*	**13** *41.9%*	**1** *3.2%*	**31** *100.0%*
2	I suggest it is preferable to develop Math Model Exam in the school by a group of selected math teachers from the department	**9** *29.0%*	**10** *32.3%*	**1** *3.2%*	**10** *32.3%*	**1** *3.2%*	**31** *100.0%*
3	I believe it is more advantageous to prepare Mathematics Model Exam at the Woreda or Sub-city or Zone education bureau by the subject experts	**4** *12.9%*	**6** *19.4%*	**0** *0.0%*	**14** *45.2%*	**7** *22.6%*	**31** *100.0%*
4	I think it is better if Mathematics Model Exam will be prepared at the level of Regional education bureaus by selected subject experts.	**14** *45.2%*	**5** *16.1%*	**1** *3.2%*	**6** *19.4%*	**5** *16.1%*	**31** *100.0%*

As indicated in Table 1 above, for Item 4 the responses of majority of the teachers are either “Strongly Agree” or “Agree”. The table of Item Statistics (Appendix B) demonstrates that from the first four items of Part I, Item 4 has the higher mean scores. This implies that the teachers’ response for the question were at the agreement end of the rating scale. In other words the highest mean value for this item clearly shows the importance that the math teachers place on the regionally prepared Math Model Exam.

Among the teachers there is a widely shared perception that the Grade 10 Math Model Exam should cover contents of Grade 9 and Grade 10.

The next table shows the frequency of responses on questions, for Part I, that deal with the units of lesson or content area that the model exam should cover.

Table 2: *Responses to Items Number 5 – 10 of Part I*

No	Item	SA	A	Und	D	SD	Total
5	I think it is quite possible, if Grade 10 Math Model Exam covers all units of Grade 10 only	**4** *12.9%*	**1** *3.2%*	**0** *0.0%*	**15** *48.4%*	**11** *35.5%*	**31** *100.0%*
6	Since Model Exam is a Practice Test I suggest that the items should be selected based on the depth or breadth of contents in each chapter because students can understand the concepts and become confident on EGSECME.	**15** *48.4%*	**13** *41.9%*	**1** *3.2%*	**2** *6.5%*	**0** *0.0%*	**31** *100.0%*
7	Since it is a Model Exam, the items must be prepared from both the contents of Grade 9 and Grade 10 in proportion to the total number periods allotted.	**22** *71.0%*	**6** *19.4%*	**2** *6.5%*	**1** *3.2%*	**0** *0.0%*	**31** *100.0%*
8	I generally believe that items of Grade 10 Math Model Exams should be prepared from both the contents of grades 9 & 10 based on a table of specifications.	**19** *61.3%*	**8** *25.8%*	**3** *9.7%*	**1** *3.2%*	**0** *0.0%*	**31** *100.0%*
9	Since Model Exam is a Practice Test, I strongly suggest that the items should be selected from chapters that the students feel difficult because they can practice how to solve them and become confident on EGSECME.	**2** *6.5%*	**11** *35.5%*	**3** *9.7%*	**9** *29.0%*	**6** *19.4%*	**31** *100.0%*
10	Because it is a Model Exam, I suggest 50% of the items in the exam should be selected from the contents of Grade 9 and the remaining 50% from Grade 10.	**13** *41.9%*	**7** *22.6%*	**3** *9.7%*	**8** *25.8%*	**0** *0.0%*	**31** *100.0%*

For this part of the questionnaire there are six items regarding the Units of Lesson/Content Areas that should be included in the Model Exam. The Item Statistics table of the questionnaire (Appendix: B) shows that four of the questions have mean values that are much greater than the ideal mean 3 with relatively smaller standard deviation. This implies that the respondents have shared almost the same idea that “the Grade 10 Math Model Exam should cover Units of Lesson from Grade 9 and Grade 10.

Moreover there was no great variation in the teachers’ response on how to determine the number of item from a given chapter for the Model Exam. Table 3, which is shown below, indicates the number of teachers responded to items that suggest different options on how to determine the number of questions from a given chapter that the Model Exam suppose to cover.

Table 3: *Responses to Items Number 11- 15 of Part I*

No	Item	SA	A	Und	D	SD	Total
11	I suggest that, for Grade 10 Math Model Exam the number of items from a given chapter should depend on the number of periods allotted, the nature of the learning outcomes to be assessed and the depth or breadth of contents of that chapter.	**11** *35.5%*	**15** *48.4%*	**3** *9.7%*	**1** *3.2%*	**1** *3.2%*	**31** *100.0%*
12	It is recommended by some teachers that the number of items selected from a given chapter should be based on the difficulty or easiness of the concepts in the chapter.	**4** *12.9%*	**11** *35.5%*	**6** *19.4%*	**8** *25.8%*	**2** *6.5%*	**31** *100.0%*
13	I think it is better if the number of items from a given chapter depend on whether that chapter is totally or partially covered in the academic year.	**7** *22.6%*	**9** *29.0%*	**4** *12.9%*	**6** *19.4%*	**5** *16.1%*	**31** *100.0%*
14	Some people say the number of items chosen from a given chapter should be based on the students' ability to answer the questions from that particular chapter.	**3** *9.7%*	**4** *12.9%*	**4** *12.9%*	**15** *48.4%*	**5** *16.1%*	**31** *100.0%*
15	It is possible to determine the number of item from a given chapter after consulting or reviewing the EGSECE booklets of the previous years.	**4** *12.9%*	**12** *38.7%*	**5** *16.1%*	**8** *25.8%*	**2** *6.5%*	**31** *100.0%*

Even if the teachers have the same opinion about the contents of the Grade 10 Math Model Exam, there is an observable difference, as shown on the table, in their opinion on the number of items that should be selected from a given chapter so as to be included in the Model

Exam. This can also be seen from the relatively higher standard deviation shown in the table of Item Statistics (Appendix: B).

In addition to this the respondents have given higher value for the role of curriculum in preparation of Grade 10 Math Model Exam. The next table of frequency, Table 4 for Part I, shows the number of teachers responded to items that provides different options on the role of curriculum in the preparation of Math Model Exam.

Table 4: *Responses to Item Number 16 - 20 of Part I*

No	Item	SA	A	Und	D	SD	Total
16	As much as possible the Grade 10 Math Model Exam should assess all the intended contents and the corresponding depth of knowledge indicated in the "*Formal Curriculum*" or National Curriculum.	**14** *45.2%*	**17** *54.8%*	**0** *0.0%*	**0** *0.0%*	**0** *0.0%*	**31** *100.0%*
17	Since the Grade 10 Math Model Exam is considered as a "Practice Test", I think it is very important if the Model Exam contains items from that portion of the curriculum over which teachers "drill" students on test taking techniques using released items and item format of EGSECME.	**4** *12.9%*	**16** *51.6%*	**7** *22.6%*	**4** *12.9%*	**0** *0.0%*	**31** *100.0%*
18	I suggest that the Model Exam should be prepared by considering both the "*Implemented Curriculum*", i.e., what is actually taught, and how it is actually being taught and the "*Achieved Curriculum*" , i.e., what the students have actually learned as well as their attitudes toward math.	**10** *32.3%*	**17** *54.8%*	**2** *6.5%*	**2** *6.5%*	**0** *0.0%*	**31** *100.0%*
19	During the preparation of the Model Exam at the outset care should be taken to avoid the negative effect of the "*Hidden-Curriculum*" such as test frauds committed by individual or systematically organized group of students on the testing days.	**5** *16.1%*	**20** *64.5%*	**6** *19.4%*	**0** *0.0%*	**0** *0.0%*	**31** *100.0%*
20	Since the Grade 10 Math Model Exam is considered as a "Practice Test", I think it is not necessary to make alignment of items or questions with curriculum objectives.	**0** *0.0%*	**5** *16.1%*	**7** *22.6%*	**15** *48.4%*	**4** *12.9%*	**31** *100.0%*

As can be seen on this table there are five options which suggest possible role of curriculum (from conjecture without firm evidence) in the test development process of the Grade

10 Math Model Exam. According to the responses of the teachers it can be seen that, they highly recognized the role of curriculum in the preparation of the Math Model Exam. Moreover, as indicated on the Table of Item Statistics (Appendix: B) the highest mean value with a minimum standard deviation for Item 16 and the least mean value with highest standard deviation for Item 20, clearly shows the importance teachers place on the role of curriculum during the preparation of this Model Exam.

Furthermore there is an observable variation of perception among the respondents on the cognitive dimensions to be sampled by the Math Model Exam. In relation to this, Table 5 exhibits the number of respondents to the given ways in which the cognitive levels are sampled by the Math Model Exam.

Table 5: *Responses to Items Number 21- 25 of Part I*

No.	Item	SA	A	Und	D	SD	Total
21	I think, in math exam, items that ask for *Knowledge* or *Recall* of facts are easy, and hence such items should be given priority when writing items for Math Model Exam and the number of these items should be kept as many as possible.	**4** *12.9%*	**15** *48.4%*	**2** *6.5%*	**10** *32.3%*	**0** *0.0%*	**31** *100.0%*
22	I suggest that during the preparation of Math Model Exam questions that ask for the *Comprehension* part should be kept to the minimum.	**0** *0.0%*	**15** *48.4%*	**7** *22.6%*	**7** *22.6%*	**2** *6.5%*	**31** *100.0%*
23	Since it is a Math Exam, I generally believe that questions on the *Application* of Theorems, Postulates, Laws and Formulas are very important, and hence such kind of items must be larger in number and given more priority than the others	**9** *29.0%*	**12** *38.7%*	**2** *6.5%*	**8** *25.8%*	**0** *0.0%*	**31** *100.0%*
24	Items that ask for high order thinking ability (*Analysis, Synthesis* and *Evaluation*) are complicated for students and difficult to prepare, therefore it is better to avoid such kind of questions when writing items for Grade 10 Math Model Exam	**2** *6.5%*	**8** *25.8%*	**6** *19.4%*	**12** *38.7%*	**3** *9.7%*	**31** *100.0%*
25	I generally believe that, Grade 10 Math Model Exam is a "Practice Test" so when preparing this exam equal emphasis should be given to each cognitive level in proportion to the related learning outcomes intended to be measured.	**6** *19.4%*	**21** *67.7%*	**3** *9.7%*	**1** *3.2%*	**0** *0.0%*	**31** *100.0%*

As it is shown on the table the items were dealing with possible ways of sampling the behavioral dimensions that were supposed to be sampled by the Grade 10 Math Model Exam. This table of frequency distribution of responses indicates that for each of the items, except item 24, the response with the highest frequency is “Agree” the next higher one is that of “Disagree”, except items 24 and 25. This implies that there is a clear variation of perception among the teachers on the cognitive dimensions to be sampled by this Model Exam. The greater standard deviation (Appendix B) also asserts the existence of this variation.

Moreover there is a detectable difference of perception among teachers on the use of “Table of Specification” to prepare the Math Model Exam. The next table of frequency, the last for Part I, contains possible options on the probable uses of “Test Blueprint” (or “Item Specification”) during the preparation of Grade 10 Math Model Exam.

Table 6: *Responses to Items Number 26- 30 of Part I*

N<u>o</u>.	Item	SA	A	Und	D	SD	Total
26	I think *Test Blueprint* (or *Item Specification*) is used to prepare the Grade 10 Math Model Exam in many high schools.	**10** *32.3%*	**13** *41.9%*	**3** *9.7%*	**5** *16.1%*	**0** *0.0%*	**31** *100.0%*
27	I think you have heard of this document and its usage either as a student in the Teacher Training College/ University or as a teacher on a certain work-shop.	**5** *16.1%*	**15** *48.4%*	**9** *29.0%*	**2** *6.5%*	**0** *0.0%*	**31** *100.0%*
28	I believe *Item Specification* is not as such important to prepare Grade 10 Math Model Exam.	**0** *0.0%*	**3** *9.7%*	**3** *9.7%*	**21** *67.7%*	**4** *12.9%*	**31** *100.0%*
29	I think it is very difficult to prepare *Test Blueprint* every year, so it is not used to prepare Grade 10 Math Model Exam.	**0** *0.0%*	**7** *22.6%*	**1** *3.2%*	**17** *54.8%*	**6** *19.4%*	**31** *100.0%*
30	Even if teachers are well aware of its importance, I think they do not use it to prepare Grade 10 Math Model Exam in your school	**0** *0.0%*	**13** *41.9%*	**5** *16.1%*	**9** *29.0%*	**4** *12.9%*	**31** *100.0%*

As to the “Table of Specification” majority of the teachers had responded that they have some ideas about it (item 27). Hence 73.2% of them said, it is used to prepare the Grade 10 Math Model Exam in many high schools (item 26) and 41.9% of the teachers said that they use it, but equal percentage, i.e., 41.9% of them said they don’t use it to prepare this Model Exam in their own schools. This variation of responses for the respective items is also indicated by the relatively larger standard deviation on the table of Item Statistics (Appendix B).

The next table, for Part II, shows the number of teachers responded to items that dealt with the administration process of Grade 10 Math Model Exam.

Table 7: *Responses to Item Number 1- 4 of Part II*

No	Item	SA	A	Und	D	SD	Total
1	To create conducive environment for the administration of Math Model Exam in high schools, I suggest effort should be made in order to make special seating arrangement of the students. For example, **not** to allow more than two students to sit on one desk as in *GSECME.*	**19** *61.3%*	**11** *35.5%*	**0** *0.0%*	**1** *3.2%*	**0** *0.0%*	**31** *100.0%*
2	If students are found cheating or helping each other while they are taking the Model Exam and since this is unethical, I strongly suggest that disciplinary measures should be taken.	**21** *67.7%*	**10** *32.3%*	**0** *0.0%*	**0** *0.0%*	**0** *0.0%*	**31** *100.0%*
3	I think during the exam period, usually students complete Math Model Exam and leave the exam hall in less than 30minutes or $\frac{1}{5}$ of the total given time for the exam.	**3** *9.7%*	**11** *35.5%*	**4** *12.9%*	**8** *25.8%*	**5** *16.1%*	**31** *100.0%*
4	Since the Grade 10 Math Model Exam replicates or models the *GSECME*, I suggest special orientation should be given to students about the rules and regulations that they should follow during the exam period of Math Model as well as the purpose of this Model Exam.	**16** *51.6%*	**14** *45.2%*	**1** *3.2%*	**0** *0.0%*	**0** *0.0%*	**31** *100.0%*

It was found out that the respondents were well aware of the importance of creating conducive environment for the administration of Grade 10 Math Model Exam. As can be seen on this table of frequency of responses to items in Part II of the questionnaire, except for item 4, the response with the highest frequency is either “Strongly Agree” or “Agree”. This indicates that the respondents (Math Teachers) seem to have a strong belief in a proper administration process of the Model Exam and support the idea of taking disciplinary measures when test fraud is observed during the exam period.

This can also be seen from the higher mean values for the various features and principles of the administration process of educational testing mentioned in the four items of Part II. Surprisingly enough the highest mean values for item 2 of this part of the questionnaire reveals the fact that the teachers give more emphasis on the need for protecting and avoiding test frauds.

However regarding the scoring and reporting processes of students' result on the Grade 10 Math Model Exam some useful methods seem to be neglected. In relation to Part III of the questionnaire Table 9, which is shown below, indicates the number of teachers responded to items that suggested for different options on how the scoring process of the Grade 10 Math Model Exam had taken place. As this part was very crucial in the investigation, the items were modified so as to make each of them as explicit as possible in accordance with the feedback obtained during the piloting the questionnaire.

Table 8: *Responses about the Scoring Processes of the Grade 10 Math Model Exam.*

No	Item	SA	A	Und	D	SD	Total
1	I think a student's obtained score on Math Model Exam can be added to or combined with the student's scores on classroom performance so as to get the total score of the student in the semester.	**9** *29.0%*	**13** *41.9%*	**3** *9.7%*	**5** *16.1%*	**1** *3.2%*	**31** *100.0%*
2	I suggest that the score which a student obtained on Grade 10 Math Model Exam should be reported independent of any other score, because this helps to predict his/her efficiency on the *GSECME*.	**6** *19.4%*	**17** *54.8%*	**6** *19.4%*	**2** *6.5%*	**0** *0.0%*	**31** *100.0%*
3	After the correction of Math Model Exam, when assigning score and reporting the student's result, I think there is nothing wrong if the reported score is converted to 100. (For example, *out of 60 questions of Math Model Exam if a student correctly answers 20 questions, then the student's obtained score is always reported as* $\frac{33}{100}$)	**3** *9.7%*	**15** *48.4%*	**7** *22.6%*	**5** *16.1%*	**1** *3.2%*	**31** *100.0%*
4	Student's reported score on Grade 10 Math Model Exam truly reflects his/her mastery of the subject matter.	**2** *6.5%*	**15** *48.4%*	**10** *32.3%*	**4** *12.9%*	**0** *0.0%*	**31** *100.0%*
5	If the results of the previous several years are studied through detail analysis of school records, I think one can see that students who scored high on Grade 10 Math Model Exam also got high grades on the *GSECME*.	**7** *22.6%*	**16** *51.6%*	**6** *19.4%*	**2** *6.5%*	**0** *0.0%*	**31** *100.0%*
6	If the results of the previous several years are studied through detail analysis of school records, I think one can see that students who scored high on Grade 10 Math Model Exam also scored high on their University Entrance Math Examinations at the Preparatory levels.	**6** *19.4%*	**15** *48.4%*	**8** *25.8%*	**2** *6.5%*	**0** *0.0%*	**31** *100.0%*
7	As a matter of fact Grade 10 Math Model Exam is considered as "Practice Test" so the student's score should be taken as an indicator of the student's preparedness for EGSECME and hence it is better not to report it as achievement score.	**2** *6.5%*	**14** *45.2%*	**3** *9.7%*	**10** *32.3%*	**2** *6.5%*	**31** *100.0%*

Thus it can be seen from the nature of the items that if some of them are taken pair-wise, they convey conflicting ideas with respect to the principles of educational testing, for example item 1 and 6, and item 2 and 3. However the respondents almost equally agreed with these items, because the number of respondents as well as the mean values for all the items are much greater than the ideal mean 3 (see Appendix B). Therefore the difference between true score, obtained score, reported score and transformed score are found to be ignored or given no attention at all.

Moreover regarding the decision that has to be made based on the results of the model exam, there was an observable difference of perception among the respondents.

Table 9 below exhibits the frequency of responses to the given options, in Part IV, regarding to the ways in which academic decisions are made based on the scores of students on the Grade 10 Math Model Exam.

Table 9: *Responses about Academic Decision based on Grade 10 Math Model Exam Scores.*

No	Item	SA	A	Und	D	SD	Total
1	Due to any reason if a student does not take Math Model Exam, I strongly suggest that he/she should be considered as *incomplete* and hence should not be allowed to take the *GSECME*.	**7** *22.6%*	**7** *22.6%*	**4** *12.9%*	**6** *19.4%*	**7** *22.6%*	**31** *100.0%*
2	No matter how high or low a student scored on the Math Model Exam, I think the primary issue in taking the Model Exam is to make decision about that student (i.e., either he/she is allowed or denied to take the *GSECME*).	**1** *3.2%*	**7** *22.6%*	**6** *19.4%*	**15** *48.4%*	**2** *6.5%*	**31** *100.0%*
3	If a student fails in the Model Exam but gets a passing grade on the GSECE, then I suggest that this student should leave his/her present school but can continue his/her schooling in another school.	**2** *6.5%*	**8** *25.8%*	**8** *25.8%*	**11** *35.5%*	**2** *6.5%*	**31** *100.0%*
4	From my previous observations, students who had got a passing score on Grade 10 Math Model Exam in one school and proceeded their Preparatory education in another school, did **not** get passing scores either on Grade 12 Math Model Exam or University Entrance Math Examinations or on both.	**0** *0.0%*	**7** *22.6%*	**13** *41.9%*	**8** *25.8%*	**3** *9.7%*	**31** *100.0%*
5	It is necessary to design and implement a kind of system among all schools that enables them to make uniform, sound and informed academic decisions based on scores on the Model Exam	**10** *32.3%*	**17** *54.8%*	**3** *9.7%*	**1** *3.2%*	**0** *0.0%*	**31** *100.0%*

Both the table of frequency distribution of responses, above, and table of Item Statistics (Appendix B) revealed that there is a noticeable difference on the responses of the teachers to the items about the decision making process.

4.2 From the Interviews

4.2.1 Main Themes

When asked the first question on the availability of an official document that describes the purpose and function of the Model Exams, with full certainty Respondent #1 said:

> "*No, there is no such document and I never heard of it, but currently it is prepared at the regional education bureaus by consensus*"

The others also shared the same idea.

The interviewees have different views and opinion about the second question on the nature of the Model Exams. In this regard Respondent #1 said;

> "*The objective of this Model Exam is to help students, to let them practice and prepare themselves for the national exam so that they do not worry when they take that exam...*"

and after a few seconds pause he continued to say

> *"...the name 'Model Exam' is given to it because it looks like the national exam, that is to say a "Practice Test" which gives students an idea of what to expect on the national exams and will increase their confidence in taking these centrally prepared exams*".

Others said that it is an "Achievement Test" which is used to measure what the students achieve as the result their schooling. Respondent #3 who said "It is an Achievement Test" further argued and said;

When answering this question Respondent #3 said;

> *"...it is named after "Model Exam" because it models the national exam with respect to three aspects namely its content coverage, number of items and time allowed. Only modeling these three parts of the national exam cannot qualify our exam (i.e., the Model Exams) as a replica or model of the national exams. The name 'Model Exam' is too much and heavier to our exams when its nature and purpose are compared to that of the national exam which it is supposed to model. This is true not only for Grade 10 Model Exams but also for Grade 12 Model Exams. Both Exams do not deserve the name 'Model Exam.'"*

When the respondent was explaining this, he amazingly gestured by putting both hands over his shoulder and bending his head, to show as if he is carrying a heavier load on his shoulder, just to emphasize his argument and the degree of the difference between the Model Exams and the corresponding national Exams.

When answering questions that are related to the scoring processes and the consequences Respondent #3 said;

> *"...after school, when students apply for a job those who come from schools where the teachers/scorers are reluctant are more advantageous than students who come from those schools where the scorers are strict."*

In relation to the question about the preparation of the Model Exam all the respondents repeatedly said;

> *"...during the preparation of both the School Based Model Exam (i.e., before 2004 E.C.) and Regional Based Model Exam (after 2004 E.C.) there are many items with conceptual and editorial problems."*

In particular Respondent #2 said;

> *"...for the School Based Model Exam (i.e., before 2004 E.C.) there was a chance for correcting these problems even as the exam session went on*

but for the Regional Based Model Exam (i.e., after 2004 E.C) there is no such chance."

4.2.1 Findings from the interview

From the interview session with three principals and an official at the Woreda education bureaus the interviewees equivocally confirmed that, there is no official document that describes the purpose or function of the Model Exams as a whole. They said that these exams were begun a long time ago and even they themselves took it as a student of secondary school at the level of Grade 12 in the old curriculum. All the interviewees also said that in their current professional life they do not even see evidence or a document anywhere that mentioned the purpose of these Model Exams.

Through the interview it is also found that, still there is confusion about the nature of these Model Exam, because some of them said it is a "Practice Test" and the others said it is an "Achievement Test".

Moreover the interviewees revealed that scoring process is undertaken by teachers in the schools and the students' score are reported out 100% where 40% of which constitutes the students' classroom performance while 60% of the score is the students' obtained score on the Model Exam. However, there is no consistency on the reported scores. In some schools, where the teachers /scorers are reluctant, the students' reported scores are very inflated or exaggerated but in other schools where the scorers are relatively strict, the students' reported scores are reasonably normal. Consequently problems have been observed when students, who were unable to pursue their education in higher institutions, are joining the labour market.

Furthermore the interviewees confirmed that since 2004E.C to date, the Oromia Regional Education Bureau has taken the responsibility of preparing the Model Exams for all secondary schools in the region while the administration, the marking and the scoring processes are left for the respective schools.

4.3 Results from Experts' Evaluation of Grade 10 Math Model Exams Booklets

After critical evaluation of Grade 10 Math Model Exams Booklets that were collected from the sampled schools and which were administered at the end of the 2003E.C academic year, the findings of the subject matter experts are summarized and listed as follows. For each question the numbers under "Yes" or "No" columns represent the number of experts whose answers to the corresponding question was Yes or No respectively.

Table 10: *Evaluation Results about Adequacy of the Assessment Plan*

Q. No		***Yes***	***No***
1	Is the preparation of this Grade 10 Math Model Exam adequately planned?	0	3
2	During the preparation of this Math Model Exam, is appropriate emphasis given to the objectives and each content area to be measured at this grade level?	1	2

As shown on the table all experts found that the Grade 10 Math Model Exam was not adequately planed. On their remark (that is consolidated and presented on Appendix G) they said that the Model Exam was not adequately planed because they found most of the questions give emphasis to measure specific learning outcome (i.e., application) and there are some questions which measure the same learning out come from the same content area. As to them the preparation process of this Model Exam ignored the correct proportion of specific learning outcomes to be sampled.

Table 12, which is given below, indicates the number of experts in accordance with their answer to the corresponding question related to the adequacy of items and the corresponding tasks. The experts' evaluation results were also supported by their remark given against each of the question is given on Appendix G.

Table 11: *Evaluation Results about Adequacy of Test Items and the Corresponding Tasks*

Q. No		***Yes***	***No***
3	Is the format of the item and task suitable for the learning outcome being measured ***(appropriateness)***?	0	3
4	Does each item or task require students to demonstrate the performance described in the specific learning outcome it measures **(*relevance*)**?	1	2
5	Does each item or task present a clear and definite task to be performed ***(clarity)***	0	3
6	Is each item or task presented in simple readable language and free fron excessive verbiages ***(conciseness)***?	0	3
7	Does each item or task provide an appropriate challenge ***(ideal difficulty)***?	2	1
8	Does each item or task have an answer that would be agreed upon by different mathematics experts ***(correctness)***?	1	2
9	Is there a clear basis for awarding partial credit on items or tasks with multiple points ***(scoring rubric)?***	1	2
10	Is each item or task free from technical errors and irrelevant clues ***(technical soundness)***?	1	2
11	Is each item free from ethnic, religious and gender bias ***(cultural fairness)***?	3	0
12	Is each test item independent of the other item in the test ***(independence)***?	2	1
13	Is there an adequate number of test item for each learning outcome ***(sample adequacy)***	0	3

As indicated by Item 3, 5, 6, and 13 above the experts said that the Model Exam preparation processes lacks *appropriateness, clarity, conciseness,* and *sample adequacy* whereas the items are found free from ethnic, religious and gender biases, Item 11.

Table 12: *Evaluation Results about the Adequacy of Test Format and Editorial Processes*

Q. No		***Yes***	***No***
14	Are test items of the same type grouped together in the test (or within sections of the test)?	3	0
15	Are the test items arranged from easy to more difficult within Sections of the test and the test as a whole?	2	1
16	Are the test items numbered in sequence?	3	0
17	Are the correct answer distributed in such a way that there is no detectable pattern?	1	2
18	Is the test material well spaced, legible and free of typographical errors?	1	2
19	Are there directions for each section of the test and the test as a whole?	2	1
20	Are the directions clear and concise?	1	2

The experts also found that items of the same type were grouped together (or within sections of the test) in the Model Exams as well as the items in the exam were numbered in sequence Item 14 and Item 16.

Table 13: *Evaluation Results about the Math Model Exam in General*

Q. No		***Yes***	***No***
21	Can this Model Exam be considered as “Practice Tests” which gives students an idea of what to expect on the EGSECE and increases their confidence in taking the national exam by practicing on the items of this Model Exam?	1	2
22	Does this Model Exams really serve as an “Achievement Test”, i.e., used to measure what the students achieve as the result their schooling?	1	2
23	Does this Model Exams really model or copy the national EGSECE?	0	3

The experts could not classify the Math Model Exam either as “Practice Tests” or “Achievement Test” but they agree that it does not model or copy the national EGSEC Math Exam.

4.4 Relationship between Variables (results from Statistical Analysis)

Scores of 230 students who took all the four exams, namely the 2003 E.C. Grade 10 Math Model Exam and EGSECE as well as the 2005 E.C. Grade 12 Math Model Exam and EUEE, were collected and analysed.

4.4.1 The Correlation Matrix

After the scores were collected and before checking whether the Criterion-related validity was maintained, it was necessary to examine whether the assumptions of Correlation and Linear Regression Analyses were fulfilled. Consequently when the score (or variables) were studied pair-wise it was found out that the Pearson correlation coefficients for all pairs of scores were positive and were significant at the 0.01 level (1-tailed), N = 231. Moreover, some of these coefficients indicate the existence of a strong correlation while the others indicate weak correlation. The situation is demonstrated on Table 15 given below.

Table 14: *Correlation Matrix of Scores (Variables)*

Correlations

		EUE Math Exam for 2005 E.C	Grade 10 Math Mode Exam for 2003 E.C	EGSEC Math Exam for 2003 E.C	Grade 12 Math Model Exam for 2005 E.C
Pearson Correlation	EUE Math Exam for 2005 E.C	1.000	$.350^{**}$	$.266^{**}$	$.300^{**}$
	Grade 10 Math Mode Exam for 2003 E.C	.	1.000	$.761^{**}$	$.593^{**}$
	EGSEC Math Exam for 2003 E.C	.	.	1.000	$.484^{**}$
	Grade 12 Math Model Exam for 2005 E.C	.	.	.	1.000
Sig. (1-tailed)	EUE Math Exam for 2005 E.C	.	.000	.000	.000
	Grade 10 Math Mode Exam for 2003 E.C	.	.	.000	.000
	EGSEC Math Exam for 2003 E.C	.		.	.000

Thus there was a linear relationship between the variables under consideration. This relationship is also depicted by the scatter plot matrix shown on Appendix D.

As the main objective of this study was to gather and anlalyse validity evidences of the 2003 E.C Grade 10 Math Model Exam from selected secondary schools in Oromia, one of these validity evidences under investigation was the Criterion-related validity evidence of the exam. Hence the two types of Criterion-related validity, namely, the *concurrent validity* and the *predictive validity* can be performed through the correlation and regression analyses accordingly.

4.4.2 Finding on the Concurrent Validity

The analysis on the concurrent validity evidences of the Grade 10 Math Model Exam in relation to the corresponding centrally prepared EGSEC Math Exam, since both exams were given in 2003 E.C, was done through close examination of the strength of the relationship explained by the Pearson correlation coefficient. The analysis was performed by considering the Grade 10 Math Model Exam as an estimator variable and the corresponding centrally prepared of EGSEC Math Exam as the criterion variable. Therefore, the obtained Pearson correlation coefficient (r = .761) indicated that there is a strong positive relationship between the variables. This implies that the Grade 10 Math Model Exam which was prepared and administer for the 2003 E.C in the selected secondary schools was a very good indicator of the current performance of the students on the corresponding EGSEC Math Exam of 2003 E.C.

In addition to this it was also found that the sample mean scores and the corresponding standard deviations were given on Table 15 below.

Table 15: *Descriptive statistics of the Math Model Exam and EGSEC Math Exam (for2003 E.C)*

Descriptive Statistics			
	Mean	Std. Deviation	N
EGSEC Math Exam for 2003 E.C	26.50	8.372	230
Grade 10 Math Model Exam for 2003 E.C	61.85	12.326	230

Furthermore, the ANOVA table which is given below indicates the statistical significance at which the performance of students on the EGSEC Math Exam for 2003 E. C was predicted from the students' performance on the 2003 E. C Grade 10 Math Model Exam.

Table 16: *Output for ANOVA.*

ANOVA[b]					
	Sum of Squares	df.	Mean Square	F	Sig.
Regression	9287.239	1	9287.239	313.041	.000[a]
Residual	6764.261	228	29.668		
Total	16051.500	229			

a. Independent Variable: Grade 10 Math Model Exam for 2003 E.C

b. Dependent Variable: EGSEC Math Exam for 2003 E.C

From the values indicated on Table 16 above it was found that the Grade 10 Math Model Exam was accounted for 57.9% (r^2 = .579, F (1, 228) = 313.04, p = .000) of the variation in the scores of EGSEC Math Exam. Hence the Grade 10 Math Model Exam for 2003 E.C predicted a significant proportion of the score variation on EGSECE Math for the same year.

Moreover as to the second type of Criterion-related validity, namely *predictive validity* of Grade 10 Math Model Exam, the analysis is done through multiple linear regression analysis. At the outset this analysis employed hierarchical method so as to get three models and to see the contribution of Grade 10 Math Model Exam in predicting EUE Math Exam. Thus in Model 1, the first stage in the hierarchy, only the Grade 10 Math Model Exams is used as a predictor to estimate EUE Math Exam. In Model 2 both Grade 10 Math Model Exams and the EGSECE Mathematics Exam are used as the predictor variables to predict EUE Math Exam. In Model 3 the scores on all three exams (i.e., Grade 10 and Grade 12 Math Model Exams and the EGSEC

Mathematics Exams) are used to estimate scores on EUE Math Exam. Consequently the outputs of this multiple linear regression analysis are shown on the tables below.

Table 17: *Model Summary for the Predictive Validity of Grade 10 Math Model Exam (2003 E.C)*

Model Summary[d]										
					Change Statistics					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change	Durbin-Watson
1	.350[a]	.123	.119	10.452	.123	31.843	1	228	.000	
2	.350[b]	.123	.115	10.475	.000	.000	1	227	.999	

a. Predictors: (Constant), Grade 10 Math Model Exam for 2003 E.C

b. Predictors: (Constant), Grade 10 Math Model Exam for 2003 E.C, EGSEC Math Exam for 2003 E.C

c. Dependent Variable: EUE Math Exam for 2005 E.C

The regression correlation coefficient (R_1 = .350) was the same as the Pearson correlation coefficient, as there was one predictor variable in the linear regression Model-1. Moreover, 12.3% of score variation (R_1^2 = .123, F (1, 228) = 31.843, p = .000) in EUE Math Exam score were predictable from scores variation in the Grade 10 Math Model Exam. However, when one predictor variable, i.e., EGSEC Math Exam for 2003 E.C, is included in this model so as to get Multiple Linear Regression Model-2; the multiple correlation coefficient remained the same, i.e., R_2 =.350, consequently the coefficient of determination was unchanged too (i.e., R_2^2 = .123). Moreover, under the “Adjusted R Square” column the value decreased from .119 in Model-1 to .115 in Mode- 2 instead of increase. This happened due to the inclusion of EGSEC Math Exam to Grade 10 Math Model Exam so as to get two predictor variables in Mode- 2. This tells us that the inclusion of EGSEC Math Exam as a predictor variable (with Grade 10 Math Model Exam) had no contribution in predicting the proportion of variation of scores on EUE Math Exam. This was also confirmed by the absence of change in R Square, that is R Square Change = .000 (see on the table under column 8 for Model - 2). In spite of indicated the presence of a special case; this situation was of considerable interest. From statistical point of view the occurrence of this situation indicated that the information about EUE Math Exam carried by EGSEC Math Exam was completely redundant with the information that was carried by Grade 10 Math Model Exam. This was also confirmed by the result explained in sub-section 4.4.2 above, which demonstrated the presence of a strong positive linear relationship between these two variables (r = .761 even that is significant at the 0.01 level for 1-tailed). Hence including two predictor variables in the regression analysis that were strongly collinear was not sensible, considering that once one of the variables was included, the inclusion of the second variable would offer no information that was not already provided by the first.

Thus, the option left for the next possible multiple linear regression model, was given as shown below.

Table 18: *Model Summary of Multiple Linear Regression Analysis*

Model Summary[c]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics: R Square Change	F Change	df1	df2	Sig. F Change	Durbin-Watson
1	.350[a]	.123	.119	10.452	.123	31.843	1	228	.000	
2	.368[b]	.136	.128	10.396	.013	3.465	1	227	.064	1.656

a. Predictors: (Constant), Grade 10 Math Model Exam for 2003 E.C

b. Predictors: (Constant), Grade 10 Math Model Exam for 2003 E.C and Grade 12 Math Model Exam for 2005 E.C

c. Dependent Variable: EUE Math Exam for 2005 E.C

Likewise result of the analysis of score variance, in the predictor variables along with the criterion variable in each of the multiple linear regression models, was summarized on the following ANOVA table.

Table 19: *Analysis Of Variance (ANOVA) For The Predictive Validation*

ANOVA[c]						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3478.558	1	3478.558	31.843	.000[a]
	Residual	24906.589	228	109.239		
	Total	28385.148	229			
2	Regression	3853.014	2	1926.507	17.826	.000[b]
	Residual	24532.134	227	108.071		
	Total	28385.148	229			

a. Predictors: (Constant), Grade 10 Math Model Exam for 2003 E.C

b. Predictors: (Constant), Grade 10 Math Model Exam for 2003 E.C, and Grade 12 Math Model Exam for 2005 E.C

c. Dependent Variable: EUE Math Exam for 2005 E.C

In the Model-1 linear regression 12.3% of score variation in Math EUE Math Exam scores ($R^2_{\mathbf{1}}$ = .123, F (1, 228) = 31.843, p = .000) was predicted by the score variation in Grade 10 Math Model Exam. Hence the 2003 E.C Grade 10 Math Model Exam predicted a significant proportion of variation in the scores on EUE Math Exam of the year 2005 E.C.

However when one predictor, i.e., Grade 12 Math Model Exam for 2005 E.C, was included in Model-1 so as to get the second model, i.e., Model-2, for multiple linear regression analysis. As it was expected the regression correlation coefficient (R_1 = .350 for Model-1) increased and became R_2 = .368 for Model-2.

Moreover, it was found that there was an incremental change in the proportion of score variation due to the inclusion of the second predictor variable (Grade 12 Math Model Exam) to get Model-2. Therefore, 13.6% of variation (R^2_2 = .136, F (1, 227) = 17.826, p = .000) in EUE Math Exam score was accounted for the score variations in both predictor variables. Therefore the 2003 E.C Grade 10 Math Model Exam together with the 2005 E.C Grade 12 Math Model Exam predict a significant proportion of scores variation in EUE Math Exam given in the year 2005 E.C.

The multiple linear regression coefficients for the two models are given on the following table.

Table 20: *The Multiple Linear Regression Coefficients*

Model		Unstandardized Coefficients		Standardized Coefficients			Correlations		
		B	Std. Error	Beta	T	Sig.	Zero-orde	Partial	Part
1	(Constant)	21.905	3.533		6.199	.000			
	Grade 10 Math Model Exam For 2003 E.C	.316	.056	.350	5.643	.000	.350	.350	.350
2	(Constant)	19.124	3.819		5.008	.000			
	Grade 10 Math Model Exam For 2003 E.C	.240	.069	.266	3.467	.001	.350	.224	.214
	Grade 12 Math Model Exam For 2005 E.C	.128	.069	.143	1.861	.064	.300	.123	.115

Therefore, in Model-1while testing the significance of the regression coefficient (B = .316), the computed values were β = .350, *t* (228), *p* = .000 < 0.05. Hence, the slope was indeed significantly different from zero. Consequently the 95% CI (B) was (0.206, 0.427). As this interval does not contain zero, the corresponding population parameter was significantly different from zero. Therefore this test asserts that the 2003 E.C Grade 10 Math Model Exams had significant contribution in predicting the scores on EUE Math Exam given for 2005 E.C academic year.

In Model-2 the first partial slope of the regression line was related to the first predictor variable, i.e., Grade 10 Math Model Exam. This partial slope was significantly different from zero, (β_1 = .266, *t* (227) = 3.467, *p* =.001) and the 95% CI for β_1 is (0.1048, 0.3752). Since this interval does not contain zero, the corresponding population parameter was significantly different from zero. However, for the second partial slope of the regression line i.e., related to the Grade 12 Math Model Exams scores, (β_2 = .128, *t* (227) =1.861, *p* = .064) the *p* value exceeded α = .05. It was also found that the 95% confidence interval of this partial slope happened to be (-.008, .264) which certainly contains zero. This implies that the corresponding population parameter was not indeed significantly different from zero. However the 90% CI for β_2 is (.0145, .2415) which of course does not contain zero and hence the corresponding population parameter was indeed significantly different from zero at α = .10 significance level.

Finally, based on the results presented on the above tables, when checking whether the assumptions for Multiple Linear Regression were satisfied, it was found that:

a) the variables were continuous and measured at the interval level, **b)** the predictors had apparently non-zero variance as shown on Table 15, **c)** the variance inflation factor (VIF) and

Tolerance values confirmed that multicollinearity was not as such a serious problem that resulted in biasing the regression of the EUE Math Exam scores (criterion variable) on the predictors described by the model. **d)** the standard error of the estimates for each model were almost the same and this implies that at each level of the predictor variables the residuals had the same variance, i.e., there existed homoscedasticity, **e)** the scatter plot of the standardized residuals vs. the standardized predicted values showed the points were randomly scattered and confined between -2 and 2 and more concentrated around the mean which is 0 and this indicated that the distribution of the residuals was normal Appendix E, **f)** for any two observations the residuals were independent or uncorrelated and the this confirmed by the value of Durbin-Watson test statistic, i.e., 1.657 which is closer to 2 as shown on Table 7. Collectively all these statistical results implied that the assumptions for multiple linear regression were fulfilled for the chosen model.

CHAPTER 5
DISCUSSION

In this part of the study the findings are discuss in accordance with the order of the research questions that were posed at the beginning of the study.

Primarily the first crucial research question that the study investigated was the availability and whereabouts of any clearly stated official document (either in the schools or in any concerned bureaus or institution) that describes the intended purpose/s or main function of preparing and administrating Grade 10 Model Exams. In relation to this, the findings showed there is no such document. Thus it seems that these Model Exams are exams without clearly intended purpose and uniform implementation procedures. Moreover, as can be seen from the results of the questionnaire, the overall perception of the teachers is that the contents of the Grade 10 Math Model Exams should be similar to that of EGSEC Math Exam and they considered this Model Exam as an “Achievement Test”. Even if these teachers need the Math Model Exam to model EGSEC Math Exam, the subject matter experts (who evaluated the Grade 10 Math Model Exam booklets) were agreed and confirmed that, in terms of the major aspects of content validity evidences; this Model Exam does not model or copy the national EGSEC Math Exam. However, the classification of the Model exam as an “Achievement Test” is supported by these experts too. On the other hand the findings from both the questionnaire and the interview revealed that there is also a second task that this Model Exam has to serve for, i.e., to serve as a “Practice Test”. However, without having any concrete evidence about these very important components of educational tests (i.e., intended purpose and uniform implementation procedures), this study was proceeded so as to investigate at least some common evidences that are shared by schools and support the validity of Grade 10 Math Model Exam.

The second and the third research questions are logically related and dealt with the content validity evidences. Therefore, integrating and discussing the corresponding findings and their implications may give a better picture of the issue of maintaining content validity of the Grade 10 Math Model Exam during the preparation processes. The findings imply that, due to the long standing tradition of administering these Model Exams is so deep rooted in our schools. As expected it is found that the teachers have shared almost the same idea (through their responses to the questionnaire) that "the Grade 10 Math Model Exam should cover contents of Grade 9 and Grade 10". This means they need the contents of the Model Exam to be similar to that of the nationally prepared EGSEC Math Exam. However when it comes to the method of determining number of items to be selected from the respective contents, there is an observable difference in their opinion. Some of them need the number of items to be determined based on the depth or breadth of contents in each chapter because students can understand the concepts and become confident on EGSEC Math Exam. The others suggested that the number of items has to be determined based on the number of allotted period or else on table of specification. Moreover, a significant proportion of them agreed to do this by consulting or reviewing the EGSECE booklets of the previous years. In any ways, either directly or indirectly, the value that they give to maintain the similarity between the Model Exams and EGSECE is very high.

Furthermore, the findings from the questionnaire showed that, the teachers highly recognized the role of curriculum in the preparation of Grade 10 Math Model Exam, though there was an observable difference on the emphasis they give to the cognitive levels (or behavioural dimensions) to be sampled by the Math Model Exam. Also it should be noted that there is a tendency of agreement among the teachers on two different activities. From practical point of view this situation puts the Math Model Exam preparation process in question. For instance, 61% of the teachers “Strongly Agree” to “*I generally believe that items of Grade 10 Math Model Exams should be prepared from both the contents of grades 9 & 10 based on a table of specifications*” while 42% of them said that they “Agree” to “*Even if teachers are well aware of its importance, I think they do not use it to prepare Grade 10 Math Model Exam in my school*”. Therefore this implies that the important role that the table of specifications plays in the preparation of the Model Exams is recognized by the teachers, but its implementation is ignored.

Moreover the subject matter experts’ evaluation result pointed out some of the factors that affect the content validity of Grade 10 Math Model Exam such as, lack of plan in the preparation, inadequacy of some of the items and the corresponding tasks, inadequate number of items that represent important learning outcomes. All that has been discussed so far indicate that the Math Model Exams preparation process has suffered from lack of important factors that lessen its content validity. As a matter of fact in this study, the essence of the assessing the content validity of Grade 10 Math Model Exam, then, is in evaluating the adequacy of the sampling of the contents based on which the scores on this Model Exam are interpreted. More formally, the goal in the content validation of Grade 10 Math Model Exam is to determine the extent to which a set of tasks or items of the exam provides a relevant and representative of the domain of tasks based on which interpretations of scores on the Model Exam are made. Therefore, inadequate representativeness of content coverage can play in undermining both the validity and reliability of content-referenced test scores.

The results obtained from the interview make possible to assess the extent to which the reported scores of Grade 10 Math Model Exam in Oromia enabled the test users to make accurate decision, an issue that was posed by the fourth research question. According to these results the scoring process of the Model Exams has taken place in the respective schools. Students' scores are reported after combining their classroom performance with the obtained scores on the Model Exams. Because performances are evaluated by the classroom teachers and these evaluation results are highly prone to subjectivity, the reported scores are certainly contaminated by this subjective judgment. This by its turn creates a problem on the independent analysis of students' scores on the Model Exam. As the result of this the reliability of the reported scores as well as the scores on the Model Exam becomes questionable.

In addition to this, when evaluating the Grade 10 Math Model Exams in terms of their value implications which refers to individual and social consequences of test use (known as *consequential validity* or *Utility of Testing*), the results of the interview pointed out an interesting and exemplary situation. This situation (i.e., when students join the labour market) emanated from the scoring process of the Model Exam and can be detected when the scores are reported or else when decision is made based on these reported scores. As indicated in the result of the interview this decision either positively or negatively affects the students' future life, even after schooling. Since "Utility" is dependent on the extent to which the use of test scores can increase the rate of accuracy of the inferences and decisions we wish to make, once again the Grade 10 Math Model Exams suffered from low consequential validity which refers to the benefits that scores on the Model Exams bring to decision making.

Concerning the criterion related validity of Grade 10 Math Model Exam supported by logically related validity evidences, the statistical results were obtained through the correlation and multiple linear regression analysis respectively. Thus, with respect to the fifth research question these results and their implications are discussed accordingly.

As indicated, the concurrent validity of the school based Grade 10 Math Model Exam is high where the criterion measure is the nationally prepared EGSEC Math Exam. In other words, the 2003 E.C Grade 10 Math Model Exam prepared at school level in Bishoftu town of Oromia Region is useful as a predictor of the then/current status of the students or as a substitute (concurrent) measure of the national EGSEC Math Exam of the same year. Statistically the Grade 10 Math Model Exam has higher concurrent validity, but from the practical point of views this does not always mean it is more suitable for a substitute (or concurrent) measure of the nationally prepared EGSEC Math Exam. This is because as discussed above the Model Exam under consideration suffers from low content validity. For instance the scoring process is highly influenced by subjectivity which results in high reported scores. This can be seen easily by comparing the sample mean sores of the Model Exam (M = 61.85, SD = 12.33) and that of EGSEC Math Exam (M = 26.50, SD = 8.37).

Furthermore, the findings of this study that are related to the predictive validity were obtained through multiple linear regression analysis. As a matter of fact the variables were selected for consideration in the analysis because they are logically interrelated and have theoretical as well as practical importance, but not because they are available. Thus the study undertook a variety of approaches to improve the fit of the chosen models to the available data so as to enhance the interpretability of the results, because some aspects of the regression results were not consistent with expectation.

Therefore, it is found out that there is a tendency for those students who have had higher score on Grade 10 Math Model Exam to get higher score on EGSEC Math Exam than those who have had lower score on the Model Exam (Pearson Correlation coefficient, r = .761 which is significant at the 0.01 level, for 2-tailed test). This means the scores on these exams are somewhat strongly correlated. Thus, taking these two variables (Grade10 Model Math Exam and EGSEC Math Exam) as predictors means to some extent redundant, and necessarily their respective estimates of the dependent variable (or predicted score on EUE Math Exam) will also be redundant. In other words what is need to be explained by EGSEC Math Exam for 2003 E.C, is already explained by the 2003 E.C Grade10 Model Math Exam administered in Bishoftu town. This is described by the fact that the sample regression coefficient and both the sample partial correlation coefficients (i.e., B = -.015, sp = -.007, pr = -.008) are approximately zero for EGSEC Math Exam in the regression Model (in which these exams were taken together as predictor variables).

Ultimately the required result, to estimate the scores on EUE Math Exam optimally from both these predictors, is an equation in which their redundancy (or more generally the strong relationship between Gr.10 Model Math Exam and EGSEC Math Exam) is taken into account. As the result of this, it was found necessary to exclude one of these two redundant variables so as to get the next possible multiple linear regression model that best describes the data. In particular, the researcher sought additional high school data that would enhance the validity of the Grade10 Model Math Exam. Unfortunately, there was no access to other additional data. The only option was a model in which once again two variables are used to predict the score on the EUE Math Exam. In this case Grade 10 Model Math Exam and Grade 12 Model Math Exam are chosen as predictor variables, because the aim of this study is to gather and analyse validity evidences that enable us to establish a logical argument about the validity of the Gr.10 Model Math Exam.

However the limitation of this study is that, the degree to which these findings are generalizable is unknown because the sample size and area of study was so restricted due to budget and time constraints. Furthermore, the lack of the reliability analysis of these Math Model Exams is the second limitation of the study. Future work in this area, apart from taking large sample size and including samples from different regional states, should focus on the development of an improved high school index, ideally by including some of such variables as reliability indices of the Model Exams, teachers' attitude towards employing different types of classroom performances, students' attitude towards exam, level of administration involvement in testing programs, level of parents follow up of their kids' performance at school, and the likes.

CHAPTER 6

SUMMARY, CONCLUSION AND RECOMMENDATION

6.1 Summary

The overall purpose of this study was to identify, gather and analyze the validity evidences of Grade 10 Math Model Exam so as to establish a logical argument whether some aspects of validity, such as content validity as well as criterion related validity, were maintained. It focused on the study of validity of this Model Exam with respect to educational test development and measurement principles particularly prepared and administered in the selected schools from the Oromia Regional State. To meet its objectives and answer specific research questions, 31 secondary school mathematics teachers, 3 secondary school principals, an official from City Education Bureaus as well as 3 mathematics subject experts were directly involved in the study.

The results obtained from the interview revealed that there was no official document that describes the purpose or function of the Model Exams in our education system as a whole. As the result of this, there was no clear understanding about the purpose and nature of the Model Exam. Some of the respondents said "It is a Practice Test" which gives students an idea of what to expect on the national exams and will increase their confidence in taking these exam. Others said "It is an Achievement Test" which is used to measure what the students achieve as the result their schooling. The name "Model" is given to this exam because it copies some aspects of the national exam, such as the grade levels from which content areas are selected and sampled by items of the exam, the number of items in the exam and the given time for completing the exam.

As the result of this the whole processes, from preparation to decision making, are highly dominated by unaltered and traditional practices of examination that has been fixed through being widely held belief, and hence prone to subjectivity and lack of accountability in the testing program. All these and related situations, unless corrected, will seriously impair the validity and reliability of the Model Exam under investigation. Furthermore the results obtained from the evaluation of the subject matter experts as well as the statistical findings satisfactorily provided answers to the research questions posed by this study. In short the findings of the study revealed the weakness of the content and criterion related validity of the Grade 10 Math Model Exams.

Moreover in answering the research question related to the scoring process the results obtained from interview session provided sufficient answer by revealing the existence of a potential problem. This problem is the inconsistency of the scoring process of the Model Exam which also has negative consequences. More generally the available evidences and all the methods employed by this study revealed that the current practice of preparing and administering this Model Exam is not compatible with the contemporary concept of validity, that is, the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests.

6.2 conclusion

Based on the summary of the findings explained above, the researcher draws the following conclusions, and ***their corresponding implications***:

1. The results of the present study showed that the preparation and administration of these exams were begun in our secondary schools a long-long period of time ago that no one knows and continue to be given to date. Moreover, there is no any clearly stated official

document that describes the intended purpose/s of preparing and administrating these Grade 10 Model Exams in our education system.

This implies that, these Model Exams that have no clearly stated or known purpose may become sources of unintended outcomes or unhealthy practices in the educational assessment system in the country. Moreover, this is the very first reason why discrepancies have been observed in our schools in the preparation, administration, scoring and decision making processes based on scores of these Model Exams. For instance, Why Grade 9 final exams do not include contents from Grade 8? Why Grade 11 final exams do not include contents from Grade 10? Why only this Grade 10 Model Math Exam includes contents from previous grade level i.e., contents from Grade 9? For that matter Grade 12 Model exam also includes contents from Grade 11. But why? Is it because students who completed Grade 10 and Grade 12 suppose to take the nationally prepared high stake EGSECE and EUEE? Logically convincing replies should be given to these questions.

2. The finding also indicated that the Grade 10 Math Model Exam suffered from lower content validity. It also revealed that teachers, school administrators, even officials at different education bureaus level need this exam to measure students' achievement at the end of each academic year, they also need it to be a practice test.

This implies that the Model Exam is required to serve two quite different purposes, i.e., to measure achievement and to be used for practice simultaneously. But scholars in the field like, Cureton (cited in Shepard, 1997) argued that, the essential question of test validity is how well a test does the job it is employed to do. The same test may be used for several different purposes, and its validity may be high for one, moderate for another and low for a third (1997: 5).

Moreover, according to the AERA (as cited in Haladyna, 2004) "*validity*" refers to the degree to which evidence support the interpretations of test scores entailed by proposed uses of the tests (2004: 9). Therefore, when we examine the reported scores on the Model Exam with respect to the above explanations, they may lack credibility and lead test users to make wrong conclusion or decision about whether the students have acquired the necessary knowledge, skill or attitude. Even this decision may affect the students' future life after school.

3. It was found out that there is inconsistency in the predictive validity of the Math Model Exam, in short it was found out that what is need to be explained by EGSEC Math Exam is already explained by Grade10 Model Math Exam.

The statistical analysis revealed that the 2003 E.C Grade 10 Math Model Exam and the corresponding nationally prepared EGSEC Math Exams were redundant in the analysis of the predictive validity evidences. Furthermore, the statistical findings indicated the presence of a strong positive correlation of scores between these two exams. One who may closely see the results of the interview also expect this to happen, because, as the results pointed out, the scoring process is done by school teachers and hence the reported scores highly prone to subjectivity. This subjectivity may results in high inflation of the reported scores. Actually this is of great interest and a call for further study on the situation. Furthermore, this can be considered as an indication of a situation that is taking place in the schools. This situation is that, as EGSECE is a high stake, the school administration and the teachers are working together to the level of their effort so as to make the Model Exams a practice test, to familiarize students with test taking skills, even to enforce teachers to "teach-to-the test" so that students will get a passing mark on this centrally prepared EGSECE.

6.3 Recommendation

Based on the conclusions explained above, the researcher forwards the following suggestions. Moreover the suggestions presented here are generally based on the study and meant to be informative rather than authoritative.

- There should be a research based and well organized policy document or standard that exhaustively describes the purpose of every exam at each grade level of our school system. In this regard the MoE, NEAEA and AAU should be the pioneer.
- Teachers should be given continuous training on modern methods of assessment that enable them to be efficient and effective in developing valid and reliable tests and implementing different kinds of continuous classroom assessment.
- Teachers, school administrators and officials at different levels should facilitate the implementation of different and effective method of teachings that enable students to acquire the necessary knowledge and skill. Only then students are able to score passing mark on the national exams and will be successful in their further education and future endeavour after school, but not by “Teaching-to-the Test” or by “Coaching” that narrows the curriculum or by teaching test taking skills through different techniques or tools.
- MoE, NEAEA, Regional Education bureaus and Universities in the country should work in collaboration to enhance the quality of the educational assessment and examination in our education system.

References

Amrein, A. L., & Berliner, D.C. (2002). *An Analysis of Some Unintended and Negative Consequences of High-Stakes Testing*, Great Lakes Center for Education Research & Practice Retrieved from: http://www.greatlakescenter.org

Anastasi, A., &Urbina, S. (2003*). Psychological Testing*

India, New Delhi: Prentice /Hall

Cohen, J. Cohen, P. West, S. G. & Aiken, L. S. (2003) *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences (3rd. ed.)* New Jersey, London: Lawrence / Mahwah.

Dame Abera (2013). *Introduction to research and research designs [Class handout]*.

School of Psychology, AAU, Addis Ababa

Desalegn Chalchisa (2012). *Standardized Achievement Tests [Class handout]*

Institute of Educational Research, AAU, Addis Ababa

Domino, G., & Domino, M. (2006).*Psychological Testing: An Introduction*

Cambridge, UK: The Edinburgh Building / Cambridge University Press,.

Ethiopia. MoE. (1994). *Education and Training Policy* (1st ed.). Ethiopia. Addis Ababa:

ST. George Printing Press.

Freeman, F.S. (2006). *Theory and Practice of Psychological Testing*

Kolhapur Road, Delhi: Surjeet Publications, Kamla Nagar

Gronlund, N. E. (1981). *Measurement and Evaluation in Teaching*

Third Avenue, New York: Macmillan, Inc.

Guilfordm, J. P. (2008). *Psychometric Methods*

Surjeet Publications, Delhi: Kolhapur Road.

Hailu Sime. (1990). *Teacher's Estimation of Test Item Difficulty as an Indicator of Their Prior Knowledge of Students' Academic Performance*

Unpublished, MA Thesis, Addis Ababa University

Kubiszyn, T. & Borich, G. (2004). *Educational Testing and Measurement: Classroom Application and Practice.*

New York Third Avenue, NY: John Wiley & Sons, Inc.

Kebede Adera. (1991). *The Prediction of the Academic Achievement of Students from High School Grade Point Average, ESLCE and Selected Aptitude Tests in Junior College of Commerce.* Unpublished, MA Thesis, Addis Ababa University

Legesse Getachew. (2006). *Assessing the predictive validity of the Ethiopian higher education entrance certificate examination*. Unpublished, MA Thesis, Addis Ababa University.

Looney, J. W. (2011), "Integrating Formative and Summative Assessment: Progress Toward a Seamless System?" *OECD Education Working Papers*, No. 58,

OECD Publishing. doi: 10.1787/5kghx3kbl734-en

Miller, M.D., Linn, R.L., Gronlund, E.N. (2009). *Measurement and Assessment in Teaching* (10th ed.). Upper Saddle River, New Jersey: Pearson Education, Inc.

O'Leary, Z. (2004). *The Essential Guide to Doing Research*

Oliver's Yard, 55 City Road, London: SAGE Publications Ltd

Salkind, N. J. ed, (2007). *Encyclopedia of Measurement and Statistics*.

Thosand Oaks, California: SAGE Publication Inc.,

Statistics, Ethiopia. (2009). *Statistical Abstract*: Ethiopia. Addis Ababa: Central Statistical Agency.

Surafel Tilahun. (2004). *Assessment of the Relationship between General Secondary Education, EGSECE and a Year One Preparatory Students' Academic Performance in East Wollega Zone Secondary Schools* Unpublished, MA Thesis, Addis Ababa University

Urbina, S. (2004). *Essentials of Psychological Testing*.

John Wiley & Sons, Inc., NJ: Hoboken.

Wosen Woga. (2007). *High School Average Score, EGSECE GPA, and Entrance Exam Results as Predictors of College Performance (The Case of Arbaminch College of Teachers' Education)* Unpublished, MA Thesis, Addis Ababa University.

Appendices

Appendix A: Questionnaire

Thank you very much for your willingness to complete this questionnaire.

It will probably take you about fifteen to twenty minutes only to fill it in.

Before you begin to look at the questions, please take a few moments to read the following information as it will help you to understand why it is required to conduct this study and what is expected from you.

Why Do a Study on the Mathematics Model Exams for Grade 10?

Although it has been a long staying practice of conducting model exams in our high schools, very little is actually known about these exams in relation to their nature, uniformity as well as their clear-cut purpose in the education system. Moreover there is also lack of empirical analysis concerning whether they model (replicate) in every aspect those exams that they are suppose to model, i.e., the centrally (nationally) prepared exams. At best, it is possible to make assumptions about their usefulness and, at worst, we rely on the image or idea of these exams that has become fixed through being widely held and that are passed from one generation to the next. Thus I hope that your feedback to this questionnaire will give a great deal of contribution in analysing the nature, usefulness and values of the Grade 10 Mathematics Model Exams that are administered in our high schools to date.

General Direction

This questionnaire is designed just to get your feedback on the Grade 10 Mathematics Model Exam which is either prepared by the regional education bureau or the zone education bureau or math department at your school or by individual teacher at the school level accordingly. Your answers will help not only for this study but also for a broader research in the future! This survey is not like a test. There is no right or wrong answers - you will only be asked to give your opinions about a range of ideas. You don't have to be really smart to take part, because this is about only your views not your knowledge. This questionnaire has five parts. Part I – Part IV deal with the process of item writing and assembling, administering, scoring and making academic decisions based on the students' scores on this model exam respectively while the last one, Part V, asks you to give a "Personal Profile" of yourself. This will help to ensure that the study has managed to get responses from a good cross-section of mathematics teachers. If this is not done, then this study will be biased.

Part I (Item writing and assembling process of the Grade 10 Math Model Exam)

Below is a list of 34 statements. Some of these you might agree with and others you might not. These statements are designed to provoke a response. They do not necessarily reflect my opinions. You are kindly requested to tell what your response is to each statement by putting a tick mark (✓) ***inside the box*** against the statement and below your corresponding response (to the statement) given on the top (first) raw.

Key**: SA=** *Strongly Agree***, A=** *Agree*, **Und=** *Undecided***, D=** *Disagree***, SD=** *Strongly Disagree*

EGSECME= *Ethiopian General Education Certificate Mathematics Examination.*

NEAEA= *National Educational Assessment and Examinations Agency.*

About the Place where the Mathematics Model Exam of your school is prepared						
No	**Item**	**SA**	**A**	**Und**	**D**	**SD**
1	I believe it is recommendable to prepare Mathematics Model Exam at the school level by capable and experienced individual teacher					
2	I suggest it is preferable to develop Math Model Exam in the school by a group of selected math teachers from the department					
3	I believe it is more advantageous to prepare Mathematics Model Exam at the Woreda or Sub-city or Zone education bureau by the subject experts					
4	I think it is better if Mathematics Model Exam will be prepared at the level of Regional education bureaus by selected subject experts.					
Regarding the Units of Lesson/Content Area that the Model Exam Covers						
No	**Item**	**SA**	**A**	**Und**	**D**	**SD**
5	I think it is quite possible, if Grade 10 Math Model Exam covers all units of Grade 10 only					
6	Since Model Exam is a Practice Test I suggest that the items should be selected based on the depth or breadth of contents in each chapter because students can understand the concepts and become confident on EGSECME.					
7	Since it is a Model Exam, the items must be prepared from both the contents of Grade 9 and Grade 10 in proportion to the total number periods allotted.					
8	I generally believe that items of Grade 10 Math Model Exams should be prepared from both the contents of grades 9 & 10 based on a table of specifications.					
9	Since Model Exam is a Practice Test I strongly suggest that the items should be selected from chapters that the students feel difficult because they can practice how to solve them and become confident on EGSECME.					
10	Because it is a Model Exam, I suggest 50% of the items in the exam should be selected from the contents of Grade 9 and the remaining 50% from Grade 10.					

How to determined the number of item from a given chapter for the Model Exam						
No	**Item**	**SA**	**A**	**Und**	**D**	**SD**
11	I suggest that, for Grade 10 Math Model Exam the number of items from a given chapter should depend on the number of periods allotted, the nature of the learning outcomes to be assessed and the depth or breadth of contents of that chapter.					
12	It is recommended by some teachers that the number of items selected from a given chapter should be based on the difficulty or easiness of the concepts in the chapter.					
13	I think it is better if the number of items from a given chapter depend on whether that chapter is totally or partially covered in the academic year.					
14	Some people say the number of items chosen from a given chapter should be based on the students' ability to answer the questions from that particular chapter.					
15	It is possible to determine the number of item from a given chapter after consulting or reviewing the EGSECE booklets of the previous years.					
The Role of Curriculum during the preparation of Math Model Exam						
No	**Item**	**SA**	**A**	**Und**	**D**	**SD**
16	As much as possible the Grade 10 Math Model Exam should assess all the intended contents and the corresponding depth of knowledge indicated in the "*Formal Curriculum*" or national curriculum.					
17	Since the Grade 10 Math Model Exam is considered as a "Practice Test", I think it is very important if the Model Exam contains items from that portion of the curriculum over which teachers "drill" students on test taking techniques using released items and item format of EGSECME.					
18	I suggest that the Model Exam should be prepared by considering both the "*Implemented Curriculum*", i.e., what is actually taught, and how it is actually being taught and the "*Achieved Curriculum*" , i.e., what the students have actually learned as well as their attitudes toward math.					
19	During the preparation of the Model Exam at the outset care should be taken to avoid the negative effect of the "*Hidden-Curriculum*" such as test frauds committed by individual or systematically organized group of students on the testing days.					
20	Since the Grade 10 Math Model Exam is considered as a "Practice Test", I think it is not necessary to make alignment of items or questions with curriculum objectives.					

The cognitive level on which priority should be given when writing items for Math Model Exam.						
No.	**Item**	**SA**	**A**	**Und**	**D**	**SD**
21	I think, in math exam, items that ask for *Knowledge* or *Recall* of facts are easy, and hence such items should be given priority when writing items for Math Model Exam and the number of these items should be kept as large as possible.					
22	I suggest that during the preparation of Math Model Exam questions that ask for the *Comprehension* part should be kept to the minimum.					
23	Since it is a Math Exam, I generally believe that questions on the *Application* of Theorems, Postulates, Laws and Formulas are very important, and hence such kind of items must be larger in number and given more priority than the others					
24	Items that ask for high order thinking ability (*Analysis, Synthesis* and *Evaluation*) are complicated for students and difficult to prepare, therefore it is better to avoid such kind of questions when writing items for Grade 10 Math Model Exam					
25	I generally believe that, Grade 10 Math Model Exam is a “Practice Test” so when preparing this exam equal emphasis should be given to each cognitive level in proportion the related learning outcomes intended to be measured.					
On the use of “Test Blueprint (or Item Specification)” in the preparation of Math Model Exam						
No.	**Item**	**SA**	**A**	**Und**	**D**	**SD**
26	I think *Test Blueprint* (or *Item Specification*) is used to prepare the Grade 10 Math Model Exam in many high schools.					
27	I think you have heard of this document and its usage either as a student in the Teacher Training College/ University or as a teacher on a certain work-shop.					
28	I believe *Item Specification* is not as such important to prepare Grade 10 Math Model Exam.					
29	I think it is very difficult to prepare *Test Blueprint* every year, so it is not used to prepare Grade 10 Math Model Exam.					
30	Even if teachers are well aware of its importance, I think they do not use it to prepare Grade 10 Math Model Exam in your school					

Concerning students' reaction towards the contents of Mathematics Model Exam						
No.	**Item**	**SA**	**A**	**Und**	**D**	**SD**
31	I think students are always complaining that the questions in the Grade 10 Math Model Exam were not selected from content areas that they had learnt in the classroom					
32	I believe that students kept silent to give their opinion on Model Exam because they don't give much attention to it at all or they don't assume this exam as significant as the EGSECME.					
33	Majority of the students said that the Model Exam was easier than the GSECE.					
34	I think the students are satisfied by the contents of the Math Model Exam because it fulfills their expectation in preparing them for EGSECME.					

Part II (Administering the Grade 10 Math Model Exam)

Below is a list of 4 items. Some of these you might agree with and others you might not. These statements are designed to provoke a response. They do not necessarily reflect my opinion. You are kindly requested to tell what your response is to each statement by putting a tick mark (✓) ***inside the box*** against the statement and below your corresponding response (to the statement) given on the top (first) raw

No	**Item**	**SA**	**A**	**Und**	**D**	**SD**
1	To create conducive environment for the administration of Math Model Exam in high schools, I suggest effort should be made in order to make special seating arrangement of the students. For example, **not** to allow more than two students to sit on one desk as in *GSECME.*					
2	If students are found cheating or helping each other while they are taking the Model Exam and since this is unethical, I strongly suggest that disciplinary measures should be taken.					
3	I think during the exam period, usually students complete Math Model Exam and leave the exam hall in less than 30minutes or $\frac{1}{5}$ of the total given time for the exam.					
4	Since the Grade 10 Math Model Exam replicates or models the *GSECME*, I suggest special orientation should be given to students about the rules and regulations that they should follow during the exam period of Math Model as well as the purpose of this Model Exam.					

Part III (Scoring the Grade 10 Math Model Exam)

Below is a list of 7 statements. Some of these you might agree with and others you might not. These statements are designed to provoke a response. They do not necessarily reflect my opinions. You are kindly requested to tell what your response is to each statement by putting a tick mark (✓) ***inside the box*** against the statement and below your corresponding response (to the statement) given on the top (first) raw

No	**Item**	**SA**	**A**	**Und**	**D**	**SD**
1	I think a student's obtained score on Math Model Exam can be added to or combined with the student's scores on classroom performance so as to get the total score of the student in the semester.					
2	I suggest that the score which a student obtained on Grade 10 Math Model Exam should be reported independent of any other score, because this helps to predict his/her efficiency on the *GSECME*.					
3	After the correction of Math Model Exam, when assigning score and reporting the student's result, I think there is nothing wrong if the reported score is converted to 100. (For example, *out of 60 questions of Math Model Exam if a student correctly answers 20 questions, then the student's obtained score is always reported as* $\frac{33}{100}$)					
4	Student's reported score on Grade 10 Math Model Exam truly reflects his/her mastery of the subject matter.					
5	If the results of the previous several years are studied through detail analysis of school records, I think one can see that students who scored high on Grade 10 Math Model Exam also got high grades on the *GSECME*..					
6	If the results of the previous several years are studied through detail analysis of school records, I think one can see that students who scored high on Grade 10 Math Model Exam also scored high on their University Entrance Math Examinations at the Preparatory levels.					
7	As a matter of fact Grade 10 Math Model Exam is considered as "Practice Test" so the student's score should be taken as an indicator of the student's preparedness for EGSECME and hence it is better not to report it as achievement score.					

Part IV (Academic decisions based on scores on the Grade 10 Math Model Exam)

Below is a list of 5 statements. Some of these you might agree with and others you might not. These statements are designed to provoke a response. They do not necessarily reflect my opinions. You are kindly requested to tell what your response is to each statement by putting a tick mark (✓) ***inside the box*** against the statement and below your corresponding response (to the statement) given on the top (first) raw

No	**Item**	**SA**	**A**	**Und**	**D**	**SD**
1	Due to any reason if a student does not take Math Model Exam, I strongly suggest that he/she should be considered as *incomplete* and hence should not be allowed to take the *GSECME*.					
2	No matter how high or low a student scored on the Math Model Exam, I think the primary issue in taking the Model Exam is to make decision about that student (i.e., either he/she is allowed or denied to take the *GSECME*).					
3	If a student fails in the Model Exam but gets a passing grade on the GSECE, then I suggest that this student should leave his/her present school but can continue his/her schooling in another school.					
4	From my previous observations, students who had got a passing score on Grade 10 Math Model Exam in one school and proceeded their Preparatory education in another school, did **not** get passing scores either on Grade 12 Math Model Exam or University Entrance Math Examinations or on both.					
5	It is necessary to design and implement a kind of system among all schools that enables them to make uniform, sound and informed academic decisions based on scores on the Model Exam					

Part V

Here you are not asked for your name, so don't write your name and you will remain anonymous.

Age ___________ Sex_______________

Name of the School where you are teaching now_______________________________________

Service year(s) in teaching (Mathematics) _____________

Educational qualification ____________________________

Did you take any Measurement and Evaluation Courses?

Yes ____ No______

THANK YOU VERY MUCH !!!

Appendix B: Item Statistics

	Mean	Std. Deviation	N
Part I- Item 1	3.26	1.365	31
Part I- Item 2	3.52	1.313	31
Part I- Item 3	2.55	1.387	31
Part I- Item 4	3.55	1.609	31
Part I- Item 5	2.10	1.300	31
Part I- Item 6	4.32	.832	31
Part I- Item 7	4.58	.765	31
Part I- Item 8	4.45	.810	31
Part I- Item 9	2.81	1.302	31
Part I- Item 10	3.81	1.250	31
Part I- Item 11	4.10	.944	31
Part I- Item 12	3.23	1.175	31
Part I- Item 13	3.23	1.431	31
Part I- Item 14	2.52	1.208	31
Part I- Item 15	3.26	1.182	31
Part I- Item 16	4.45	.506	31
Part I- Item 17	3.65	.877	31
Part I- Item 18	4.13	.806	31
Part I- Item 19	3.97	.605	31
Part I- Item 20	2.42	.923	31
Part I- Item 21	3.42	1.089	31
Part I- Item 22	3.13	.991	31
Part I- Item 23	3.71	1.160	31
Part I- Item 24	2.81	1.138	31
Part I- Item 25	4.03	.657	31

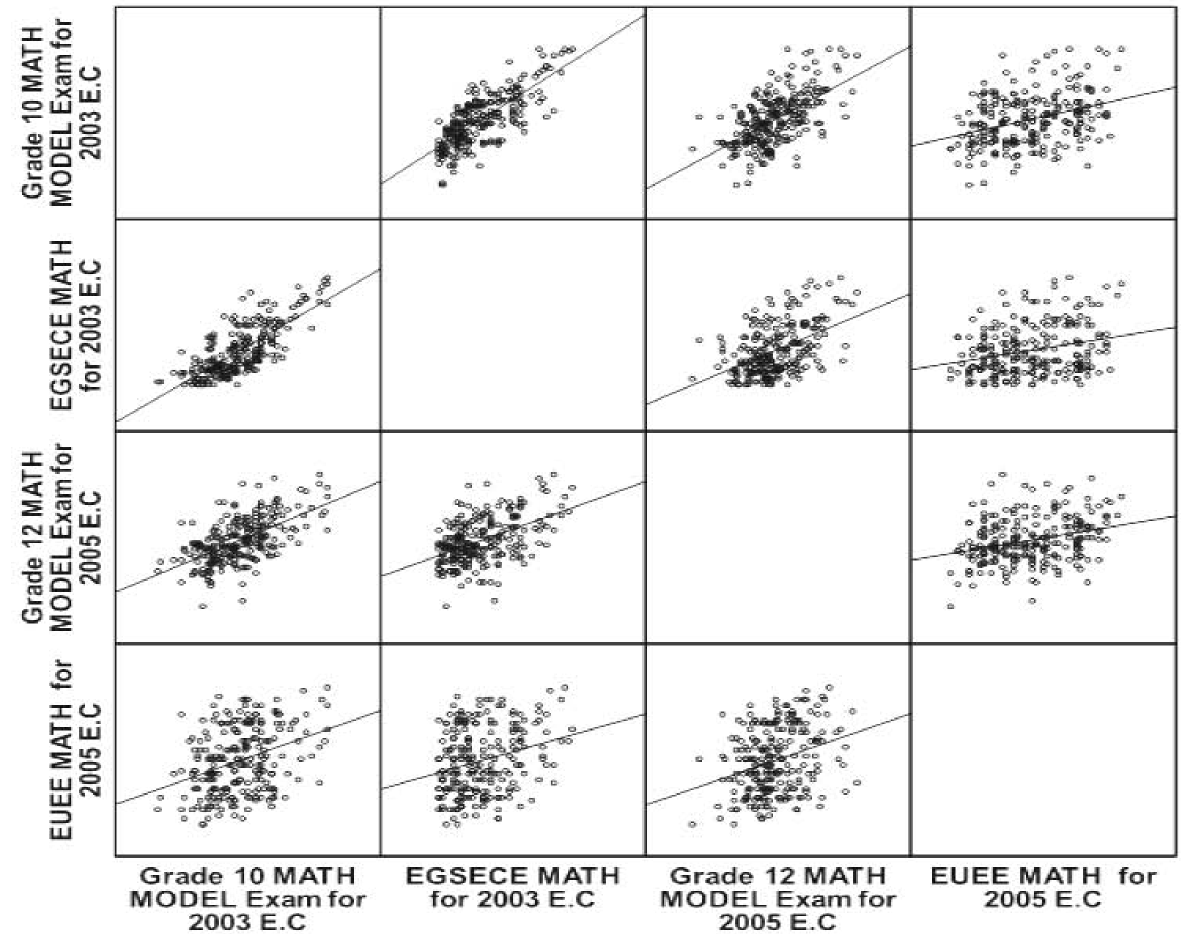
	Mean	Std. Deviation	N
Part I- Item 26	3.90	1.044	31
Part I- Item 27	3.74	.815	31
Part I- Item 28	2.16	.779	31
Part I- Item 29	2.29	1.039	31
Part I- Item 30	2.87	1.118	31
Part I- Item 31	2.65	1.082	31
Part I- Item 32	3.35	1.170	31
Part I- Item 33	2.90	1.076	31
Part I- Item 34	3.77	.805	31
Part II- Item 1	4.55	.675	31
Part II- Item 2	4.68	.475	31
Part II- Item 3	2.97	1.303	31
Part II- Item 4	4.48	.570	31
Part III- Item 1	3.77	1.146	31
Part III- Item 2	3.87	.806	31
Part III- Item 3	3.45	.995	31
Part III- Item 4	3.48	.811	31
Part III- Item 5	3.90	.831	31
Part III- Item 6	3.81	.833	31
Part III- Item 7	3.13	1.147	31
Part IV- Item 1	3.03	1.516	31
Part IV- Item 2	2.68	1.013	31
Part IV- Item 3	2.90	1.076	31
Part IV- Item 4	2.77	.920	31
Part IV- Item 5	4.16	.735	31

Appendix C: Item-Total Statistics of the Questionnaire

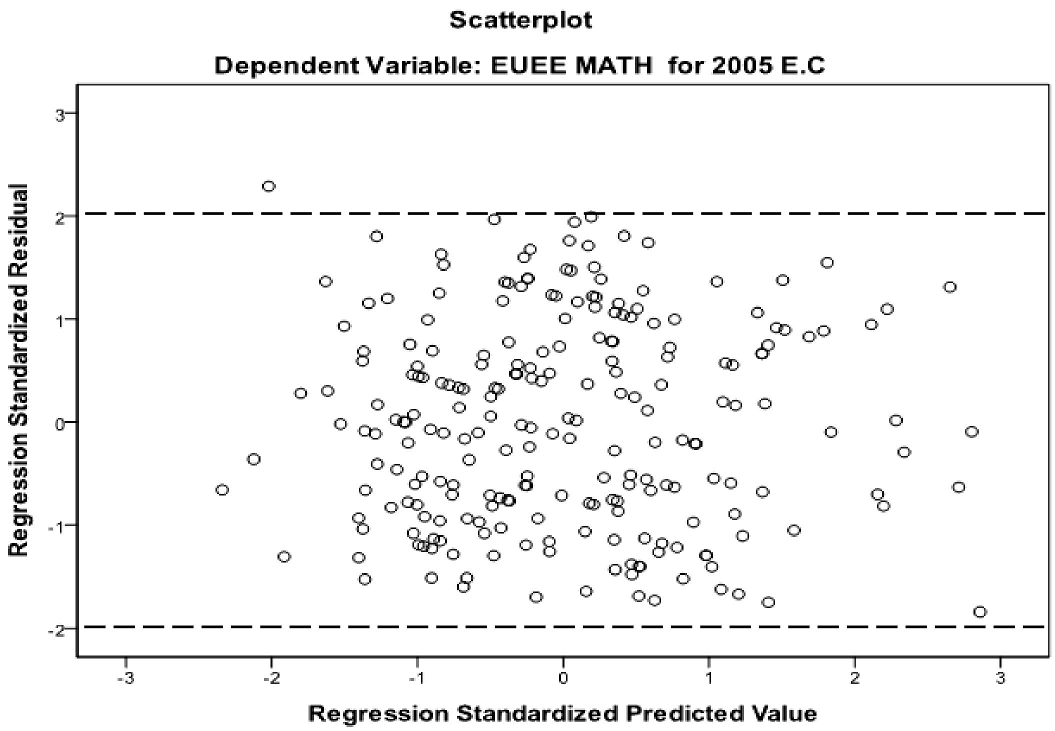
	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
***Part I*- Item 1**	169.00	234.400	.242	.793
***Part I*- Item 2**	168.74	239.598	.125	.797
***Part I*- Item 3**	169.71	223.680	.501	.782
***Part I*- Item 4**	168.71	245.280	-.029	.806
***Part I*- Item 5**	170.16	231.806	.326	.790
***Part I*- Item 6**	167.94	237.796	.308	.791
***Part I*- Item 7**	167.68	234.026	.504	.787
***Part I*- Item 8**	167.81	233.761	.484	.787
***Part I*- Item 9**	169.45	227.923	.427	.786
***Part I*- Item 10**	168.45	230.856	.368	.788
***Part I*- Item 11**	168.16	246.940	-.048	.800
***Part I*- Item 12**	169.03	234.499	.292	.791
***Part I*- Item 13**	169.03	228.166	.374	.787
***Part I*- Item 14**	169.74	230.531	.393	.787
***Part I*- Item 15**	169.00	227.400	.494	.784
***Part I*- Item 16**	167.81	244.428	.108	.796
***Part I*- Item 17**	168.61	234.912	.398	.789
***Part I*- Item 18**	168.13	236.116	.389	.789
***Part I*- Item 19**	168.29	241.280	.253	.793
***Part I*- Item 20**	169.84	239.006	.229	.793
***Part I*- Item 21**	168.84	231.406	.417	.787
***Part I*- Item 22**	169.13	252.649	-.230	.806
***Part I*- Item 23**	168.55	239.389	.158	.795
***Part I*- Item 24**	169.45	234.523	.304	.791
***Part I*- Item 25**	168.23	240.047	.290	.792

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
***Part I*- Item 26**	168.35	231.103	.447	.786
***Part I*- Item 27**	168.52	243.725	.079	.797
***Part I*- Item 28**	170.10	244.824	.040	.797
***Part I*- Item 29**	169.97	238.966	.198	.794
***Part I*- Item 30**	169.39	244.178	.028	.799
***Part I*- Item 31**	169.61	234.312	.330	.790
***Part I*- Item 32**	168.90	241.357	.101	.797
***Part I*- Item 33**	169.35	231.570	.418	.787
***Part I*- Item 34**	168.48	242.725	.121	.796
***Part II*- Item 1**	167.71	251.280	-.249	.802
***Part II*- Item 2**	167.58	249.118	-.196	.800
***Part II*- Item 3**	169.29	232.413	.309	.790
***Part II*- Item 4**	167.77	248.047	-.110	.799
***Part III*- Item 1**	168.48	238.591	.183	.795
***Part III*- Item 2**	168.39	241.178	.183	.794
***Part III*- Item 3**	168.81	240.828	.148	.795
***Part III*- Item 4**	168.77	235.314	.419	.789
***Part III*- Item 5**	168.35	233.437	.483	.787
***Part III*- Item 6**	168.45	230.523	.600	.784
***Part III*- Item 7**	169.13	245.049	.001	.800
***Part IV*- Item 1**	169.23	236.047	.173	.796
***Part IV*- Item 2**	169.58	234.385	.354	.789
***Part IV*- Item 3**	169.35	235.303	.301	.791
***Part IV*- Item 4**	169.48	234.658	.386	.789
***Part IV*- Item 5**	168.10	241.957	.171	.794

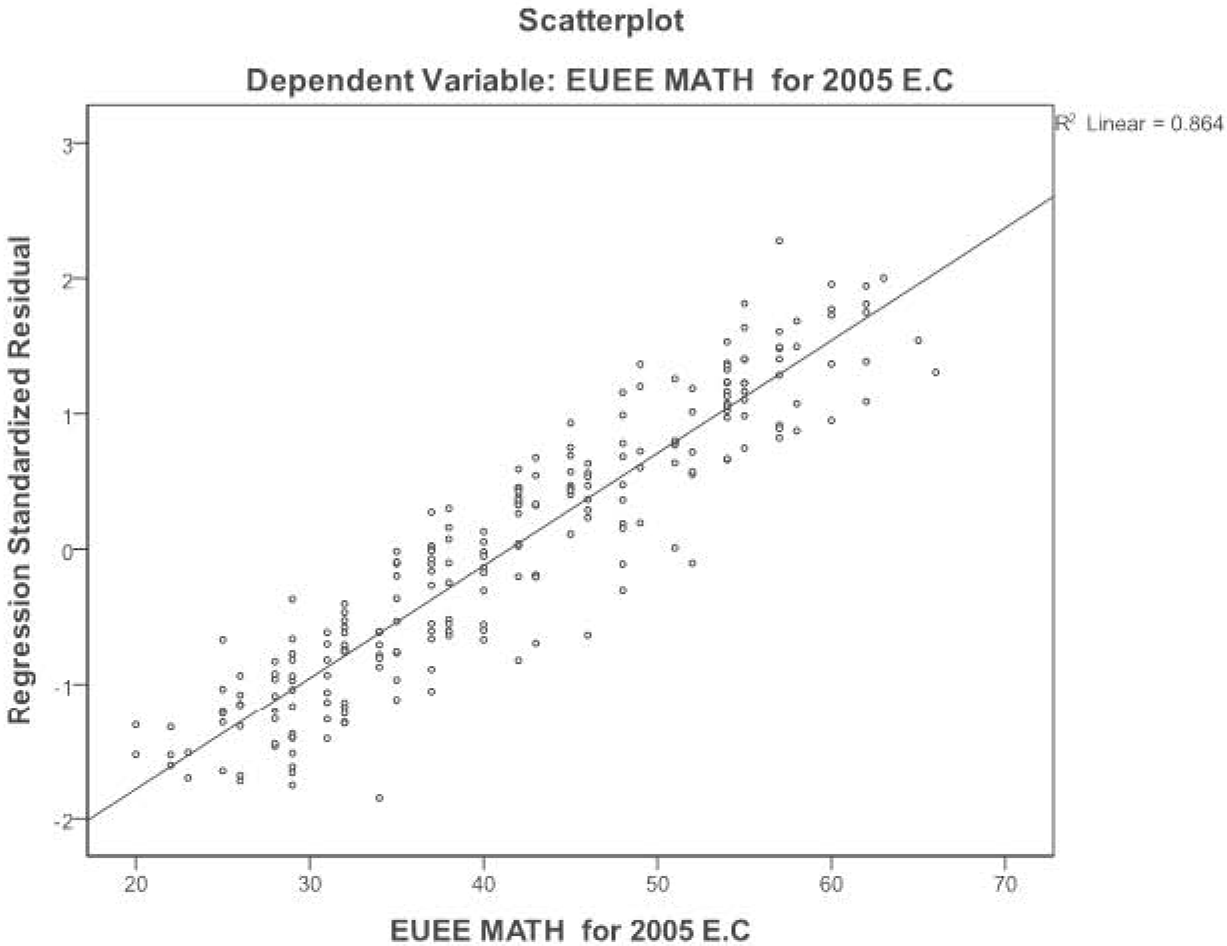
Appendix D: ***The scatter plot matrix***



Appendix E: ***The scatter plot showing the distribution standardized residuals vs. standardized predicted scores (for Normality Assumption)***



Appendix F: ***The scatter plot showing the distribution of standardized residuals vs. criterion raw scores (for Colinearity Assumption)***



Appendix G: Consolidated General Remarks of the subject matter experts on the Grade 10 Math Model Exams on

The preparation of this Grade 10 Math Model Exam was not adequately planned

1. There are questions in which students can guess the answer from the distracters
2. There are questions which have either multiple answers or no answer at all
3. There are some observable language problems
4. There are contents not included in the exam
5. All domains are not proportionally distributed and hence there are questions which measures the same learning out comes from the same topic
6. There are questions which may not be understood by the students
7. Letter on the figures are not clearly indicated and there is also missed picture
8. There are questions which lead students to mistake or to chose the wrong answer
9. Letters of the correct answer or answer keys are not equally distributed
10. The questions are not well sequenced
11. The exams were not edited or the exams' item writing process was not assisted by mathematics equation editor available on every desktop or personal computer.

Appendix H: Bio-data of Teachers Responded to the Questionnaire

Age	Ans. Not Given	**25 - 29**	**30 - 34**	**35 - 39**	**40 - 44**	**45 - 49**	**50 and above**
No.	***2***	***10***	***6***	***2***	***5***	***3***	***3***

Sex	Ans. Not Given	**Female**	**Male**
No.	***1***	***2***	***28***

Service year(s) in teaching Mathematics	Ans. Not Given	**1 – 8**	**9 - 16**	**17 - 24**	**25 - 33**
No.	***2***	***13***	***7***	***3***	***6***

Educational qualification	Ans. Not Given	**12 + 4**	**B.Ed.**	**B.Sc.**	**Degree (Not Specified)**
No.	***4***	***2***	***6***	***4***	***15***

About taking Measurement and Evaluation Courses	Ans. Not Given	**Yes**	**No**
No.	***1***	***25***	***5***

Appendix I: Interview Questions

For School Administrators, Woreda and Regional Education Bureaus Officials.

First I would like to thank you for giving me your time for this interview.

The main objective of this interview is to know more about the Grade 10 Math Model Exam and to share from your experience in this regard.

Would you please tell me your

Name ______________________. Position_______________

Work experience in: - the current position _______________

- another position/s _______________

I. Let us begin with the history of Model Exam in our school system.

So just tell me:

1- What was the main purpose or function of Model Exams?

2. Is there any official document issued by the concerned bodies like MoE or others in your region that explains the role of these Model Exams?

When was this document issued?

3- Can we take it as a "Practice Test" so that it gives students an idea of what to expect on the exam days of the national exam, (i.e., EGSECE) and hence it will increase their confidence in taking this national exam?

4- Can we consider it as an "Achievement Test" to measure what the students achieve as the result their schooling?

5- Does it deserve the name "Model Exam"?

II. On Item writing and assembling process of the Grade 10 Math Model Exam)

1) Where do you suggest the Mathematics Model Exam of your school should be prepared?

- *At the school level* (by capable and experienced individual teacher). Why?
- *By the department of Math* in the school (by a group of selected math teachers)
- *At the Woreda or Zone education bureau* (by the subject experts or selected teachers from schools in the Woreda or Zone)
- *At the level of Regional education bureaus* (by selected subject experts).

2) Can you tell me about the Units of Lesson/Content Area that the Model Exam Covers?

3) How does the number of item from a *given chapter* for the Model Exam is determined?

4) What is the *Role of the Curriculum* during the preparation of Math Model Exam?

5) Is it necessary to use "*Test Blueprint "(*or *"Item Specification*) in the preparation of Math Model Exam?

6) What is the *students' reaction* towards the contents of Math Model Exam *and /or* the Exam itself?

***III.* In Administering the Grade 10 Math Model Exam**

1- Do you think it is necessary to make any special *seating arrangement* of students during the Model Exam?

2- If students are *found cheating or helping each other* while they are taking the Model, what is your reaction towards such students?

3- Since the Grade 10 Math Model Exam *replicates or models the GSECME*, do you think it is necessary to give special *orientation* to students about the rules and regulations that they should follow during the exam period of Math Model ?

***IV.* On the scoring Process of the Grade 10 Math Model Exam.**

1- *Where do you correct* the Grade 10 Math Model Exam?

2- Can student's *obtained score* on Math Model Exam be *added to or combined with* the student's *scores on classroom performance* so as to get the total score of the student in the semester?

3- Do you agree if the *score which a student obtained* on Grade 10 Math Model Exam *be reported independent of any other score*, because this helps to predict his/her efficiency on the GSECME? Can you explain me why?

4- After the correction of Math Model Exam, when *assigning score* and *reporting the student's result*, what do you think, if *it is reported out of 100?* Is it *wrong* or *right*? *Why?*

5- Do you think a *student's reported score* on Grade 10 Math Model Exam *truly reflects his/her mastery of the subject matter? Why?*

6- Do you think students *who scored high* on Grade 10 Math Model Exam *also got high grades* on the *GSECME? Why?*

7-Did you observe that majority of the students *who score high* on Grade 10 Math Model Exam also score high on their *EUE* Math Exam at the Preparatory levels? Is this happened frequently, I mean every year or occasionally?

***V*. Academic decisions based on scores on the Grade 10 Math Model Exam**

1- Due to any reason if a student is found unable to take Math Model Exam, do you allow this student to take the *GSECME*?

2- No matter how high or low a student scored on the Math Model Exam, what do you think you think about the primary issue in taking the Model Exam? Is to make decision about that student (i.e., either he/she is allowed or denied to take the *GSECME*)?

3- From your previous observations, have you ever seen that students who had got a passing mark on Grade 10 Math Model Exam in one school, did **not** get passing scores either on Grade 12 Math Model Exam or University Entrance Math Examinations or on both, when they proceed their Preparatory education in another school?

5- Do you support the idea of designing and implementing a kind of system among all schools that enables them to make uniform, sound and informed academic decisions based on scores on the Model Exam?

☞ ***VI***. So from what you told me so far, can you summarize and tell me the clear-cut purpose and function of the Grade 10 Math Model Exam?

THANK YOU VERY MUCH!!!

Appendix I: Check List *(Adapted from "Measurement and Assessment in Teaching" by Miller, Linn & Gronlund, E.N. (10th ed.). (2009, p. 150)*

	Adequacy of the Assessment Plan	***Yes***	***No***
1	Is the preparation of this Grade 10 Math Model Exam adequately planned?		
2	During the preparation of this Math Model Exam, is appropriate emphasis given to the objectives and each content area to be measured at this grade level?		
	Adequacy of Test Items and the Corresponding Tasks	***Yes***	***No***
3	Is the format of the item and task suitable for the learning outcome being measured ***(appropriateness)***?		
4	Does each item or task require students to demonstrate the performance described in the specific learning outcome it measures **(*relevance*)**?		
5	Does each item or task present a clear and definite task to be performed ***(clarity)***		
6	Is each item or task presented in simple readable language and free from excessive verbiages ***(conciseness)***?		
7	Does each item or task provide an appropriate challenge ***(ideal difficulty)***?		
8	Does each item or task have an answer that would be agreed upon by different mathematics experts ***(correctness)***?		
9	Is there a clear basis for awarding partial credit on items or tasks with multiple points ***(scoring rubric)?***		
10	Is each item or task free from technical errors and irrelevant clues ***(technical soundness)***?		
11	Is each item free from ethnic, religious and gender bias ***(cultural fairness)***?		
12	Is each test item independent of the other item in the test ***(independence)***?		
13	Is there an adequate number of test item for each learning outcome ***(sample adequacy)***		

	Adequacy of Test Format and Directions	***Yes***	***No***
14	Are test items of the same type grouped together in the test (or within sections of the test)?		
15	Are the test items arranged from easy to more difficult within Sections of the test and the test as a whole?		
16	Are the test items numbered in sequence?		
17	Are the correct answer distributed in such a way that there is no detectable pattern?		
18	Is the test material well spaced, legible and free of typographical errors?		
19	Are there directions for each section of the test and the test as a whole?		
20	Are the directions clear and concise?		
	About this Math Model Exam in General	***Yes***	***No***
21	Can this Model Exam be considered as “Practice Tests” which gives students an idea of what to expect on the EGSECE and increases their confidence in taking the national exam by practicing on the items of this Model Exam?		
22	Does this Model Exams really serve as an “Achievement Test”, i.e., used to measure what the students achieve as the result their schooling?		
23	Does this Model Exams really model or copy the national EGSECE?		

Declaration

I hereby declare that this thesis is my original work done under the guidance of Dr. Dame Abera. All relevant sources used for the thesis are duly acknowledged.

Name ______________________________

Signature ______________________________

Place ______________________________

Date of Submission ______________________________

This thesis has been submitted for examination with my approval as University advisor.

____________________	____________________	____________________
Advisor	**signature**	**Date of Approval**