Ethiopian Primary School Students' Mathematics Achievement through the lens of the Trends International Mathematics and Science Study

Effa Gurmu Bati

Addis Ababa University
Addis Ababa, Ethiopia

September, 2020

Ethiopian Primary School Students' Mathematics Achievement through the lens of the Trends International Mathematics and Science Study

By

Effa Gurmu Bati

Supervisors: Kajisa Hansen Yang (Prof.)

Temesgen Fereja (PhD)

A Dissertation Submitted to the Center for Comparative Education and Policy Studies in Fulfillment of the Requirements for Degree of Doctor of Philosophy in International and Comparative Education

Addis Ababa University

Addis Ababa, Ethiopia

September, 2020

Declaration of the Candidate

I, the undersigned, declare that this dissertation entitle:

"Ethiopian Primary School Students' Mathematics Achievement through the lens of the Trends International Mathematics and Science Study" is my original work. It has not been presented for any academic study at any other university. All sources and materials used for this work are duly acknowledged.

Name: Effa Gurmu

Signature:

Date: 2020-06-19

Place: Addis Ababa, Ethiopia

Confirmation of the Advisor

I certify that this dissertation embodies the results of research carried out by the candidate under my supervision. It has been submitted for examination with my approval as university supervisor.

Name of the supervisor: Kajsa Yang Hansen

Signature:

Date: 2020-06-16

Place: Gothenburg, Sweden

Name of Supervisor: Temesgen Fereja

Signature:

Date: 2020-06-19

Place: Addis Ababa University,Ethiopia

Acknowledgments

Many people have contributed in different ways throughout the process of carrying out this study, and I am indebted to all of them. My first thanks go to Professor Kajisa Hansen Yang, University of Gothenburg and the main supervisor of this study, who has extended the utmost professional support, unreserved affection, and encouragement. I am grateful to my co-supervisor Dr. Temesgen Fereja of Addis Ababa University, for his follow up and constructive comments since the inception stages of this study. I am also very grateful to Dr. Teshome Nekatibeb and Dr. Alebachew Kemisso for their special support and comments.

I am very much indebted to the National Educational Assessment and Examinations Agency for the research grant and the material supports provided. I extend my deepest thanks to Ato Araya Geberegziabher and Dr. Zerihun Duressa, the General Manager and Deputy General Manager of NEAEA respectively, for their encouragements, understanding, and tolerance. I would like to say thank you to my colleagues Bekele Geleta, Assefa Leta, Yilkal Wondimeneh, Arega Mamaru, Wondimagegn Hailu, and Eleni Mengistu as well as the rest NEAD staff members for all of their cooperation and support in various ways.

The contribution of Zonal Examination Experts is highly praiseworthy for collecting the required data. Many thanks also go to the sample schools, school principals, teachers, and students who enthusiastically participated in this study. Finally, I would like to extend my sincere gratitude to my family. My wife Mestawot Hailu, I thank you from the bottom of my heart for your unwavering love and encouragement. I extend my deepest gratitude to

my mother Urgo Aboye who always prays about me and sacrifices a lot as well as to all my brothers and sisters who were on my sides at every difficult circumstance.

Dedication

This dissertation is dedicated to my father, the late Gurmu Bati who did not only raise and nurture me but also sacrificed himself dearly over the years for my education and intellectual development, though he did not attend any formal education. Incidentally, he met his demise about the same moment I began my PhD study program on 15 October, 2014. My father was a source of motivation and strength during the moments of despair and discouragement. Recently, his fatherly care and support have been shown in incredible ways. Thank you for teaching me the core values of hard work and commitment.

Abstract

Ethiopia has established the national learning assessment system since 2000. However, it never participated in any regional or large-scale international assessment, though, planned to join it as of the Fifth Education Sector Development Program. The purpose of this study is, therefore, to assess grade eight students' mathematics achievement through the lens of Trends International Mathematics and Science Study (TIMSS). This was a quantitative study, in which the released TIMSS mathematics achievement test items and context questionnaires were employed. The participants were 2,120 students selected through two-stage cluster sampling techniques from 53 schools of the three regions (Addis Ababa, Benishangul Gumuz, and Oromia) and school principals of the respective sample schools. Descriptive and inferential statistics were used to summarize the findings and generalize about a population from the samples. The results indicated that the overall achievements of students' in mathematics were about 32% and 404 points in the raw and scale scores respectively. The mean scale score for Ethiopian students was significantly better than the mean score of four African countries that participated in TIMSS study in 2011 but lower than the international mean score ($p < 0.05$). There was a statistically significant achievement mean difference across gender and school location, in favour of boys and rural schools at $p < 0.05$. The highest regional average achievement was observed in Oromia and the lowest in Benshagul Gumuz, while Addis Ababa was in-between. There was no statistically significant difference observed between the achievements of Ethiopian students' in this study and National Learning Assessment (NLA) study in 2016, with the mean scale scores of 302.05 and 300.36 respectively ($p < 0.05$). The main significant factors that affect the achievement of students were: gender, school location family SES, students' attitude to school and mathematics, instructional language, school feeding, school emphasis on academic success, and principals' qualification. In general, the achievement of students in mathematics was found below the expected national and international standards. Therefore, policy-makers and concerned stakeholders should give special attention to enhance the achievement of students in mathematics, particularly for those disadvantaged low achieving groups. Finally, to participate in International Large-scale Assessment (ILSAs) and make the right decision, further studies should be conducted countrywide as this study has left with unanswered questions.

Table of Contents **Page**

List of Figures

List of Tables

List of Abbreviations and Acronyms

ANOVA	Analysis of Variance
CFI	Comparative Fit Indices
CTT	Classical Test Theory
EGMA	Early Grade Mathematics Assessment
EGRA	Early Grade Reading Assessment
ESDP	Education Sector Development Program
ETS	Educational Testing Service
FIMS	First International Mathematics Study
GEQIP	General Education Quality Improvement Program
HSD	Honestly Significant Difference
IAE	International Academy of Education
IATA	Item And Test Analysis
ICC	Intraclass Correlation Coefficient
ICMI	International Commission on Mathematical Instruction
IEA	International Association for the Evaluation of Educational Achievement
IIEP	International Institute for Educational Planning
ILSA	International Large-scale Assessment
IRT	Item Response Theory

JICA	Japan International Cooperation Agency
LaNa	Literacy and Numeracy Assessment
MD	Mean Difference
MELQO	Measuring Education Learning Quality Outcomes
MLC	Minimum Learning Competency
MOE	Ministry of Education
NAE	National Agency for Examination
NCTM	National Council of Teachers of Mathematics
NEAD	National Educational Assessment Directorate
NEAEA	National Educational Assessment and Examinations Agency
NLA	National Learning Assessment
OECD	Organization for Economic Cooperation and Development
RMSEA	Root Mean Square Error of Approximation
PIRLS	Progress in International Reading Literacy Study
PISA	Progress of International students Achievement
SABER	Systems Approach for Better Education Results
SACMEQ	Southern African Consortium for the Monitoring of Education Quality
SD	Standard Deviations
SDG	Sustainable Development Goals
SEM	Structural Equation Modelling
SES	Socio-Economic Status

SIMS	Second International Mathematics Study
SMASEE	Strengthening Mathematics and Science Education in Ethiopia
SNNP	Southern Nations, Nationalities, and People
SPSS	Statistical Package for Social Sciences
SRMR	Standardized Root Mean Square Residual
TAP	Test Analysis Program
TIMSS	Trends International Mathematics and Science Studies
UNESCO	United Nations Educational, Scientific and Cultural Organization
US	United States

CHAPTER ONE: INTRODUCTION

1.1. Background of the Study

In education, assessment refers to the broad kinds of techniques that educators use to evaluate, measure, and record the instructional readiness, learning improvement, skill acquisition, or educational desires of students. It is a process of gathering and analyzing information on what students know, understand, and can do to make an informed decision about the next steps in the educational process (Clarke, 2012). Thus, assessing students' academic performance is a vital part of the education process and has been a component of education for as long as education has been a part of civilization. The ultimate rationale for assessing students' achievement is to improve student's learning by monitoring the students' acquisition of knowledge and progress in mastering the curriculum. According to Salvia and Ysseldyke (1995), the main reason to conduct an academic assessment is to identify and verify the challenges in education to make the right decision. Lawrence (2013) also clarifies the purpose of academic assessment as identifying student strengths and weaknesses, monitoring student progress, and evaluating student's prior knowledge.

The literature on education classifies the types of assessments into four broad parts based on their purposes (Clarke, 2012). These are classroom assessments, examinations, national assessments, and international/ regional large-scale assessments. Classroom assessments are usually devised and administered by teachers, and they are aligned with the delivered curriculum. According to Braun, Kanjee, Bettinger, & Kremer (2006), the primary goals of classroom assessment are to understand students' learning and improve teaching in which the results are immediately available to the teachers and can influence the subjects of instruction. According to Kaur and Wong (2011), classroom assessment is considered from three perspectives. The first one is Assessment of Learning, which informs the learners about how much they have learned, and identifies the gap between the intended learning goal and current achievement. The second is Assessment for Learning, which is a vehicle for informing the learner to enhance future learning and provide immediate feedback. The last one is, Assessment as Learning, in which the learner internalizes assessment as part of learning and becomes a self-directed learner.

National examinations are standardized tests that are designed, developed, and administered centrally with the subject matter specialists. The purposes of such examinations are to select and certify individual students. According to Braun and et al. (2006), the role of a national examination is selecting learners for admission to secondary or tertiary education, credentialing learners for the world of work, and providing data for holding school staff accountable for their performance. National examination is an essential component of every nation's education system. Notably, it is critical for developing countries, where the number of candidates for advancement is usually greater than the number of places available for the next educational levels.

National assessments are studies focused on generating specific information that policy-makers need to evaluate various aspects of the educational system within the country. They usually used for accountability purposes to make resource allocation decisions, and even to heighten public awareness of education issues (Braun & et al., 2006; Lockheed, 2016). These assessments may be administered to the entire or a sample of the population, and usually include background questionnaires for different participant groups to provide a meaningful context for interpreting test results.

International assessments assess learners in different countries to provide cross-national comparisons that can illuminate a variety of educational policy issues to monitor and evaluate cross-national education systems (Braun & et al., 2006; Lockheed, 2016). Similar to national assessments, it also includes contextual questionnaires for various participants that help to clearly understand the learning outcomes of the students. According to Braun and et al. (2006), these studies are planned and implemented by various organizations, such as TIMSS and PIRLS[1] conducted by IEA[2], PISA[3] by OECD[4], and SACMEQ[5] by UNESCO. According to Clarke (2012), for a country's assessment system to be successful, these various types of assessments need to be aligned with each other and with the country's education learning outcomes and policies. In this study, international large-scale assessment and large-scale international assessment are used interchangeably.

Today, international large-scale assessments (ILSAs) are rapidly expanding and receiving enormous attention around the world. The origin of modern-day international assessments can be traced back to the First International Mathematics Study (FIMS), encouraged by the results from the pilot study of twelve countries. Its purpose was to investigate the feasibility of undertaking more extensive assessments, which was conducted by the International Association for the Evaluation of Educational Achievement (IEA) in the early 1960s (Rutkowski, Davier, & Rutkowski, 2014). International large-scale assessment is a crucial tool in education policy decisions, accountability, and education planning worldwide (Rutkowski, Davier, & Rutkowski, 2014). According to Medrich and Griffith (1992), the main gain of countries to participate in international assessment is to understand

[1] the Progress in International Reading Literacy Study
[2] the International Association for the Evaluation of Educational Achievement
[3] the Program for International Student Achievement
[4] the Organization for Economic Cooperation and Development
[5] the Southern African Consortium for the Monitoring of Education Quality

a much wider variety of education policies, programs, and practices that can help them to improve their educational system. International large-scale assessment studies play a significant role in the public discussion of education in many parts of the world and its results help educational authorities to improve their understanding of education systems and draw conclusions about the strengths or weaknesses of their own with others to reform and manage their educational systems (Säljö & Radišić, 2018). As Torney-Purta and Amadeo (2013), noted a primary reason to conduct ILSA assessments is to shed light on the processes, and contexts of education by examining these processes and contexts comparatively.

Moreover, Bray, Adamson, and Mason (2007) indicated that the reason to compare the achievement of students across countries is to observe what countries can learn from each other in their education systems. Additionally, Lehmann (2011) expressed that, international learning assessments such as TIMSS, as a necessary condition for policy-making. As a result, the participation of countries in international assessment increased over time. Currently, the number of participating countries increased from 12 in the FIMS pilot study to 57 in TIMSS 2015. In this regard, several countries used TIMSS results as a necessary means of informing their educational reform and improvement strategies. For instance, Kuwait, New Zealand, Norway, Romania, and South Africa served the TIMSS result as a catalyst for curricular review and change, and Iceland used it to increase the teaching hours for mathematics and science instruction at the elementary level (Wagemaker, 2014).

As Kellaghan (1996) indicated, despite distinct advantages, there are some problems and limitations associated with the country's participation in ILSA assessments. According to

Lockheed (2015), the participation rates for high-income and upper-middle-income countries are many times higher than participation rates for lower-middle-income and low-income countries. The main reasons for the low participation of developing countries are due to their differences in social, economic, cultural, and political situations as compared to developed countries (Ouane & Singh, 2004). According to the authors, the difference in education systems such as the age at which children start school or a high selectivity in the transition from primary to secondary education, distribution of wealth within countries, the cultural contexts, and feasibility of conducting such assessments are the main challenges. The validity of using country rankings to estimate the educational achievement of students within countries is yet another problem (Torney-Purta & Amadeo, 2013).

Similarly, Backhoff (2013) agreed as inter-country comparisons cannot be based on simple rankings because of many validity threats in nations where there are cultural and linguistic variations. Furthermore, the misinterpretations of findings have negative consequences by spreading misinformation in broader national and societal contexts. In this regard, Lin, Bumgarner, and Chatterji (2014) indicated, ignoring the contextual factors and analyzing the data that only focused on countries' ranking have little meaning.

However, according to the principle of the IEA, variety of instructional viewpoints, models, and tactics that characterize the world's educational structures and systems, establish a natural laboratory in which nations can learn from one another (Rutkowski & Prusinski, 2011). Chmielewski and Dhuey (2016) also showed that the ILSA assessment is used not only to compare performance, curricula, instructional, and learning strategies across countries, it is also used to understand how international variations in education policies (i.e., the structure, administration, legal, economic, and political contexts) shape differences in student achievement and other outcomes.

On top of that, TIMSS is one of the international assessments that has gotten attention and conducted in grades four and eight students that come from a diverse set of educational systems in every four-year cycle to date. TIMSS is a curriculum-based assessment that refers to the evaluation of a student's achievement in the country's contextual curriculum as a basis for collecting data that help to make educational decisions (Deno, 2003). The content assessed in TIMSS is based on an internationally agreed upon standard curriculum in mathematics (Rutkowski & Prusinski, 2011). The focus of curriculum-based assessment is on how well students perform on the materials/textbooks used in the classrooms. A curriculum content-based assessment measures the extent to which students know the contents or standards of a particular subject matter (UNESCO, 2017). Thus, the most benefit of curriculum-based assessment is its alignment with the curriculum materials. That means the students tested on what they are already learned. According to Mullis, Martin, Foy, and Hooper (2016:2), "TIMSS has the goal of helping countries make informed decisions about how to improve teaching and learning in mathematics and science with its strong curricular focus and emphasis on policy relevant information about the home, school, and classroom contexts for learning". Hence, TIMSS is a valuable tool that countries can use to evaluate and monitor their students' achievement in international contexts.

In Ethiopia, the Education and Training Policy of the Federal Democratic Republic Government of Ethiopia accentuated assessment under the title *Educational Measurement and Examination in article 3.3* (MOE, 1994). Here, it specified that continuous assessment in academic and practical subjects conducted to ascertain the formation of the all-round profile of students at all levels, and national examinations are carried out to certify the completion of students at the various levels of education. The policy also showed the importance of the establishment of a National Organization of Educational Measurement and Examination body, which will overtake central professional guidance and coordination of educational assessments. Accordingly, the National Educational Assessment and Examinations Agency (NEAEA), established independently as an organ with the Council of Ministers Regulation No.260/2012 and one of the main objectives of the agency is to conduct the educational assessment that includes both national and international/regional large-scale assessments.

Ethiopia launched the first national baseline learning assessment in 2000 with the initiation of the Ministry of Education to evaluate the status of student learning before the ongoing reform in the education system (NOE, 2004). At that moment, the then National Organization for Examinations and the Basic Education System Overhaul project played a significant role in the setup of national learning assessment systems in Ethiopia. Still, the program is mostly operating with aid from various donors. For about the last two decades, the national learning assessment was continuously conducted on grade four and eight pupils' in Mathematics, English, Mother Tongue, and Environmental Science subjects for six rounds every four years. Similarly, in grade 10 and 12 students, it was conducted three times in Mathematics, English, Biology, Chemistry, and Physics subjects from 2010 to 2018. The main objective of the national learning assessment was to determine the students' academic achievement in the key academic subject areas and find out home and school factors related to the students' performance (NEAEA, 2012).

In addition, Early Grade Reading Assessment (EGRA) and Early Grade Mathematics Assessment (EGMA) that conducted on grade two and three children were part of the national large-scale as of 2010 to 2014. The focus of these assessments was to assess children's early literacy and numeracy skills. Furthermore, currently, to check the readiness of the children for primary school, the project known as Measuring Education Learning Quality Outcomes (MELQO) that focuses on pre-primary education is in the pipeline after its pilot study had taken place in 2018. In this regard, in Ethiopia, the practice of large-scale national assessment is well established through pre-primary to preparatory education. As the structure of education in Ethiopia comprises four stages: kindergarten or preprimary education, first cycle primary education (grade 1 to 4), second cycle primary education (grade 5 to 8), general secondary education (grade 9 and 10), and preparatory education (grade 11 and 12) according to the education and training policy of Ethiopia in 1994. Because, monitoring and evaluating learning outcomes through assessments and analysis of data is critically vital to guiding the country's quality improvement of education policies and strategies (Joshi & Verspoor, 2013).

On the other hand, though, the system of large-scale national assessment is promising and contributing a lot to the improvement of the quality education in Ethiopia, the country has not yet participated in any ILSA assessments. However, several studies indicated the benefits of participating in international assessments from diverse perspectives (Rutkowski & Prusinski, 2011; Wagemaker, 2014; Chmielewski & Dhuey, 2016). Therefore, although Ethiopia has not participated in any international assessments, it is crucial to assess the academic achievement of Ethiopian primary school students, using the instruments from ILSA studies to look at what the performance of the learners looks like and what factors are related to their achievement. Since TIMSS has released some of the mathematics achievement test items, together with the contextual questionnaires, these instruments can be used to gather the information that will be valuable for policy-makers' and stakeholders to make decisions to participate in such international assessments.

1.2. Statement of the Problem

Assessment is the best way to monitor the quality and equity of the education system to give feedback to all concerned stakeholders. It is a decisive tool to know whether an education system is producing the expected outcomes for the pupils, the economy, and society at large (Clarke, 2012). The assessment seeks to determine how well students are learning, in addition, to provide feedback to students, educators, parents, policy-makers, and the public about the success of educational services (Pellegrino, Chomsky & Glaser, 2001). In other words, it is an instrument used to inform policy-makers, help teachers to improve their teaching practices, and empower parents with information about how well their children are taught. So, the emphasis on education quality for all and learning outcomes has led to increased interest in and demand for national, regional, and international large-scale learning assessments. Appiah (2011) indicated that governments, educators, and parents worldwide are increasingly becoming interested in international comparisons of student achievement.

International assessments like TIMSS are essential for policy-makers, practitioners, and other stakeholders to gain adequate knowledge about the educational system and such knowledge can be used in discussion, planning, and improvement of the situation in schools. The number of participant countries thus arises from time to time and become part of the global scene. Currently, the number of developing countries participating in international assessment has increased as compared to previous studies (Lockheed & Wagemaker, 2013). In TIMSS mathematics assessment, some Asian countries consistently outperformed, while others strived to keep up their standing. For example, African countries such as South Africa, Ghana, Tunisia, Morocco, and Botswana have participated in TIMSS from 1995 to 2015, though; their performances were abysmal as compared with the other countries. Ndlovu and Mji (2012) argued that the poor achievement of students in ILSAs raises questions about the relevance of their curricula and educational reforms in most African countries that participated in the study. In this regard, it is difficult to know what the performance of students and the quality of education in Ethiopia looks like as compared to the other countries since it never participated in any ILSAs. Nevertheless, according to the Fifth Education Sector Development Program (ESDP V), besides national assessments, Ethiopia is planning to take part in regional and international assessments to gather information about the performance of Ethiopian students in comparison with their peers around the world (MOE, 2015). The document also explains that the information obtained from the assessment would be useful for targeting teacher training and make curriculum adjustments, to increase the international competitiveness of learners. Moreover, the involvement of Ethiopia in the international assessment will increase the opportunities to share knowledge and adopt best practices from consistently high-performing countries, or those exhibiting rapid progress in a short period (MOE, 2015).

However, though many countries could gain benefit from their participation in large-scale international assessments, it is important to conduct survey studies that depict to what extent their curricular contexts are aligned with the others and how their students are performing. While international comparative assessment can offer valuable information to obtain the greatest educational and social benefits, however, it requires minimizing its negative consequences on a country's contextual situations (Di Giacomo, Fishbein & Buckley, 2013; Lin, Bumgarner & Chatterji, 2013). Several scholars in the field of education have already contributed to the discussion about international large scale assessments and their impact on global education. While these are important contributions and valuable sources of information that can help to improve the quality and equity of education, Ethiopia is still lacking internationally comparable first-hand information from schools that is collected systematically and scientifically, since it has neither participated in any ILSAs nor conducted significant studies in this area, as already mentioned earlier. To address this knowledge gap, it is imperative to carry out such a kind of research before its participation, as an individual country's decision whether or not to participate in such ILSAs requires a careful evaluation of its suitability.

Therefore, the purpose of this dissertation is to assess the Ethiopian primary school students' Mathematics achievement through the lens of the Trends International Mathematics and Science Study by adapting the TIMSS instruments. That means, it attempts to look at how Ethiopian students perform in TIMSS released mathematics achievement test items overall, compare their performance across different subgroups and compare Ethiopian mean scores with the mean score of African countries participated in TIMSS study and international TIMSS mean score as well as to examine factors that affect the achievement of students' in mathematics.

1.3. Objectives of the Study

The general objective of this dissertation is to assess the mathematics achievement of Ethiopian grade eight students in primary school, measured by the TIMSS released test items. In addition, factors related to the students ' achievement are investigated to understand how much they account for the difference in the achievement of students.

- Find out how Ethiopian students perform in TIMSS released mathematics achievement test items and compare their performance across different subgroups such as gender, school location, language, and regions.
- Compare the achievement mean score of Ethiopian students in TIMSS released test items with that of TIMSS international and African countries mean scores participated in TIMSS.
- Compare the achievement mean score of Ethiopian students in NLA across their achievement in TIMSS mathematics released test items.
- Find out student and school related factors that significantly affect the achievement of students in mathematics.

1.4. Basic Research Questions

- What does the overall achievement of grade eight Ethiopian students in mathematics look like in TIMSS released test items?
- What statistically significant mean differences observed in mathematics achievement across various subgroups (gender, location, language, and region) in Ethiopia?

- Compared to the international TIMSS mean score, what does the achievement of Ethiopian students in released mathematics test items of TIMSS look like?
- What is the difference between Ethiopian students' achievement mean score in TIMSS mathematics released test items and the international TIMSS mean score?
- What does the achievement of Ethiopian students in TIMSS mathematics released test items look like as compared to the achievement of African countries that participated in TIMSS?
- What does the achievement of Ethiopian students in TIMSS mathematics released items look like as compared to achievement in National Learning Assessment?
- What are the major students and school related variables that significantly affect the achievement of students' in mathematics?

1.5. Significance of the Study

Large-scale international assessments are a powerful tool for researching global trends and changing or improving the education systems. Therefore, this study enables policy-makers, educators, and all concerned stakeholders to have reliable information concerning the achievement of Ethiopian students in mathematics in relation to TIMSS released test items and factors that affect students' achievement in mathematics. Specifically, the study has the following main significance:

- Help to give information for policy-makers and educators on how well the Ethiopian students are performing in TIMSS mathematics released test items.
- It helps policy-makers, stakeholders, and educational actors to recognize the main factors that influence the achievement of students in mathematics, focus their

collective attention, and create a shared academic culture dedicated to assuring and improving the quality of education for all.

- Help to identify factors that can reduce educational inequality and promote equity. So that students with disadvantaged backgrounds can get support from schools and to offer these students a life chance to manage better schooling.
- Schools and students got an experience to deal with large-scale international assessments and evaluate their performance since some copies of the mathematics achievement tests were provided to them after the completion of the data collection process.
- Used to increase the scope of knowledge in this area for those who want to study further and initiate other researchers to give attention to this study area.

1.6. Delimitation of the Study

Most of the time, TIMSS assessment is conducted at the primary level on four- and eight-years schooling or at the fourth and eighth graders. However, this study focused only on grade eight students, because of the existence of many instructional mediums at fourth grade than eighth in Ethiopia that is challenging to translate the instruments into the students' language of instruction due to lack of finance. Similarly, though TIMSS focuses on assessing Mathematics and Sciences achievements, this study is limited to only assessing student's mathematics achievement. It is because of that mathematics notations and symbols are universal that it makes the question of test translation easier and feasible than sciences. Countries had similar curricula and approaches in teaching mathematics so that there was a good chance that a mathematics test could develop in such a way that would be an acceptable match to countries' curricula (Loveless, 2007). In addition, it is because of the researcher's educational background and long experience in the mathematics field of study

1.7. Limitation of the Study

This study overall used 62 multiple-choice responses and constructed- response items of released TIMSS Mathematics achievement test 2013, by the IEA. Therefore, the length of the test was not the same as the actual test items of IEA as some test items reserved for equating when the next TIMSS assessment will be taken place. As a result, the number of items across each content domain was very small as compared to the main international assessment of TIMSS. Moreover, there were only eight constructed response items included in the test, as most of the items were not aligned with the content of the Ethiopian mathematics syllabus and difficult to correct them in such a large sample size due to financial constraints.

Also, the contextual questionnaires were only limited to students and principals. It did not incorporate the questionnaires for the mathematics teachers, since students from different sections could be taught by various teachers, which is technically challenging to manage the analysis of the collected data. Private schools were also not part of the study, since they usually resisted participating in such kinds of assessments, though, the official letter of cooperation was written to them according to the researcher’s previous experience. The reason was that they often assumed such a kind of assessment as an evaluation tool, which may have an impact on their business.

CHAPTER TWO: REVIEW OF LITERATURE

2.1. An Overview of Assessment

Developed and developing countries alike understand that providing equitable quality education for all students is essential not only to their economic growth and social stability, but also to maintain peaceful and equitable nations in the global community. The term assessment used in education refers to any procedure or activity that designed to obtain information about the knowledge, attitudes, or skills of a learner or group of learners (Allen, 2004). That is, assessment is the process of collecting and sharing information from multiple and different sources to build a clear understanding of what students learn, understand, and can do about their knowledge, as a result of their educational experiences. According to Nelson and Dawson, (2014), in education, assessment is relatively a recent phenomenon that came into use in an educational context, where different views, strategies, and concerns over assessment have been proliferating.

Assessment commonly focuses on the individual learner, the learning atmosphere, the school, or the whole education system. Walvoord (2004) describes assessment as the comprehensive way of gathering information on student learning, using the available time, knowledge, skills, and resources to inform decision-makers on how to improve instructional processes. Furthermore, the author also clarified assessment as a way of making decisions on instruction, pedagogy, staffing, guidance, and counseling as well as student support on the best possible data, concerning student learning and the factors that influence it. Similarly, assessment, according to Palomba and Banta (1999), is the systematic gathering, analysis, and use of evidence about educational programs undertaken to enhance the mental and physical development of the students. The authors clarified the meaning of assessment, as it has been driven by the questions it seeks to answer about what students should know, what the institution contributes to student growth, how learning can improve. Assessment is any systematic method of collecting evidence from questioning to draw inferences about students' understanding, attitudes, and other attributes for a particular reason, such as diagnosis, feedback, needs, reaction, and evaluative purposes (Shepherd & Godwin, 2004). Assessment is the process by which we ascertain through data collection, if students have learned the skills, content, and habits of mind that will make them successful and decide to adjust the curriculum or teaching strategy to improve learning (Dwyer, 2008).

An assessment system is a collection of policies, frameworks, procedures, and techniques for producing and using evidence for student learning and achievement. Thus, the information derived from assessments is used for a great variety of purposes. According to Nelson and Dawson (2014), the intent of educational assessment activities depends on the practitioners' and researchers' theoretical context, their assumptions and perceptions about the essence of the human mind, the roots of knowledge, and the process of learning. Effective assessment is useful in delivering feedback with appropriate quality and quantity to meet the need of stakeholder with evidence in decision-making needs in favour of enhancing the quality of education and learning outcomes for students (Ravela, Arregui, Valverde, Wolfe, Ferrer, Rizo, Aylwin & Wolff, 2009). According to Kellaghan (2004), assessment is essential to make educational decisions about students; offer feedback on their improvements; inspire students by setting targets; evaluate educational effectiveness and curriculum appropriateness; diagnose education system achievements; examine school effectiveness; assess student outcomes over time and direct policy formulation and decision-making. In general, assessment is a systematic approach involving the use of empirical evidence on student teaching and learning process to improve the programs and maximize the learning performance of the students.

2.2. Large-scale International Assessment in Mathematics

Mathematics is a core subject that is taught at every grade level with a great emphasis worldwide, as it is a key subject to scientific and technological developments of the country. In this regard, the National Council of Teachers of Mathematics (NCTM) identifies the use of mathematics in the following four areas:

> Mathematics for life which enables people to engage in such activities as purchasing, budgeting, or choosing insurance, health or retirement plans,
> Mathematics as a cultural heritage in which citizens can appreciate the beauty of mathematics,
> Mathematics at the workplace that enables people to apply problem-solving skills in solving workplace problems, and
> Mathematics for the scientific and technical community that enables citizens to pursue mathematics and technology-oriented fields such as computer science and engineering. (NCTM, 2000, p. 4).

The fact that skills in mathematics are so critical to economic progress in a technologically based society has led countries to seek information about what their students know and can do in mathematics. In addition, Beaton, Mullis, Martin, Michael, Gonzalez, Kelly, and Smith (1996) suggested that countries are seeking to advance their knowledge of what can be done to enhance students' understanding of mathematical principles, their potential to solve problems and their attitudes to learning. Thus, to make sensible decisions and participate effectively in a world transformed by the ability to exchange all types of information, assessing students' performance in several core subject areas, specially in mathematics, is vital.

Assessment in mathematics can be expressed as a research inquiry into the structure and content of mathematics curricula and the teaching and learning of mathematics. The goal of assessment in mathematics is to improve the quality of mathematics curricula and the quality of the teaching and learning of mathematics (Singh & Ellerton, 2013). Thus, to achieve such a goal, a robust assessment study is needed. These studies are usually large-scale assessments, which reach across national boundaries to achieve the common goal of generating scientifically validated knowledge in mathematics education that is independent of geographical and cultural boundaries. Nowadays, the idea of assessing performance in mathematics education has gained new momentum almost everywhere since computer technologies opened up substantial new possibilities.

There are various types of comparative studies in mathematics education. Some of them are International Commission on Mathematical Instruction (ICMI) 1908-1915, FIMS 1960s, Second International Mathematics Study (SIMS) in the late 1970s, and Third International Mathematics and Science Study 1990s, later Trends in Mathematics and Science Studies 1995 to date and PISA can be mentioned as examples (Singh & Ellerton, 2013).

2.3. Reasons for International Assessment in Mathematics

Scholars have different views about the reasons why countries participate in international educational assessments. Several claims that they are inspired to participate in large-scale assessment due to external pressure, while others argue that the demands and opportunities are more internal that arise from national politics and ambitions (Addey & Camilla, 2015).

According to Newton (2007), assessment is useful for numerous purposes, like providing student scores, national accountability, program mentoring and evaluation, distributing resources, selecting or placing students, articulating strategies, improving teaching and learning, or giving students individual feedback to their families. International large scale assessments are frequently used to inform systems in education. That means it helps in monitoring the systems and evaluate the education programs. Over the last few decades, international large-scale assessments have become increasingly relevant and an invaluable resource for advancement, accountability, educational planning, and policy formulation. Evidence from such large-scale assessments gives a good picture of the degree to which learning achievements are gained and what factors can influence them. School administrators, teachers, policy-makers, and parents can use assessment results to recognize some features of their school systems and home environments that are functioning well and require some improvement. Moreover, according to Lockheed (2016), large-scale international assessments are recently important for evaluating goal four targets set out at the end of primary and lower secondary education levels as shown in the Sustainable Development Goals (SDG). International assessments are primarily used for measuring and evaluating the consistency of student learning achievement (Postlethwaite & Kellaghan, 2008) and designed to make a trend comparison across time (Greaney & Kellaghan, 2008). In addition, capacity building of technical working groups in the area of measurement and assessment is one of the factors currently motivating countries, particularly, developing countries to take up participation in ILSAs (Lockheed, 2011). Thus, international large-scale assessment studies are mostly used to monitor student achievement by taking the importance of information into account on students' background and the context of learning both in and out of school.

Notably, participation in an international assessment of mathematics offers numerous benefits for both developed and developing countries. In many jurisdictions around the world, students are assessed in mathematics using international assessments. According to Suurtamm, Thompson, Kim, Moreno, Sayac, Schukajlow, Silver, Ufer, and Vos (2016), for accountability purposes, ILSAs such as mathematics are useful for tracking education systems and are playing a significantly increasing role in students' and teachers' lives concerning instructional processes. International large scale assessments are used to provide evidence that enables educators to examine the effects of the tasks, discourse, and learning environment on students' mathematical knowledge, skills, and dispositions and opportunities for all students to demonstrate their mathematical learning and be responsive to the diversity of learners (Klinger, McDivitt, Howard, Munoz, Rogers & Wylie, 2015). International studies such as TIMSS provide an opportunity to compare the performance of students not only in countries that are similar in language and culture but also those that have different contexts. For example, the continued success of some Asian countries in mathematics has caused some other countries to seek out the possible factors that may have led to their success. Australian has been used international large-scale assessment to monitor achievement differences between indigenous and non-indigenous students (Ford, 2013). Moreover, the reason for several countries decided to participate in ILSAs is to benefit from the information generated in which the demand for evidence could be driven both by the concerns of educational scholars and by macroeconomists' empirical desires for credible and accurate cross-national data of educational outcomes for the economic development (Hanushek & Woessmann, 2012; Lockheed, 2013). When describing the longstanding involvement of the Russian Federation in international large-scale assessment, Kovaleva (2011) reported that assessments such as TIMSS have played a role in capacity building and developing educational standards in Russia.

Thus, the main significant benefits arising from the use of international large-scale assessment of mathematics data are comparative and benchmarking purposes; improving country's overall educational system through directive policy; enhancing access and equity; improving teaching and learning practice; curriculum reforms; and utilizing strategies and indicators to monitor and evaluate educational processes (Torney-Purta & Amadeo, 2013). Overall, ILSAs are used to monitor the functioning, productivity, and equity of education systems; provide knowledge of factors that determine educational effectiveness, and a reliable, sustainable, comparative database that allows researchers worldwide to study policy-oriented questions.

2.4. Limitation of International Assessment in Mathematics

Although ILSAs can offer significant information for countries concerning comparison, there are cases where their participation is yet overlooked for different reasons. ILSAs are important to address the primary concerns of stakeholders in education. However, the diversity in education systems and differences in the priorities that different stakeholders place on distinguishable objectives could reduce the impact of assessment as a changing agent (Ritzen, 2013). According to Bloem (2013) and Lockheed (2010), countries may not take part in ILSAs for several reasons related to the challenges of culture, capacity, and costs.

According to Torney-Purta and Amadeo (2013), one of the main challenges of international large-scale assessment is its validity, since validity is a special concern when country rankings are compared in league tables. While ILSAs in mathematics offering valuable information, all parties need to be on the same page to maximize valid use of its results carefully according to their contexts to obtain the greatest educational and social benefits to minimize negative consequences (Lin, Bumgarner & Chatterji, 2014). The other problem of large-scale assessment was lack of having personal consequences or for its low-stakes for the test takers, therefore, some test takers might not spend the full effort throughout the test, resulting in an underestimation of their true proficiency levels. Consequently, the estimates of students' ability scores and the implications base on results can be probably biased (List, Robitzsch, Lüdtke, Köller & Nagy, 2017; Eklöf, 2010).

Backhoff (2013) also emphasized that inter-country comparisons in mathematics and other subjects cannot simply be based on rankings, as several factors can cause unexpected changes in countries' measured learning outcomes. A few of the factors that an author mentioned were cross-cultural test items' translation errors, cultural differences in survey response styles, cultural differences in reading the meaning of some questionnaire items, understanding validity issues, limitations in samples of students tested, and the students' opportunity to learn what is tested.

Many barriers prevent countries from using international large-scale assessment studies like TIMSS in formulating or revising their education policies. Some of them are the poor dissemination of information, weak assessment bodies, and fragmented government agencies, political factors such as violent conflict and political unrest, lack of political will and no acceptance of assessment results, discrepancies in findings resulting in a policy stalemate and data manipulation and corruption leading to policy inaction/misdirection (UNESCO, 2017). Critics have claimed that the social and cultural pluralism in educational construction is condensed into one single point in a large-scale international mathematics assessment, in which its achievement results are subject to methodological constraints (Strietholt & Scherer, 2018; Rutkowski & Rutkowski, 2010).

In particular, many developing countries are not or may not be interested in getting involved in cross-border assessments which they consider as they are too complicated for their students and thus do not provide sufficient information on the learning circumstances in their countries (UNESCO, 2018). Bloem (2013) indicated, there is a question of why lower and middle-income countries to decide to join international assessments, as the data from such assessments may not in their contexts. In this regard, Lockheed (2015) revealed that the participation rates for high-income and upper-middle-income countries are many times higher than participation rates for lower middle-income and low-income countries.

According to Lockheed (2013), high-income countries that participated in ILSAs look inside their education system and compare with the others to improve their education quality. Lockheed further states that while low or middle-income countries are encouraged to participate in ILSAs for improving their national assessments and monitoring their education systems on a common international scale. Furthermore, the donor community requires relevant information and quality data to inform their decisions on how best to support low-income countries to improve the learning outcomes of their children. It is, therefore, indispensable to provide the evidence desired by the international community to recognize the significance of advocating for and assisting countries to develop and conduct national and ILSAs.

2.5. Participation of Countries in TIMSS and Their Achievement

The participation of countries in IEA studies in general, increased from the 12 countries that participated in the FIMS assessment to 20 systems for the SIMS, 40 systems for TIMSS 1999 to 66 systems for TIMSS 2007 to 79 participants for the TIMSS 2011, though, the number of participant countries reduced to 57 for the TIMSS 2015 assessment. Particularly, with regard to grade eight, in 1995 about fifty-two, in 1999 fifty-five, in 2003 fifty-one, in 2007 fifty-seven, in 2011 fifty-nine and in 2015 forty-seven countries/systems participated in TIMSS studies. In all cases, the number of participating countries includes several benchmarking states in the United States, Canadian provinces, special administrative regions such as Hong Kong, and other subnational administrations such as one of the United Arab Emirates nations in Dubai. For the list of participant countries in eighth grade TIMSS from 1995 to 2015 (See Appendix I).

The participants in TIMSS are from all continents of the world. The majority of them were from European and Asian countries and they continuously participated almost in every cycle of the study. Few of them were from Latin American, Middle East, and African countries including various US states. Lockheed (2011) states that the decision of a country to participate in one ILSA is closely connected with its determination to partake in a subsequent large-scale assessment. According to Lockheed (2011), 76% of 17 developing countries participating in PISA 2006 were due to their previous participation in at least one of the IEA study such as TIMSS or PIRLS. However, the number of participant countries in ILSAs can be reduced for various reasons. According to Addey and Sellar (2019), countries may decide not to participate in ILSAs for ideological reasons that not being politically or economically affiliated with certain groups of countries; for cultural and educational reasons, where some program considered poorly suited to assess the diversity of learning contexts; due to a lack of resources and capacity to conduct the study; to focus on national or regional assessments and due to their concerns about poor performance being compared to those with better resourced education systems. Particularly that is why the participation of low-income countries including African countries, less as compared to the other developed countries.

Regarding the achievement of countries that participated in TIMSS, almost in all cycles from 1995 to 2015 East Asian countries outperformed other countries. In this case, for the last six assessment cycles all the ranks first to fourth occupied by those countries. For example, Singapore ranked first four times out of six cycles, while Taiwan and South Korea ranked first in 2007 and 2011, respectively.

On the other hand, most Middle East and African countries that participated in the TIMSS assessment were poorly performed. According to Bofah and Hannulathe (2015), the performance of students from the five African countries participating in TIMSS 2011 was amongst the very lowest, which recorded an average mathematics achievement below the TIMSS center-point of 500, particularly except Tunisia others falling below the low (400) International Benchmark score. Looking at the last six cycle's performances of countries in mathematics TIMSS, South Africa was the least performed country in the three cycles (1995, 1999 and 2003). In 2007, 2011 and 2015, Qatar, Ghana and Kuwait were the least performed countries respectively. For the detailed, list of ten high and ten low achieving countries in Grade 8 Mathematics since 1995-2015 look at Appendix II.

2.6. The Impact of TIMSS Assessment

The growth of international assessments and demand for participation in international assessments is evidence for the impact of ILSAs. In other words, the comparison of students' achievement in international large-scale assessment has had an increased impact over the last decades as countries increasingly compete in global markets. According to Isaacs, Zara, Herbert, Coombs, and Smith (2013), most countries' political leaders believe that educational achievement is the key to economic success. Therefore, the rank ordering of countries based on their students' outcomes on achievement tests can have wide effects on the education system. In many countries, the results and rankings of international large-scale assessment become politically high-stakes, so that the ministers of education and governments such as Poland, Germany and Israel perceived it as being an indicator of the success of their local education policy Beller (2013). Suurtamm et.al (2016) indicated that, though, ILSAs are used in a variety of ways, they have significant impacts on changing policy, curriculum, and instructional practices. Some of the impacts of the results from ILSAs such as TIMSS on educational systems are as follows.

2.6.1. Educational Policy Reform

Large-scale international assessments have become famous as 'whips' that used to inspire countries to take policy action in order to advance their education system (Hanushek & Woessmann, 2009; Beatty & Pritchett, 2012). There is also evidence that many countries use TIMSS results as an important means of informing their respective educational reforms and strategies for change. In Germany, for example, where involvement in ILSAs is mandated by law, Lehmann (2011) stated international large-scale assessments such as

TIMSS as “a necessary condition for policy-making”. As a result, Germany established the SINUS (increasing science and mathematics instruction efficiency) -Transfer program as a direct consequence of its unpleasant with mathematics and science results on TIMSS 1995 assessment. By 2002, the US Congress passed two major revisions in US educational legislation: *The No Child Left behind Act and the Education Sciences Reform Act*, both of which coincided with the effects of TIMSS and rendered special cases under the law to ensure continued national and state involvement in large-scale TIMSS assessment. According to Abdul-Hamid, Abu-Lebdeh, and Patrinos (2011), international assessments enabled comparisons of the educational performance of Jordan students with that of students in other countries and allowed them to benefit from other country's experiences in the design of successful education policy. International assessment greatly influenced the development of the educational system and teacher capacity building materials and used by funding organizations to monitor and evaluate the projects and programs in Jordan's education sector (Lietz & Tobin, 2016).

Furthermore, TIMSS results led to policy changes in the existing preservice and in-service teacher training programs as well as the development of instructional resources in some countries. For instance, in Canada (Ontario), instructional materials developed as a direct effect of TIMSS (Taylor, 2000). New Zealand was able to generate training resources and designed continuous professional development programs for mathematics and science teachers to resolve certain shortcomings in the areas (Chamberlain & Aalst, 2000). Based on the findings of TIMSS, Israel also focused on the teacher preparation programs rather than in-service training programs (Mevavech, 2000). Throughout Malaysia, the government has given more attention to teaching methods and teacher preparation after the repeated TIMSS study had taken place (Elley 2005). According to Cariola, Covacevich, Gubler, Lagos, and Ortiz (2011), changes to the teacher education system have been made in Chile following the dissemination of the TIMSS results for 1995.

Moreover, the government of the Philippines also initiated to create a program to train 100,000 teachers, while teacher guidance materials were prepared in England because of the TIMSS study report (Robitaille, 2000 & Wagemaker, 2010).

2.6.2. Curriculum Reform

International learning assessments also have a significant impact on curriculum changes in various countries. The analyses of international large-scale assessment data have influenced curricular and instructional reforms in many countries (Aggarwala, 2004; Mullis & Martin, 2007; Wagemaker, 2011). Similarly as mentioned earlier, in its analysis of the effect on international evaluation, Wagemaker (2011) noted that various countries such as Iceland, Kuwait, New Zealand, Norway, Romania, and South Africa have served TIMSS as input for curriculum review and improvement. Romania and the Slovak Republic also added new topics in mathematics and science curricula because of TIMSS (Noveanu & Noveanu, 2000; Berova & Matusova, 2000).

Pelgrum, Blahova, and Dukynaite (2011) state that, in Latvia, analysis of TIMSS 1995 findings identified gaps in skills related to the application of mathematical knowledge in real-world situations, which eventually led to curricular change. In Kuwait, the principal goal of participation in TIMSS was to evaluate the standard of the curriculum with reference to international benchmarks in addition to using to revise the examination regulations. In Slovenia, Klemencic (2010) describes the impact that TIMSS 1995 results had on the syllabus plans for students in their ninth year of schooling.

In Chile, according to Cariola et al. (2011), the TIMSS 1999 results identified a curricular imbalance in the teaching of mathematics and science and promoted curricular reform and the introduction of content standards in 2009. In Macedonia, due to the results of TIMSS repeated, changes have made in the curricula, national assessments, and in-service training (Elley, 2005). Similarly, Singapore consistently analyzes its TIMSS results to identify specific student strengths and limitations and is used as one of the benchmarks that inform national curriculum reviews (Wagemaker, 2014). After perceiving poor learning outcomes in TIMSS, New Zealand's Ministry of Education produced a series of books that enhance teacher’s understanding and their teaching methods in basic science concepts of primary education (Jones & Bunting, 2013). According to Bolotov, Kovaleva, Pinskaya, and Valdman (2013) and Tyumeneva (2013), Russia sustained to engage in international large-scale assessments to increase policy-makers' awareness about the significance of social and school contexts in learning for students, that were eventually leading them to change curriculum and performance standards at both primary and secondary education levels.

2.6.3. Instructional Reforms

Large-scale international assessments can also influence the instructional process of schools in various countries. According to Wagemaker (2011), Iceland has increased the teaching hours for mathematics and science instruction at the elementary level based on the findings obtained from the TIMSS study. TIMSS results contributed to the development of new educational standards in mathematics and Science in Russia. Consequently, it began to use standardized tests for assessing student achievement (Kovalyova, 2000). Correspondingly, Iran utilized the TIMSS assessment framework to construct their test items, in order to improve the system of assessment in the country (Heyneman & Lee, 2014). In Macedonia, because of TIMSS changes have been made in in-service training, which encouraged the teachers to change their method of teaching from a teacher-centered to a more student-centered style (Elley, 2005).

In order to advance the quality of their test development process, the Czech Republic used TIMSS as a model and practical reference guide (Strakova, Paleckova, & Tomasek, 2000). Nakajima (2010), reported that Japan's used TIMSS results in increasing the number of annual instructional days for schools and the amount of time given over to instruction which made it a regularly strong performer on international large assessment. Robitaille (2000) and Wagemaker (2010) also revealed that as the result of TIMSS 1995 findings spotlighted issues associated with teaching in Australia, Canada, Spain, and Japan.

In some developing countries, policy-makers began to model teaching practices of developed countries after the result of TIMSS studies. According to Elley (2005), eight among eighteen developing countries that take part in TIMSS-Repeat had altered the way of classroom learning method to focus on activity-based learning, problem-solving, reasoning, and critical thinking. The author also reported that, in two countries, policy-makers reduced the amount of teacher-centered learning and increase student engagement in lessons, with more debate, questioning, investigating, critical thinking, valuing of student attitude, and exploring students' curiosity.

Moreover, ILSAs are used to build the capacity of educational experts and provide reliable evidence for global donors. According to Lockheed (2010) and Greaney and Kellaghan (2008), the number of countries participating in ILSAs could increase as the worldwide technical capacity for undertaking large-scale assessments increased. That means participation in an international large-scale assessment appears to build the assessment capacity. For example, the share of countries participating in PISA after having previously participated in TIMSS assessment is higher than that of the countries that had not participated in TIMSS previously (Lockheed, 2015). The author further states that the longer a country has carried out assessments, the more likely for it to develop high expertise in assessment over time. Howie (2011) notes that South Africa had very few objective data on educational learning outcomes before engaging in ILSAs such as TIMSS. Getting involved in such assessments, however, has helped South Africa to build local research capacities. Thus, participation in international large-scale assessment is beneficial to build the capacity for the countries to conduct their national or regional assessments that demand high administrative and technical qualities.

2.7. African Countries Experience in Regional LSA

In Africa, large-scale assessment system is rapidly becoming a key priority for improving the quality of education. The focus on measuring learning outcomes as an indicator of education quality at the 1990 Education for All (EFA) world conference in Jomtien and its subsequent reinforcement at the 2000 EFA world conference in Dakar (UNESCO, 2000), had initiated several African countries to establish national and/or cross-national assessment systems. That means, in addition to partaking in international studies conducted by the IEA and OECD, African counties have been designed and implemented various assessment programs that have been focused on continental or regional contexts.

Most Africa countries are participating in regional assessment studies facilitated by UNESCO/UNICEF including the Southern and Eastern African Consortium for Monitoring Educational Quality (SACMEQ), Programme d'Analyse des Systèmes Educatifs de la CONFEMEN (PASEC) and Monitoring Learning Achievement (MLA) (Kellaghan & Greaney, 2004). According to Braun and Kanjee (2006) and Kellaghan and Greaney (2004), about 15 countries have participated in SACMEQ studies, 12 in PASEC studies, and 47 in MLA I and II studies. Furthermore, according to UNESCO (2018), about 40% of sub-Saharan African countries have engaged in regional and/or international learning assessment programs, where at least half of them have conducted a national level learning assessment.

SACMEQ is a type of assessment in Anglophone countries in Southern and Eastern Africa that work in collaboration with the International Institute for Educational Planning (IIEP) in Paris which is conducted on sixth grade learners. It has undertaken four data collections since 1995, with the latest assessment in 2012–2014 where 15 countries were involved. The participant countries include Angola, Botswana, Kenya, Lesotho, Malawi, Mauritius, Mozambique, Namibia, Seychelles, South Africa, Swaziland, Tanzania, Uganda, Zambia, and Zimbabwe. In addition to assessing the performance of students in literacy and numeracy, to examines the ways how the performance is related to contextual factors, questionnaires are administered to students, teachers, and school principals (Hungi, Makuwa, Ross, Saito, Dolata, Cappelle & Vellien, 2011).

Programme d'Analyse des Systèmes Éducatifs de la CONFEMEN (PASEC) was established in 1991 as a response to the Jomtien conference, which acts as a network for sharing information on educational assessment results among Francophone countries. The participant countries include Burkina Faso, Cameroon, Congo (Brazzaville), Côte d'Ivoire, Djibouti, Guinea, Madagascar, Mali, Niger, Central Africa, Senegal, Togo (Kulpoo & Coustère, 1999). Since 2014, the program has moved towards its first cycle of competency-based international assessment. This approach allows a comparison of countries across grade levels and cycles over time. Since it is conducting at the beginning of grade two and at its end of grade six PASEC (2015). PASEC relates student performance in French and Mathematics to students' background factors, conditions in schools and classrooms as well as the political context of the participant countries.

Monitoring Learning Achievement (MLA) was another project conducted with the support of UNESCO in 1992 to monitor the extent to which students actually acquire useful knowledge, reasoning ability, skills, and values according to the World Declaration on Education for All in 1990 (Chinapah et al., 2000). That means, the focus of the assessments were literacy, numeracy, and life skills. It consisted MLA I (for grade 4) and MLA II (for grade 8) students. Moreover, the data on students' backgrounds, school characteristics, and family backgrounds were collected to find out their impacts on students' performance. MLA was one of the international learning assessment programs, which was typical for many African countries since out of 70 countries 47 of them were from Africa (UNESCO, 2003a).

The achievement of African countries in these regional assessments was inconsistent. According to UNESCO (2000a), the results of SACMEQ assessments indicated, as education systems were failing to meet the performance standards of the countries where less than 30% of sixth grade pupils met specified minimum literacy standards in Namibia and Zambia and over 50% met minimum standards in Zimbabwe, and the figure did not change over time. In this case, there was a significant achievement difference found between regions, and between types of school within countries, though gender differences were not significant (Mioko, 1998). However, currently, the performance of most participant countries significantly improved through time. For example, South African grade six learners were achieved above the 500 centre point of the average score for the first time in SACMEQ IV both in the literacy and numeracy Department of Basic Education (2017). Overall, SACMEQ IV study results indicated that, the highest improvement margins in each participating countries in the region.

With regard to PASEC, in most participant countries the achievement of the learners deteriorated over time. According to PASEC 2014 results, more than 70% of students at the beginning of primary education (2nd grade) have not attained the expected skill level in reading/writing, and more than 50% have not attained the expected skill level in mathematics. At the end of primary education (6th grade), nearly 60% of students are below the expected skill level in both subjects (PASEC, 2015). Furthermore, when PASEC results transformed to the international scale, 71 percent of second graders do not have sufficient French language skills and about 59 percent of six graders lack essential mathematics competency (Diop, 2015). In MLA study also, only four of the participant countries had met the Jomtien learning target (i.e., 80% of learners should attain the defined learning competencies) for fourth grade pupils in life skills; two in literacy and none in numeracy (Chinapah et al., 2000).

Regional large-scale assessment in Africa same as ILSA played a critical role in enhancing the learning and teaching process. Some of the impacts from the regional large-scale assessment linked to the development planning, policy review, and educational reform processes. For example, based on the result of SACMEQ Kenya, Malawi, Namibia, and Zambia had made significant educational reforms, Zanzibar reviewed school sizes by split excessively large schools into smaller ones, Namibia improved its system of resource allocation, Zimbabwe used to revise its curriculum, Seychelles reviewed the policy on streaming education, Botswana used to build its capacity for the production of an accurate and up-to-date EMIS, Malawi used to generate policies concerning the provision of teaching materials and classroom furniture, Zimbabwe and Kenya used to improve the provision of inputs to schools, Uganda used the result to develop the eighth Education Sector Review, and so forth. (Dolata, Ikeda & Murimba, 2004; Leste, 2005; Murimba, 2005).

Similarly from the findings of PASEC study, Madagascar used to revise the time for teacher on continuous professional development training while in Cameroon, as a result of the first PASEC assessment, the ministry created a special evaluation unit to ensure the institutionalization of educational evaluation (Bernard & Michaelowa, n.d.). Togo and Guinea based on the result of PASEC, improved the mechanism of recruiting teachers for the education system (Bernard & Michaelowa, n.d.). In general, though, PISA, TIMSS, and

PIRLS have influenced global education policy, the value of a regional assessment in Africa goes up with the number and diversity of participating countries, since it provides reliable evidence in learning performance for the education systems with different institutional features and policies. Therefore, as Fifty-two countries in Africa have now unified behind the African Continental Free Trade Agreement to promote regional integration, a parallel effort to unite the region in a single, Africa-wide learning assessment that builds upon the capacity and experience of PASEC and SACMEQ, and extends participation to all countries in the region would powerfully useful to advance the quality of education.

2.8. Ethiopia- General Context

Ethiopia is a landlocked country located in the East Africa known as "the Horn of Africa". The country shares its borders with seven neighbouring countries: to the North Eritrea, in northeast Djibouti and Somaliland, to the East Somalia, to the southern Kenya, in west South Sudan, and in the northwest Sudan. A country has an area of 1.1 million square kilometers which is nearly three times the size of Germany. It has an estimated total population of over 110 million and the second populous country in Africa next to Nigeria. Different nations and nationalities having diversified cultures and more than 80 spoken languages inhabit the country. Of these languages, more than 30 are currently used as media of instruction in schools. There are two dominant religions, Christianity and Islam, which have a long history and many followers in the country.

The country has adopted a federal government system with nine decentralized regional states and two City Administrations. Among the nine regions, four of them have better infrastructures next to the City Administrations, while the rest four regions are considered as ‘emerging’ regions. The Prime Minister is the head of government, where executive power is also exercised by the government. The federal legislative power is vested in the government and the two chambers of parliament.

As it is located in the subtropical zone, Ethiopia possesses varying topography mainly plateaus, highland chains at the center, and low lands at the coastal areas. The climate varies from very hot to very cold temperature and from arid to high rainfall areas following the topography. The wide variations in the land structure, soil type and climate are the main causes of the diversified agricultural products, vegetation, wildlife and mineral resources in Ethiopia.

Concerning the economic situation of the country, the majority of the population's livelihood depends on agriculture. The country is one of the poorest in the world, and about 29.6% of the population lives below the poverty line, according to Central Intelligence Agency in 2014. Traditional agriculture and pastoralism are the main practices of the population in Ethiopia. Cereal grains are the most important food crops cultivated primarily for local consumption. Coffee, cotton, oil seeds, pulses, fruits, vegetables as well as leather and skin are some of the commercial products used for export and internal processing. The mining and industrial sectors are at a lower level of development. Nevertheless, some promising potentials and developments are observed around the production of hydroelectricity, airlines, and other communication sectors.

The current Education and Training Policy was enacted in 1994 to alleviate the problems of access, curriculum relevance, equity, quality, and related challenges of the education sector. As a result of the policy priority and financial input for the sector, Ethiopia made a great effort to improve the quality of education in all aspects, although it is still a crucial challenge for the nation. Education is the key to sustaining Ethiopian development and it is through education that the country can transform into a knowledge-based society embracing new technology and using it to solve the problems of today and tomorrow (MOE, 2009). In this regard, Ethiopia is recognized that the development of the country very much depends on the development of science and technology. Hence, the nation has given due emphasis on mathematics and science education to bolster student learning outcomes in those subjects.

2.9. Mathematics Education in Ethiopia

Mathematics is a discipline where students are provided with the opportunities to obtain knowledge through thinking and reasoning and use their acquired knowledge to make sense of the world around them. It is a tool in which students get the knowledge and experience about life, learn how to deal with problems, apply their knowledge to real-life problems, improve their ability, about logical thinking and reasoning, and they are getting ready for their future (Arslan, Çanli, & Sabo, 2012). Moreover, it is a useful device whereby students actively acquire knowledge through problem-solving in relation to the challenges faced in their day-to-day lives. Consequently, in Ethiopia, the mathematics subject is offered separately from pre-primary through tertiary education. In this regard, the next sections discuss the history of mathematics, its philosophy, curriculum, and the achievement of students in Ethiopia.

2.9.1. History of Mathematics Education at a Glance

Mathematics is an interesting viewpoint of human thought, and its history varies in quintessence to somewhat from other histories. The earlier system of education in East Africa, including Ethiopia, was oral. Its main objective was to ensure that the next generation inherited the culture, customs, and rules of engagement in the society (Halai & Tennant, 2016). Much emphasis has been placed on communal virtues and collective responsibilities. Individuals in the larger group were together responsible for the upbringing of children as well as guiding and orienting them about the morals, values, and norms in their family, society, and culture. These values and norms were transferred through numerous aspects of cultures, such as folk tales, songs, myths, legends, proverbs, and riddles (Halai & Tennant, 2016).

Similarly, the early history of mathematics education in Ethiopia is also either related to the religious sector or the indigenous people settled in rural areas (Hailu, 2016). In this regard, the past and present historical evidence disclosed as people had a perfect calendar different from the westerns, ideas of probability, and designs that comprise geometric, algebraic, or a combination of both and patterns of interest because of the results of the society's practice. For instance, the Obelisks of Axum in the 1700s, the rock-hewn churches at Lalibela in the 800s, and astronomical knowledge of ancient African calendars of Borena in 300 BC were evidence of the early artifacts representing high-quality architectural, technological, and sophisticated mathematical concepts. In line with this notion, Zaslavsky (1993) claimed that mathematical thoughts and practices originated from people's real desires and interests in all cultures, in all parts of the world, and all generations. However, mathematics education in most East African regions, as in different parts of the world, faces many challenges and opportunities related to regional interdependence, internationalization, and advancement in technology and innovation. In reaction to these problems and opportunities, most of the time curricular improvement and changes in mathematics education have been suggested and implemented within the regions (Halai & Tennant, 2016).

In this regard, as with many nations around the world, the modern history of mathematics education in Ethiopia was in one way or another has been related to curriculum materials that are strongly derived from other cultures (Bush, 2005). This means that Ethiopia has undergone many curriculum changes in mathematics education over the past 10 decades due to recurrent cultural influences from other countries. These could be either due to the need to cope up with the changing and competitive world or foreign political influences. For example, as soon as the Italian invasion ended in the early 1930s, the Ethiopian government established a political relationship with the British government and attempted to enhance the national capacity in mathematics education. As a result, it adopted the mathematics curriculum from Britain that entails two series known as *Durell and Hudson* and *High Way Mathematics Series* (Behute, 1991). According to the author, these series of curricula vanished shortly, because they fail to satisfy the needs of the society.

In 1967, "The Entebbe Mathematics Program" was introduced in 10 African countries, Ethiopia, Ghana, Kenya, Liberia, Malawi, Nigeria, Sierra Leone, Tanzania, Uganda, and Zambia (Woodhouse, 1973). This mathematics program was initiated and developed by Mathematicians/experts from the US, which continued until the early 1990s. In this mathematics program, the duty of the learner was to memorize facts and the task of the teachers was to repeat or present the prescribed pieces of information from the syllabus or textbook without understanding and problem-solving skills.

In 1994, the issue of learner-centered education boldly arisen in the Ethiopian Education and Training Policy (MOE, 1994). Moreover, the policy document acknowledged the ethnic diversities in the country and declared their rights to learn in their mother languages. However, still, mathematics education contents remained westerns focused, as Moreira (2007) referred it as mathematics from the "outside world". In this regard, Izmrili (2011) indicated, many Ethiopian textbook authors and curriculum experts reflected and promoted much of the Western Mathematics, which led to the erroneous interpretation of the history of mathematics merely as a progressive story of the isolated successes of the chosen few societies. Nevertheless, numerous researchers (Matang, 2002; Bush, 2005; Rosa & Orey, 2010), indicated as each society has something to contribute to the improvement of mathematics education and has the potential of mathematizing in its own culture at any cost. Therefore, although, the concept of mathematics is almost similar worldwide, it is crucial to adapt to the country's specific contexts during curriculum reform, design and development.

2.9.2. Mathematical Philosophy of Curriculum

According to Ernest (1991), the absolutist and the fallibility are the main philosophies of mathematics that shape the design of the curriculum in mathematics. Absolutists observe mathematics as it is discovered rather than look as it is possible to invent new ideas in it. For this philosophical school, the objects of mathematics have their own independent existence. The organizing principle of the mathematics curriculum for this school is the content or the subject matter than the method of instruction and known as traditional mathematics curriculum (Ernest, 1991). In such a curriculum, teaching is more explanatory and prescribing what to think and not how to think. Learning is more on memorization of facts and principles and recitation of the lesson.

Philosophy has a significant role in determining the goals of education, content, and cognitive domains, learning styles, and experiences. In any discipline, to design a curriculum, considering its philosophy is vital. The philosophy of curriculum presents points of view, which guide the development of the curriculum. For instance, to design a mathematics curriculum, one needs to depend on the philosophy of mathematics in addition to look at the historical foundations as well (Chernet, 2009).

On the contrary, the fallibilist philosophy of mathematics considers mathematics as a human endeavour (Ernest, 1991). According to the author, this school of philosophy believes, as mathematics is imbued with human culture and its objects are not free of revision. Moreover, they rely on that human beings invented mathematics; they will also invent new mathematical results in the future in order to solve their own problem. A mathematics curriculum designed based on this philosophical school helps the learner in liberating from the traditional emphasis on rote learning and lesson recitation (Ernest et al., 2016). In this case, learning is possible as the person actively engages in problem-solving which is transferable to a variety of situations, where teaching and learning are student-centered. This type of mathematics curriculum is known as an interactive curriculum. Here learners are expected to develop good habits of mind that can be transferred to solving different problems in their life.

In Ethiopia, although a student-centered approach was advocated in the curricular documents, the teaching-learning process in primary and secondary education is characterized by teacher domination as indicated in ESDP IV (MOE, 2010). Similarly, Cherinet (2009) shown as the recent mathematics curriculum in Ethiopia can be categorized under outdated curriculum, since it is predominantly organized on the principle of the content domain that needs to be covered for a given grade level in a given academic year rather than the way that the students need to develop knowledge and skills. In line with this, the author indicated as the current mathematics curriculum in Ethiopia cannot be able to produce a competent student in mathematical abilities. Hence, because of inappropriate mathematical curriculum philosophy, many Mathematicians in Ethiopia and other parts of the world have often expressed that, majority of the students do not recognize mathematical concepts, or do not realize why mathematical procedures work or do not know when to apply a given mathematical technique. Therefore, to design a suitable mathematics curriculum too much emphasis on any single philosophy of mathematics that is exclusively on the subject matter or the method is not acceptable. Instead, it is better to look for a middle road, where there is no extreme emphasis on subject matter or methods. That means it should focus on serving the needs of the students and society.

2.9.3. Mathematics Curriculum Contents

The key mathematical concepts taught in Ethiopian primary schools are about relationships of numbers, measurement, shapes, and data handling. Specifically, as indicated in the curriculum framework for Ethiopian education, the mathematical concepts included in the primary school grade eight mathematics curricula are:

> Numbers and operations: the number system introduced and gradually the ability of students to order and perform fundamental operations on numbers is developed.
> Sets: the concept of sets is treated spirally right from early defining to the development of students' abilities in operating with sets and using variables to describe sets.
> Equations and inequalities: the competencies in solving linear equations and inequalities in a given domain developed up to the stage where students are able to solve simple word problems leading to linear equations and inequalities.
> Ratios, proportions, and percentages: students are encouraged and assisted to define and apply the concepts of direct and inverse proportionality, ratios and percentages and apply their knowledge in solving real-life problems.
> Graphs, relations, and functions: the students' ability to collect, present and interpret data in a pictorial way are gradually developed. The notions of relations and functions introduced and gradually developed by applying the concepts learned in solving relevant and real-life problems.
> Measurements: the students' ability in measuring developed right from measuring the length of line segments of the stage where they are able to compute the volumes of various solid objects.
> Geometry: the students define, measure, and categorize some figures such as angles, triangles, quadrilaterals and circles. As a result, they can develop basic skills of geometric construction and identify common solids figures and their properties. (MOE, 2009, P. 16)

At this school-level, students acquire basic mathematical knowledge and develop important mathematical skills and a positive attitude towards the subject so that they are able to learn more; reason rationally, thinks creatively, make correct decisions and be able to solve real-life problems (MOE, 2008). Moreover, learning mathematics develops students' abilities to see mathematics in various situations and choose appropriate mathematics for these situations; think creatively, critically, strategically, and logically; plan, investigate, make conjectures and decide on levels of accuracy; reason inventively, analyze options and consider the consequences and implications of decisions. Overall, the objective of the second cycle of primary education is to offer basic and general education to pupils in order to prepare them for further general education.

2.9.4. Achievement of Students in Mathematics

In the increasingly global technological and innovative world, mathematics is considered as an important gatekeeper for the opportunities of social and economic development. The knowledge and skill of mathematics as a tool for use in everyday life are essential for the existence of any individual and society (Kemboi1, Githua & Changeiywo, 2014). Due to this, mathematics is considered as one of the most critical core subjects in the school curriculum. Thus, it is crucial to ensure that students adequately develop mathematical knowledge, skills, and attitude in order to compete in this global world. In this regard, policy-makers and other key decision-makers in education are looking towards improvement in mathematics, teaching, and learning as a key focus in education reform. For example, in the East African regions, several initiatives such as Strengthening Mathematics and Science Education in Ethiopia (SMASEE) in Ethiopia, Strengthening Mathematics and Science in Secondary Education (SMASSE) in Kenya, Secondary Science Education and Mathematics Teachers (SESEMAT) in Uganda and Tanzania significantly supporting the quality of mathematics education and/or mathematics teacher education in the respective country (Halai & Geoff, 2016).

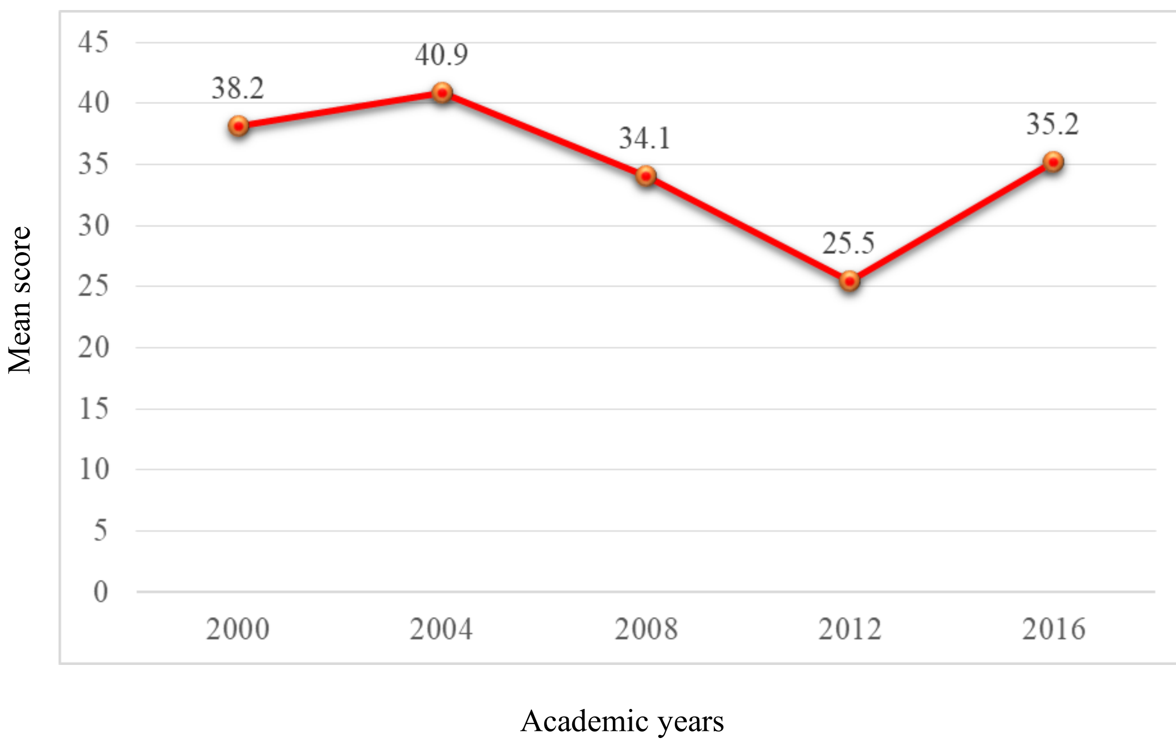
Recognizing the key role education plays in gearing up the development of skilled human power, Ethiopia has been aggressively working to ensure the quality of education. To improve the learning outcomes of students and thereby successfully pursue its educational strategies, Ethiopia has been given due emphasis to concerns about schooling quality. In this regard, it has launched the first phase of the General Education Quality Improvement Program (GEQIP) in 2008 to improve the quality of general education throughout the country. Education policies and strategies aimed to enhance the quality education drawn on the GEQIP and further develop the platform and become an integral part of the Education Sector Planning or Education Sector Development Program. Particularly, Ethiopia has given special attention to mathematics and science education to enhance students' achievement through the project for capacity development for improving learning achievement in mathematics and science education aided by JICA. The project known as SMASEE was started in 2011 with the aim to build the capacity of primary school teachers, particularly grade seven and eight mathematics and science teachers through continuous intensive training.

However, various studies are revealing that students do not perform mathematics to the expected level. The student's low achievement in mathematics is not only a concern for particular countries but has become a worldwide concern over several years (PISA, 2003). For instance, Beatty and Pritchett (2012), argued that progress in learning achievement in many of the developing countries, especially in the areas of literacy, mathematics, and science, was stagnant or even declining. These authors predicted that, at their current rates, Colombia would take 30 years and Turkey would take 194 years to reach the average level of learning achievement for the OECD countries. While countries like Indonesia, Iran, Jordan, Malaysia, Thailand, and Tunisia would never catch up, as learning levels had actually declined from one testing period to the next (Beatty & Pritchett, 2012).

Using the data at the student level obtained from Ethiopia, India, Peru, and Vietnam, Singh (2014) studied the emergence and evolution of learning gaps across four countries and found significant achievement differences among these countries, with Ethiopia at the lowest and Vietnam at the highest level. In this case, the magnitudes of the learning gap across these countries widen as the children grow up with the learners from Ethiopia still at the lower end.

Similarly, the Ethiopian National Learning Assessments (NLA) showed that, the achievement of Ethiopian students' in all subjects and mathematics specifically, below the national standards indicated in the Education and Training Policy. For instance, the NLAs that conducted in Ethiopia every four years since 2000 show the low achievement of students in mathematics subject (NAE, 2000, 2004, 2008, 2012; NEAEA, 2016).

Figure 1. Mathematics Achievement of grade 8 students 2000-2016 in percentage



As shown in figure 1, the average mean scores of eighth grade students' achievement in mathematics at every assessment cycle was below the minimum passing score (50%) set by the ministry of education in the Education and Training Policy (MOE, 1994). In this case, the highest achievement mean score was 40.9% in the year 2004. On the contrary, the least mean score was 25.5% in 2012, which is about the chance level score in tests that have four multiple choice options.

In line with this, the 2016-17 young lives longitudinal school survey on grade seven and eight students of Ethiopia also showed that students scored a low average mean (41%) on the Wave 1 (initial assessment) in mathematics test (Jack, Bridget, & Caine, 2017). Likewise, according to the study conducted in South Wollo Ethiopia, the mean score (43.7%) of government school students on regional examination was found low as compared to the national standard (Amogne, 2018).

Tassew and Aregawi (2016) identified variation in learning outcomes among children of twelve years old, nevertheless seven years apart as of 2006 to 2013, using the data from the young lives elder cohort whose ages were 12 years in 2006 and the younger cohort who were 12 years old in 2013. As a result, the analysis of the findings of the children's mathematics achievement test scores, confirmed that there was a statistically significant deterioration in mathematics achievement.

2.10. Factors Affecting Students' Achievement in Mathematics

Prominent factors related to student achievement are boundless and widely studied across different literature in which many intellectual debates and extensive discussions have taken around. Most of them include family background factors, socioeconomic status (SES), student personal characteristics, teachers' related variables, curriculum issues, and school-related factors. These factors are the major challenges and concerns of researchers and educators to assure the quality of teaching and learning in mathematics. According to Fardin, Alamolhodaei, and Radmehr (2011), international research has highlighted a number of factors influence on mathematics performance among learners, those arising from both school and home resources. Family background variables and school-related factors are significantly related to academic achievement (Engin-Demir, 2009).

In this regard, though, instructional planning is an effective way to alleviate problems related to the quality of teaching and learning mathematics, identifying the factors that affecting mathematics achievement is important to make the right decisions to improve achievement in mathematics. To identify the major factors associated with mathematics achievement, researchers have focused on many factors related to demographic, social, economic, and educational variables. For instance, international studies such as TIMSS measure student achievement and collect information on factors about schools, teachers, and curriculum that may affect this achievement as each country has its own impact within the context of its own culture (Mullis, Martin, Foy & Hooper, 2016). Several researchers find out a number of problems that are responsible for the achievement or failure in mathematics learning outcomes. However, this study only focused on some of the student and school-related factors that influence the student's achievement mean scores at individual and school levels. These are student home-related as well as school/classroom-related factors as presented below.

2.10.1. Student-Related Factors

The student-related factor is one of the aspects that influence students' achievement in mathematics. These aspects, mainly include students' personal or individual characteristics such as gender, home language, attitudes towards mathematics (interest, confidence & like), and home backgrounds such as parental education levels, SES, family possessions, family support, etc. are directly or indirectly affecting student achievement positively or negatively. Student and home variables have a direct effect on mathematics performance in which home variables also have an indirect effect on mathematics performance (Hui, 2014). Rumberger and Palardy (2004) pointed out that among many students' related factors such as students' demographic characteristics, family characteristics, and students' academic background influence students' achievement in mathematics. Student characteristics that are significantly affecting students' achievement are late to school and absenteeism, lack of assistance with studies at home and use of local language, parental support, supplementary reading materials, the interaction between students' parents and teachers (Abera, 2017). Teshome (2005) showed that studying with the mother tongue was a key factor in the achievement of students since those students who learned mathematics in their mother tongue were substantially performed higher than those who learned in the non-mother tongue. Kiwanuka, Damme, Noortgate, Anumendem, and Namusisi (2015) stated that SES, gender, past achievement in mathematics, family support, and the attitude of learners towards mathematics were important determinants for the success of students in mathematics. In order to understand mathematical concepts, the most important factor related to students' achievement in mathematics is the language of instruction, family income, and confidence in learning mathematics (Habtamu, 2017). According to Muleta and Garoma (2016), the effects of demographic factors, including gender, parents' education level, and SES on students' mathematics achievement are significant. As Mohammadpour (2012), students' self-concept, attitude towards studying mathematics, and their home educational materials are accounted for about 23 percent variations in mathematics achievement. According to Bofah and Hannula (2015) the achievement of students in mathematics can be affected by their gender, parental education levels as well as students' aspiration for further education. The higher educational aspirations of students and their parents' higher education associated with the higher achievement in mathematics (Gil-Flores, Padilla-Carmona & Suárez-Ortega, 2011). In general, demographic and individual factors are the main variables that contribute to the students' low or high achievement in mathematics (Hailu & Habtamu, 2016; Habtamu, 2017).

2.9.1.1. Home Background Factors

Home background variables such as family SES, education level, and their support are the foundation for students' academic achievement. Research has shown that family background characteristics such as SES as measured by parental education level, parental occupation, and family income influence school achievement (Hanushek & Luque, 2003; Ushie, Emeka, Ononga, & Owolabi 2012). Family background variables such SES and parents' education have a significant effect on students' overall academic achievement as well as achievement in mathematics (Farooq, Chaudhry, Shafiq, & Berhanu, 2011). Howie (2006) and Alordiah, Akpadaka, and Oviogbodu (2015) also indicated that, as family SES affects students' performance in mathematics. According to Brese, Mirazchiyski, Educational Testing Service and IEA (2013) findings from TIMSS grade 8 students, parental education and student achievement in mathematics have a strong association.

Similarly, applying the path-modeling techniques to data from the TIMSS and PIRLS 2011 assessments, Gustafsson, Hansen and Rosén (2013) indicated the indirect and direct effects of parental education and several books in the home on the students' achievement in mathematics in most participating countries. Ojimba (2013) also depicted that, as there is a significant relationship between home support for mathematics, SES of parents, and students' achievement in mathematics. As Savaş, Taş, and Duru (2010), family income, studying time, students' attitude towards mathematics and attendance had statistically significant effects on students' mathematics achievement. Daniel and Bamlaku (2017) indicated that mothers' literacy, fathers' literacy, and parents' income, had a positive and statistically significant relationship to regional primary school leaving examination scores including mathematics. According to the studies of young lives in four countries (Ethiopia, Peru, Indian, and Vietnam), if parents have more schooling and household wealth increases, then student's mathematics achievement scores improve over time (Lynn, Cheng, & Meisenberg, 2019).

Furthermore, Woessmann and Fuchs (2004), observed that students performed significantly worse in reading, mathematics and science in schools by the lack of parental support. Parental encouragement by paying for extra tuition, purchasing textbooks, promoting hard-working, participating in activities such as attending meetings of Parent-Teacher Association in schools, assisting in doing homework, and guiding learners was positively associated with the success of students in mathematics (Kiwanuka et al., 2015). Looking at Turkish TIMSS data (Yayan & Berberoglu, 2004) found that when parental education levels and numbers of books at home increased, eighth grade student achievement in mathematics also increased. Similarly, according to Brese et al. (2013) among the home properties, the number of books at home found to have a strong association with student achievement in mathematics. On the contrary, the author reported items of everyday life, such as dishwashers, cars, televisions, and cell phones had a low association with mathematics achievement. Ogunsola, Osuolale, and Ojo (2014), also depicted the education of parents and family socioeconomic background did not have any significant relationship with students' academic achievement.

2.9.1.2. Individual Student Factors

Studies have shown that individual student variables such as gender, attitude towards school environment, teacher, subject matter, involvement in school activities and effort, and their perception of parental support have significant effects on their school achievement.

Gender disparities and differences exist in the areas of schooling and academic achievement in countries all over the world. Particularly, with respect to gender differences in mathematics achievement, many research studies showed as boys outperformed girls (Kassahun & Kedir, 2006; Singh & Singh, 2007; NAE, 2004, 2008; Khatoon & Mahmood, 2 010; Kaahwa, 2012; Mohammadpour, 2012; Ochwo, 2013; Alordiah, Akpadaka, & Oviogbodu, 2015; NEAEA, 2012, 2016; Contini, Tommaso, & Mendolia, 2017). On the contrary, other researches revealed that as girls perform better than boys in the mathematics achievement (Wang, 2001; Ganihar & Wajiha, 2009; Farooq, Chaudhry, Shafiq, & Berhanu, 2011; Mubeen, Saeed, & Arif, 2013). On the other hand, some studies indicated as there was no significant difference by gender in mathematics achievement (Seleshi, 2001; Abiam & Odok, 2006; Choudhury & Kumar, 2009, Ajai & Imoko, 2014; Teketel, 2014).

A student's character can also influence academic achievement in school. That means how students perceive and think about schooling is a great predictor of student achievement. Students' self-concept, self-efficacy, attitudes, and motivation were found positively related to the achievement in mathematics (Adewuyi, Taiwo, & Olley, 2012; Chevalier, Gibbons, Thorpe, Snell, & Hoskins, 2009; Ajayi, Lawani, & Salomi, 2012). According to Chepete (2008) individual-level factors, specifically, the student attitudinal factor such as aspiration, self-confidence, and value placed on mathematics stood out as more crucial to affect mathematics achievement. For instance, according to the study conducted in Iran, student self-confidence in mathematics and attitude towards learning mathematics are highly related to the achievement in mathematics (Kiamanesh, 2005). There was a positive and significant relationship between students' attitude towards school, values of education, achievement motivation and academic achievement in mathematics (Asrat, 2017). Moreover, school attachment and commitment were positively linked with mathematics academic achievement (Stewart, 2007). One of the variables that accounted significantly for the variation of grade six students in mathematics achievement was their attitude towards learning mathematics (Seleshi, 2001). According to Tamiru (2017), the number of days students absent from school is one of the strongest influences in determining students' academic achievement. This finding is also supported by the Ethiopian NLA studies conducted on grade eight students (NEAEA, 2008, 2012, 2016).

2.10.2. School /Classroom Related Factors

School or classroom variables played an important role in affecting mathematics achievement. Particularly, the location, resources, and social atmosphere of a school often relates to its students' academic success. In Ethiopia, school factors are the main reason for differences in primary school students' achievement (Azubuike, 2015). As Kiwanuka, Damme, Van den Noortgate, Anumendem, and Namusisi (2015) reported, out of the total variance in mathematics achievement, about 14% and 17 % variation accounted for the classroom and school-level factors respectively. School climate and resources, school and classroom sizes, and good attendance in school were significant factors that affect mathematics achievement (Mohammadpour, 2012). Some of the school-related factors are discussed in detail as follows.

2.9.2.1. School Characteristics

School characteristics such as school location, class size, the economic background of the family in the school environment, instructional time, and school feeding are the determinant factors for students' achievement.

In this regard, Kyei and Nemaorani (2014) indicated that the location of schools affects students' mathematics performance. Igboegwu and Okonkwo (2012) in their study also showed the existence of significant differences in students' mathematics achievement with respect to the location of schools and concluded that urban school students achieved significantly better than rural school students. Alordiah, Akpadaka, and Ogbodu (2015) likewise showed as urban students achieved higher than rural students in mathematics subject. Correspondingly, in Ethiopia when looking at the trends of NLA of grade eight students' academic performance in mathematics by location, in most cases, students from urban performed better than students from rural areas, except in 2008 where rural students slightly perform better than urban counterparts (NAE, 2004, 2008; NEAEA, 2012, 2016). On the contrary, Yusuf and Adigun (2010) reported that there is no significant influence of school location in mathematics achievement. In line with this, Kissau (2006) found that students in urban and rural location performed in a similar manner.

Several findings also revealed the impact of classroom size on students' academic performance. According to Ito et al., (2017), an increase in class size resulted in a decrease of the assistance from teachers and peers and challenges classroom management which lead to the reduction in students' academic achievement. On the other hand, small class size helps to guide students better, to discuss openly with their teachers which creates a strong relationship between the students and teachers to have a positive effect on students' academic success (Crosnoe, Johnson, & Elder, 2004). Lindahl (2005) also found that some minorities and economically disadvantaged groups in Sweden benefited from smaller classes in their achievement in mathematics and science in TIMSS. On the contrary, analyzing the factors associated with students' mathematics achievement in the TIMSS 1999 and TIMSS 2011 studies of Finnish by Kupari and Nissinen (2013) showed as class size had a significant positive effect on mathematics achievement and suggested, as the performance was better in bigger classrooms. Similarly, Wößmann (2003) found that smaller class sizes significantly related to lesser student performance in mathematics and science in 39 countries participating in the 1995 TIMSS. Therefore, although there is an effect of class size on academic achievement, there has been no critical class size suggested by the researchers that increase or decrease the academic achievement of students. Hence, the relationship between school size and students' outcomes is controversial. In this regard, Slate and Jones (2005) concluded from their literature review; both very small and very large school and classroom size could be positively or negatively related to the students' academic achievement.

The school feeding program is another factor that could affect the achievement of students, though there were contradictory findings. According to Afridi, Barooah, and Somanathan,

(2014) schools with school feeding programs had the highest academic performance compared to those without it. Participation in the school feeding program increases test scores by 15.7 percent points (Ahmed, 2004). Similarly, Poppe, Frölich, and Haile (2019) have studied the effect of school meals on academic achievement and they found that supplementing children with school meals can be beneficial for concentration, reading, writing, and arithmetic skills. In Kenya, the school feeding program to public schools had shown to increase the academic performance of students as compared to schools with no school feeding (Chepkwony et al., 2013). However, evaluating the impact of school feeding programs in Ethiopia primary schools, Ermias (2015) found that there was statistically significant difference between the two groups in academic achievement test score. Likewise, the positive effect of school feeding program on academic achievement and attendance is not significant (Abiy, 2017). Adrogué and Orlicki (2013) also reported that while language test scores show a statistically significant improvement because of school feeding, no substantial effects were shown in mathematics scores.

The amount of time spent on study subjects and the number of instructional days in a school year affect the students' academic achievement. According to Dağlı (2019), there is a broad association between allocated instructional time and achievement. Cattaneo, Oggenfuss, and Wolter (2017), indicated that additional instruction time significantly increases the within-school variance of subject-specific test scores. The evidence from over fifty countries participated in PISA 2006 consistently showed that instructional time has a positive and significant effect on test scores (Lavy, 2015; Cattaneo et al., 2017; Tesfaledet, Beyene, & Tsegaye, 2019). However, Carr (2016) suggested that increasing instructional time did not result in higher achievement in reading and mathematics in the third grade PISA assessment.

2.9.2.2. School/Classroom Resources

School and classroom learning environment is the other variable that mainly determines the performance of students (Creemers & Kyriakides, 2008). Physical resources such as office space, playgrounds, classroom furniture, and libraries are associated with higher levels of students' achievement (Holmlund, McNally, & Viarengo, 2008; Lee & Linda Zuze, 2011). In many countries, school resources, including access to pure water and electricity, are associated with better student academic achievement (Huang, 2010; Lee, Zuze, & Ross, 2005). Moreover, mutual trust and positive interactions between staffs and students is a predictor for a better student scores (Stewart, 2008). According to Rumberger and Palardy (2004), school inputs and classroom processes contribute to students' achievement in mathematics. School contexts and facilities influence students' mathematics learning achievement significantly (Abaineh, 2012). According to Rajoo (2013) and (Ado, 2015), the quality of the learning environment is one of the important determinant factors that influences the achievement of students' mathematics. Accessibility of enough materials and facilities in the schools/classrooms are the core factors that affect the learning achievement of mathematics (Habtamu, 2017).

With regard to the relationship between school resources and students' achievement in mathematics, literature has shown inconsistent results. For instance, Hanushek & Luque (2003), Engin-Demir (2009) and Hanushek and Woessmann (2017) reported that more resources do not necessarily yield performance gains for students. On the contrary, Card and Krueger (1996) and Parcel, Dufur, and Zito (2010) have demonstrated that attending a school with a better physical environment is associated with increased mathematics scores. According to the findings of Afana, Lietz, and Tobin (2013) on grade eight students of Palestinian, Israeli Hebrew and Israeli Arab schools, shortages of computer hardware and software have a negative significant relationship with the students' achievement in mathematics. School resources, such as the presence of toilet facility have larger effects on academic achievement (Chowa, Masa, Ramos, & Ansong, 2015). Willms and Somer (2001) indicated that a statistically significant positive effect of adequate instructional materials and suitable size of school libraries on students learning outcomes in thirteen Latin American countries.

According to Amogne (2018), the reason for the differences in academic achievement between private and government schools in Ethiopia was a difference in library and laboratory resource facilities and equipment. While schools equipped with libraries and laboratories had positive effects on the achievement of students, on contrary spent more time to arrive at school had negative impacts on the achievement of students (Tesfaledet et al., 2019). Therefore, the quality and quantity of the educational facilities available within an educational system have positive relationships with the quality of the education and students' achievement. In particular, school factors are highly influential on mathematics performance (Mbugua, Kibet, Muthaa, & Nkonke, 2012).

2.9.2.3. School Emphasis on Academic Success

Teachers, parents, and students' priority and ambition for academic success reflected school emphasis. Attitude towards school emphasis is an important predictor of academic well-being (Patrick, Anderman, & Ryan, 2002). The attitudes of students towards school/classroom and their interest in learning have influenced and have been influencing students' achievements (Moè, Pazzaglia, Tressoldi, & Toso, 2009; Kpolovie, Joe, & Okoto, 2014; Veresova & Mala, 2016). In addition, the degree of teachers' success in implementing the school's curriculum, teamwork, understanding curricular goals, ability to inspire students, and their expectations for student achievement considered as school emphasis. Teachers' understanding and implementation of the school curriculum, their expectation and ability to inspire students, and their collaboration with other teachers play a significant role in improving students' academic achievement (Chu et al., 2015; Kocakaya & Kocakaya, 2014; Metzler & Woessmann, 2012; Zakharov, Carnoy & Loyalka, 2014). Çakır (2014) indicated there is significant mathematics mean score difference between low and high achiever students in terms of their attitude towards school, teachers, and motivation, where low achievers perform lesser than high achievers. Badri (2018) reported that attitudes and behavior of teachers, parents, and students towards school emphasis have a significant impact on students' TIMSS scores. Academic emphasis is important to differences among urban elementary schools in student mathematics and reading achievement (Goddard, Sweetland, & Hoy, 2000). The relationship between students and teachers is also the other component of school emphasis that is important in promoting students' achievement. Marzano (2007) has painted the relationship between students and teachers as a necessary component to enhance student learning.

In many cultures, parental involvement is a cultural conception of responsibility that influences students' academic achievement significantly (Ho, Chen, Tran, & Ko, 2010). That means, there is a significant positive relationship between parental involvement in education and students' academic performance in mathematics. Similarly, Arulmoly and Branavan, (2017) and Yohannes, Bereket, and Hailu (2017) indicated that parental involvement is an important predictor of mathematics achievement.

2.9.2.4. School Discipline and Safety

School discipline and safety, like student classroom disturbance and destruction, student exam malpractice, and lack of adequate environment is the key problem that can affect students' achievement in schools. In any country, the most effective schools are those who have a positive classroom condition with respect to the discipline of students in the classroom (Willms & Somer, 2001). Kutsyuruba, Klinger, and Hussain (2015) showed as safe and well-organized classroom environments related to the students' academic achievement. The authors boldly indicated the impact of a negative school environment like bullying, victimization, and violence as one of the significant problems in learning and students' achievement worldwide. According to Kutsyuruba, Klinger, and Hussain (2015), bullying is a serious problem throughout Latin American schools; where students who suffer from peer aggression and/or physical or verbal violence yields a significantly lower performance in reading and math than those in less violent classroom settings. Oliveira, de Menezes, Irffi, & Oliveira (2018), specifically stated that bullying has a negative impact on performance in mathematics. According to *Ending the Torment* (2016) report, students who are bullied have experienced 8% and 6% decrease in science and reading test scores on global achievement tests respectively. Correspondingly, one standard deviation increase in the within-school-level of bullying decreases individual test score by 0.55 (Van der Werf, 2014).

Student and teacher absenteeism is the other factor that reducing the achievement of students. When a teacher is absent from the classroom, student learning is disrupted (Obeng-Denteh, Yeboah, Sam, & Monkah, 2011), as a result, students' performance can be significantly impacted negatively. Absenteeism among teachers contributes immensely to the learners' poor performance, as it makes teachers not to cover the syllabus adequately (Nyabuto, 2007). The more days a teacher is out of the classroom, the lower their students tend to score low on standardized tests (Finlayson, 2009; Tesfaledet et al., 2019).

Furthermore, although this study does not discuss in detail, according to Muleta and Garoma (2016), the other significant factors that affect mathematics instruction and student achievement are curricula, instructional strategies, methods, teachers' competency, and school contexts. School contexts and facilities, students' self-directed learning, and motivation influence students' mathematics learning significantly (Abaineh, 2012). To sum up, demographic factors (gender, SES, and parents' educational level), school factors (school context and facilities), and individual factors (motivation, interest, and attitude) are the main factors that affect students' achievement in mathematics.

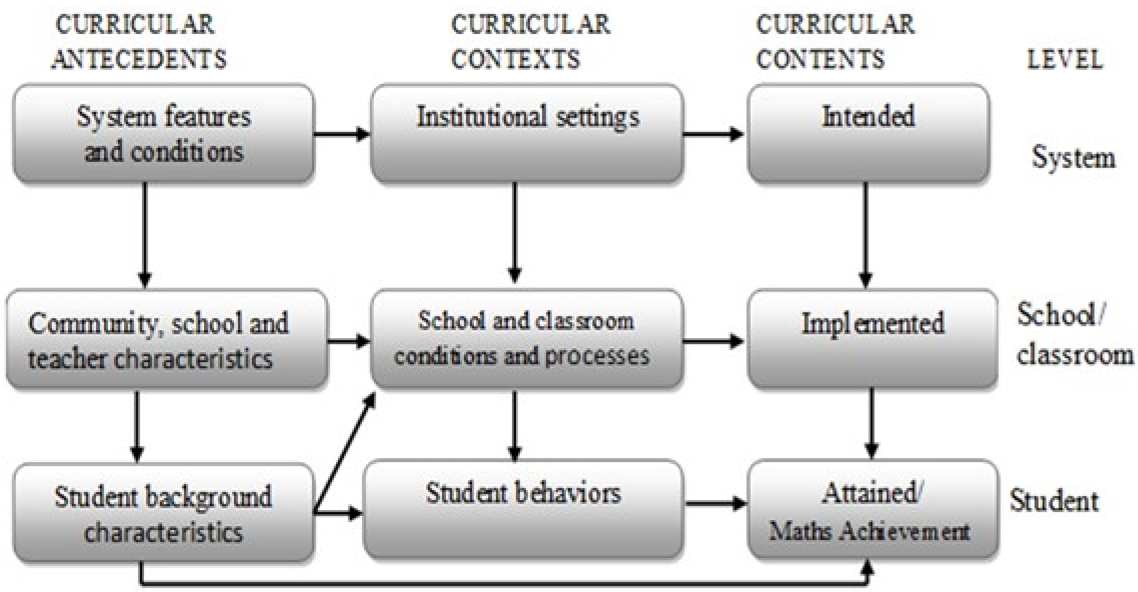
2.11. Conceptual Framework

The conceptual framework is the understanding of how the research problem is best explored, the specific direction of the research has to take, and the relationship between the different variables in the study. According to Miles & Huberman (1994), "a conceptual framework is a system of concepts, assumptions, and beliefs that support and guide the research plan". They also added as it lays out the key factors, constructs, or variables, and presumes relationships among them. The conceptual framework also allows in specifying and describing the concepts within the problems of the study for the researchers (Luse, Mennecke, & Townsend, 2012). Moreover, it provides a logical structure of related ideas that assist in offering a picture of how notions in a study connected.

The conceptual frameworks usually employed in ILSAs such as TIMSS were the three-curriculum framework, an IEA research framework, and TIMSS Model of the potential educational experience. However, Bos (2002) evaluated the three frameworks based on the quality of the tools they used for the clarification of factors, concrete definition they provided for the factors, their basis for assumptions about relationships between a cluster of factors, and their substantial theoretical basis. As a result, the author recommended that an IEA research framework was in a better position to use in large-scale assessments. This conceptual framework, mainly based upon the IEA thinking on the curriculum at three levels: the intended, implemented, the attained curriculum, and the three educational system levels (system, school/classroom, and student levels) or an IEA research framework (Bos, 2002; Travers & Westbury, 1989).

The three-curriculum levels can be described in terms of three broad sets of curricular variables: content, contexts, and antecedents. According to Mullis et al., (2009), the TIMSS curriculum model establishes a comprehensive description of effective factors that have shown effectiveness in increasing achievement in mathematics achievement. In this regard, figure 2 illustrates how the curricula are explained at the system, the school/classroom, and the student-level as well as the three curricular dimensions: curricular antecedents, contexts, and content levels.

Figure 2. An IEA Research Framework for TIMSS



Source: Travers and Westbury (1989, P.8) and Bos (2002, P.48)

The curricular content at each of the three levels (intended, implemented, and attained) is supposed to be influenced by the contexts in which it occurs and the contexts can be determined by a number of antecedent (background) conditions (Travers & Westbury, 1989). Conversely, antecedents of the curriculum influence the curricular contexts, as well as the curricular contents. In details, beginning with the curricular content, the descriptions for each of the levels in the conceptual framework are as follows.

Curricular Content: The set of curricular variables consists of the intended, implemented, and attained content. According to Rutkowski, Davier, and Rutkowski (2014), the intended curriculum is what is dictated by policy, the implemented curriculum is what is taught in the classroom, and the attained or achieved curriculum is what students know and can do.

The intended curriculum defined as a statement of society's goals for teaching and learning' and regarded as a planned curriculum for particular subjects (Travers & Westbury, 1989). At a system level, it is important to know to what extent the intended curriculum is valid and transmitted to the schools. The intended curriculum should be reflected in the curriculum documents such as goal statements, prescribed textbooks, syllabi, and other educational resources and examinations (Robitaille & Maxwell, 1996). Its contents although studied through an analysis of curriculum guides, official policy statements, and regulations.

The implemented curriculum is the second level of curricular content, which is regarded as part of the intended curriculum that is interpreted and translated by teachers and practiced at the classroom levels that can affect students learning. Usually, it refers to the concept of opportunity to learn. According to McDonnell (1995), the concept of opportunity to learn refers to schooling inputs and processes necessary for producing important student outcomes. The concept of opportunity to learn is also defined as the extent to which the subject contents needed to complete each of the items included in an achievement test has been covered (Travers & Westbury, 1989). Furthermore, Haan (1992) also indicated an opportunity to learn as that measures what learners actually taught in the classroom. Thus, the implemented curriculum is central to the educational process, in which the teacher has the responsibility for both transmitting the knowledge to the students and enabling them to acquire appropriate knowledge and skills.

Attained curriculum in a broad sense refers to curricular goals of the education system and the structures established to achieve it. Particularly, it is part of the intended and implemented curriculum, which is actually learned by the students. In this case, the knowledge and skills of a certain subject and attitudes towards that subject are usually the dependent variables of primary interest in such comparative studies. The attained or achieved curriculum can be described in terms of what students have learned, including their attitudes towards school and the particular subject under investigation (Bos, 2002).

According to Bos (2002), to some extent, the differences between intended curricula and what was actually implemented in the classroom can account for differences for the students' achievement across various countries and within countries as well. Accordingly, it is very crucial to consider the input, processes, and outcomes of education in order to understand the connections among the intended, implemented, and achieved curriculum. Therefore, the inclusion of the three curricular content levels allows the researcher to analyze possible explanations for the variables related to the students' achievement in mathematics.

Curricular Contexts: A contextual dimension is also included in the contextual framework because of not only curricular content variables, but also contextual factors can be the causes of differences in student outcomes (Travers & Westbury, 1989). At each curricular content level of the framework, a set of context variables is classified.

The contextual variables at the level of the intended curriculum are institutional settings. The good examples of such variables can be the existence of policies, central curriculum, and examinations. At the implemented curriculum level, school and classroom conditions and processes determine the curricular context. At this contextual level, features of the instructional practices, such as teachers' lesson structures and the type of assessment they apply, understanding curriculum can be mentioned. The curricular context of the third level of the curriculum framework consists of variables defined as student background and behavior. It includes the students' personal perceptions related to their misbehavior and school/class environment.

Curricular Antecedents: The antecedent dimension of the three-curriculum model contains background factors. These antecedents include system features and background characteristics of schools, teachers, and students that can affect contextual factors and attained curriculum as well.

Background factors classified as the intended curriculum level include system features such as the wealth of the society or national wealth, which may affect the retention of students within education (Travers & Westbury, 1989). At the implementation curriculum level, community characteristics (school location), school characteristics (school size, school socio-economic composition of its student intake, school types) and teacher characteristics were considered as antecedents. At the attained curriculum level, students' background factors were classified as antecedents. Travers & Westbury (1989) described this set of variables as 'the characteristics which the students bring to the classroom'. Some of the student background characteristics that can be classified at the level of the attained curriculum are the educational and socio-economic background of the student's family and the students' personal characteristics.

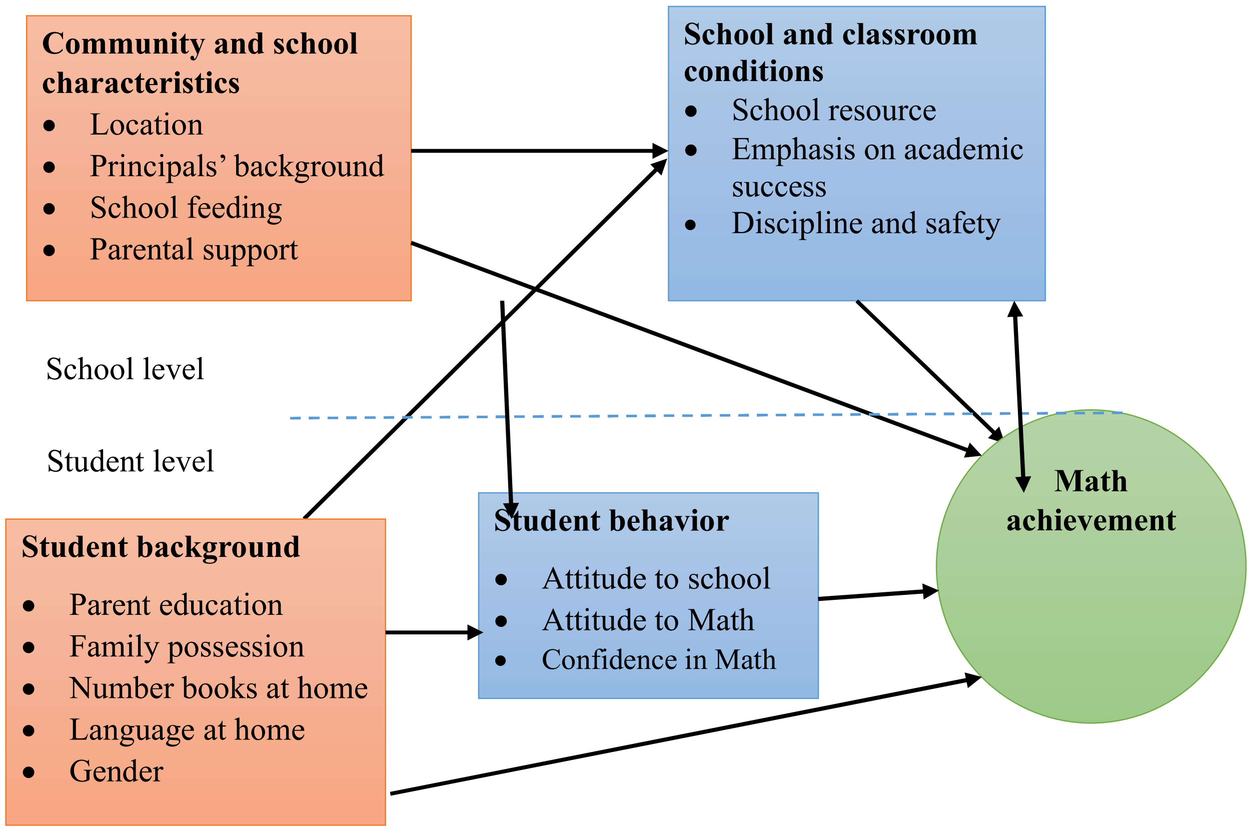
In this regard, the discussion of the empirical studies under literature reviews and the IES research framework indicated that how different factors at different levels affect the achievement of students. The focus of this study was also to find out what students' level and school level factors related and affect the achievement of students in mathematics. Thus, it was not emphasized on the factors at the system level in order to limit the scope of data collection and use two-level modeling for the data analysis. As discussed earlier, the main factors assumed to affect the mathematics achievement of students were student-related and school-related factors. Some of the student-related factors are home background or SES such as parent education, family possession, family income, meals per day, tutorial support at home and number of books at home, and student behaviors like their attitude toward the school and learning mathematics. Similarly, school or classroom-related factors include school characteristics (location, class size, free feeding, student composition, school distance) and school/classroom conditions such as school resources, academic emphasis, discipline, and safety. These factors have a direct and indirect impact on mathematics achievement as shown in the literature review and IEA conceptual of TIMSS.

In this study, therefore, taking into account the relationship among various variables with the mathematics achievement, the conceptual framework in figure 3 was adapted from the IEA studies applied in the SIMS found in Travers and Westbury 1989. This framework was preferred because of its contribution to the explanation of differences in learning outcomes better than the other TIMSS conceptual frameworks (Bos, 2002). In the design of this conceptual framework, an attempt was made to structure the variables thoroughly based on theoretical considerations and previous findings.

The left-hand side of the figure shows the antecedent variables known as community and school characteristics and student background. Community and school characteristics are not influenced by other variables in this framework, while student background variables can be affected by it. Community and school characteristics such as school location, school feeding, and principals profile variables are assumed either positively or negatively to influence a school and classroom conditions, student background, students behaviors and attitudes, and the outcome variable of mathematics achievement. The school and classroom conditions include the latent variables such as school resources, school emphasis on academic success and school discipline and safety are assumed to affect the outcome variable or mathematics achievement directly or indirectly through student behavior and it affects students' behavior directly.

Furthermore, student behavior, that is the student's motivations and attitudes towards school and mathematics learning are assumed to influence the individual students' achievement level in mathematics which is the outcome variables as well as school and classroom conditions possibly in a reciprocal relationship. Meanwhile, the antecedent variable of the student background such as parent education level, their possessions, number of books at home, student gender, and home language also influence the outcome variable directly or indirectly through the mediating variables of school and classroom conditions as well as student behavior. In this case, it is important to note that this conceptual framework measures the achievement of the student both at individual and school levels. Since one of the purposes of this study is to identify student and school-level factors that influencing mathematics achievement in Ethiopia.

Figure 3. The adapted conceptual framework of this study



CHAPTER THREE: METHOD OF THE STUDY

This chapter discusses the methodology and the design of the research study. It is separated into nine subparts. The first subsection discusses the research paradigm. The second part contains discussion on the survey research design applied in this study. In the third part, the sampling strategies and the participants of this study presented. In the fourth part, the instruments of data collection employed for the study discussed. The issues of validity and reliability of the data collection instruments discussed under the fifth subsection. The sixth section of the chapter deals with the procedure of the study. The seventh subsection of this chapter elaborates about the issue of equating and item response theory. The data analysis techniques used in the study presented under the eighth subsection. The last section of this chapter, presents the ethical considerations of this study throughout the whole process.

3.1. Research Paradigm

A paradigm is a collection of fundamental beliefs that signify a worldview that characterizes, for its holder, the essence of the world, the individual's position in it, and the degree of credible relationships to that world and its parts (Guba & Lincoln, 1994). In this regard, positivism is the main research paradigm for this study, as it is mostly related to the stated research questions and conceptual framework than the post-positivism paradigm, though both of them are well known and established worldwide. The reason to restrict to the positivism approach was that post-positivism paradigm was usually promoted the triangulation of qualitative and quantitative methods that explores the diversity of facts researchable through various kinds of investigations. However, this study was typically focused on the quantitative aspect. Thus, positivism was in a better position to conduct this study. The positivism research paradigm attempts to investigate, confirm and predict law-like patterns of behavior and is commonly used to test theories or hypotheses (Taylor & Medina, 2013). The positivism theory produced predominantly within the natural sciences and preserved an objective, patterned, and knowable truth that involves making and testing statements, as well as distinguishing and verifying causal relationships (Leavy, 2017). Particularly, it is convenient in natural science, physical science, and to some extent, in the social sciences, where large sample sizes are involved. Positivism adheres to the view that only factual knowledge gained through observation is trustworthy and depends on quantifiable observations that lead to statistical analysis. In studying positivism, the part of the researcher is restricted to information gathering, analyzing and objectively interpreting the data. In these kinds of research, the findings of the studies are most of the time observable and quantifiable. Therefore, this worldview offers the values for scientific objectivity, researcher impartiality, and replication (Hesse-Biber & Leavy, 2011).

As Kaboub (2008), within the late 19th century, the thinking of positivism originated as a truth-seeking philosophy through the critique of metaphysics by Auguste Comte and his claim that hardly specialized and rational facts can expose the reality-related truth. The positivist worldview emphasizes that real, honest, and factual events can be systematically and empirically interpreted and investigated as well as explained through articulation and logical observation and inquiry (Aliyu, Bello, Kasim, & Martin, 2014). According to Taylor and Medina (2013), the three basic beliefs that define a research paradigm including positivism are the ontological, epistemological, and methodological questions.

As this study is in a positivist position, ontologically it accepts the world as an outside (Carson, Gilmore, Perry, & Gronhaug, 2001) in which there is a single objective reality of this inquire about the phenomenon, notwithstanding the researcher point of view or perspective (Hudson & Ozanne, 1988). Slevitch (2011) states ontology as the study of reality or things that comprise reality. In this case, the researcher stays detached from the study participants by creating a distance, which is important to make clear distinctions between reason and feeling in keeping emotionally neutral (Carson et al., 2001), and also necessary to pursue objectivity and to use consistently real and logical approaches to this study. Statistical and mathematical strategies are dominant to this type of research, which follows particular systems to organize the procedures of the studies to reveal the single and objective reality (Carson et al., 2001). The ontological questions raised in such kind of research include what is the form and nature of reality and, therefore, what is there to be known about it? Hence, in assessing the achievement of the students in mathematics and identifying factors that affect their success in primary schools, truth is an outsider to the researcher and depicted by the objects in space.

According to Guba and Lincoln (1998) and Hesse-Biber and Leavy (2011), epistemology is a philosophical principle in how the studies proceeds, how one characterizes the researcher's position, and how one understands the researcher's relationship with the participants of the study. Typically, it poses the question of what is the essence of the knower's relationship, and what is known? The epistemological thought is asked about "What is the nature of the relationship between the inquirer and the knowledge?" as Allison and Hobbs, (2006) indicated. According to Bryman and Bell (2013), an epistemological issue concerns the discourse of what ought to be respected as acceptable information in a discipline of study. Therefore, the epistemological assumption of this approach where the theory of knowledge is believed to be objective and scientific realism, search for truth, and justification by empirical confirmation of hypothesis or research questions and universal scientific standard upheld in this study.

Methodologically, this research is quantitative in nature characterized by deductive approaches to the research process aimed at disproving or lending credence to existing theories; involves measuring variables and testing relationships between variables in order to reveal patterns, correlations, or causal relationships; results in statistical data from a large sample (Livey, 2017). The main methodological question for quantitative research is how the researchers can go about finding out whatever he/she believes can be known. Quantitative researchers try to work their studies under the assumption of objectivity. They believe that there is a truth to be observed and that rational observers looking at the same phenomenon will agree on its presence and characteristics. They try to remain as impartial or value-free as they can, and the effort to avoid human tendency at whatever point plausible. In a sense, quantitative researchers attempt to study the phenomena that are of interest to them “from a distance.” For example, standardized questionnaires and other quantitative measuring tools are often used to measure carefully what is observed in reality.

This method is particularly useful to quantify attitudes, opinions, behaviors, and other defined variables by way of generating numerical data and generalize results from a larger sample population. For these reasons, to assess the achievement of grade eight students’ in mathematics through the lens of TIMSS released test items and find out factors related to the students and schools, that affect the achievement of students’ using a quantitative research approach is appropriate.

3.2. Research Design

In this study, a survey research design was involved in order to assess what the achievement of Ethiopian students looks like in TIMSS released mathematics test items and identify factors that affect students' achievement. Particularly, a cross-sectional survey study is used to capture information based on data gathered for a specific point in time. A cross-sectional survey is used to gather information to make inferences about a population of interest at a point in time from the samples. It can be explained as snapshots of the populations about which the data were gathered. The data are gathered from a group of participants with varied characteristics and demographic variables such as age, gender, income, education, and geographical locations. Survey research is a widely used type of non-experimental research. Survey research is a research method that applies to a wide range of problems. It is a research design where individuals fill out a questionnaire or interviewed about their attitudes, activities, opinions, and beliefs. Survey research is also commonly used to test theoretical models constructed by researchers based on past literature, experimental results, and other factors (Pettigrew et al., 2008).

This research design is especially useful when one needs to make direct statistical generalizations about attitudes, opinions, and beliefs from a single sample to a population (Christensen, Johnson, & Turner, 2013). According to Leedy and Ormrod (2005), survey research is utilized to collect data about the population to understand their characteristics, outlook, states of mind, or trend. It provides a strategy for a quantitative or numeric explanation of experiences, attitudes, or opinions of a population by conducting research on a sample of that population (Cress well, 2014). Survey research has an expressive nature so that the analyzer does not manipulate variables, instead describes and makes inferences from frequency counts and other types of investigation (Burns & Grove, 2005). According to McMillan and Schumacher (2006), the reason for the researchers' turn face to the survey research was, because of the nature of its flexibility and more appropriate for collecting a large data from diversified type of populations. Lastly, survey research is the concept for working with large data on geographically scattered populations, when the other methods of research are not always feasible (Rubin & Babbie, 2008, Brewer, 2009). However, the limitation of survey research is the oversimplification of the reality in the society in which the arbitrary design of questionnaires with pre-conceived categories represent a predisposed and overly simple view of truth. Nevertheless, completing these seemingly simple steps needs a lot of thought and work, since, without this thought and work, the questions asked will prompt unreliable answers.

3.3. Participants and Sampling Techniques

The target population was grade eight students in Ethiopian primary schools. It only includes students from government schools found in urban and rural areas. Therefore, eighth grade students and school principals were the participants of this study.

Regarding the sampling and the sampling techniques, first Ethiopia divided into three strata. These are City Administrations, ‘emerging’ regions, and ‘non-emerging’ regions. ‘Emerging’ regions are regions found at peripheral areas and more disadvantaged in SES and ‘non-emerging’ regions are those in a better position relatively. The two regions (Benishangul Gumuz and Oromia) and one City Administration (Addis Ababa) were selected randomly among the nine autonomous regional states and the two City Administrations, since it is difficult to cover all the regions due time and financial problems. In this regard, Addis Ababa, Benishangul Gumuz and Oromia have represented the City Administration (Dire Dawa), the 'emerging' regions (Gambella, Somali and Afar), and the other remaining 'non-emerging' regions (Amhara, Harari, SNNP, Tigray) respectively. That means the stratified simple random sampling technique was employed to select one sample group from each stratum. In this case, based on the number of school proportions, this study was attempted to incorporate all schools from each zone in the sample regions and sub-cities of Addis Ababa. The stratified simple random sampling was also used to select participant students by gender and school location.

Moreover, using two-stage cluster sampling, each sample schools and students were selected randomly based on the sample size determination formula of a two-stage cluster sample design, which would provide an equivalent sampling accuracy for a simple random sample of 400 subjects (Ross, 2007).

The formula for calculating a sample size for two-stage cluster is $a = \frac{400}{b} \times [1 + (b-1)\rho]$.

Where: a = number of clusters

b = the size of selected cluster

ρ (rho) = coefficient of interclass correlation

In this case, the cluster size (b) was 40 students, which is equivalent to the average number of students-classroom ratio of second cycle primary schools, according to the Ethiopian Education Statistics Annual Abstract (MOE, 2018). Moreover, taking the data of mathematics test scores as the dependent variable and the schools as a random factor from the fifth national learning assessment of Ethiopia in grade eight 2016, the variance component analysis resulted in an interclass correlation rho was 0.11. The value of rho indicates a measure of the tendency of student characteristics to be more homogeneous within schools than would be the case if the students were assigned to the schools at random. Under this two-stage cluster sampling design, in the first stage, the sample schools selected randomly from each sample region using probability proportional to their size by school location, and then 40 students (20 boys and 20 girls) selected randomly from each school during the second stage. Equal number of boys and girls were selected as their number was almost the same in the sample frame of the target population. Finally, using the formula indicated above, 53 primary schools, 2,120 grade eight students, and 53 school principals participated in this study. At this point, the school principals are automatically selected from randomly selected sample schools.

Table 1. Planed and achieved number of the participants

Regions	Planed		Achieved		Response Rate (%)	
	Students	Principals	Students	Principals	Students	Principals
Addis Ababa	600	15	592	15	98.66	100
Benishangul Gumuz	480	12	448	12	93.33	100
Oromia	1040	26	1037	25	99.71	96.15
Total	2120	53	2077	52	97.98	98.11

Table 1 presents the number of planned and achieved sample participants and their response rate. The total response rates for the sample of students and school principals were 97.98% and 98.11% respectively. The response rates of the students by region were about 98.7%, 93.3% and 99.7% for Addis Ababa, Benshangule Gumuz, and Oromia respectively. Similarly, the response rate of the school principals for both Addis Ababa was 100% while that of Oromia was 96.15%.

Figure 4. Percentage of participant students by gender, location, and age category

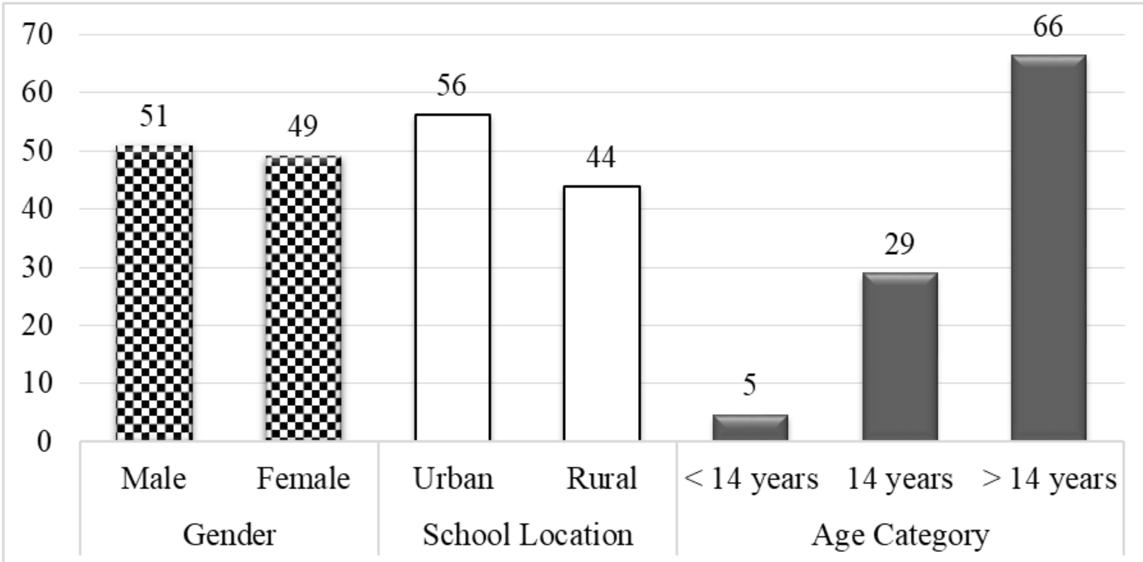


Figure 4 presents, demographic characteristics of the students who participated in this study. In this regard, among the number of sample students participated in the study, 51% of them were male and 49% of them were females. When looking at the participant students across school locations, 56% and 44% of them were from urban and rural schools respectively. Concerning the age of the respondent students, the majority (66%) of them were above 14 years old, 29% of them were 14 years old and the remaining 5% of them were under the age of 14 years.

Figure 5. Percentage of sample school principals by gender, age, and qualification

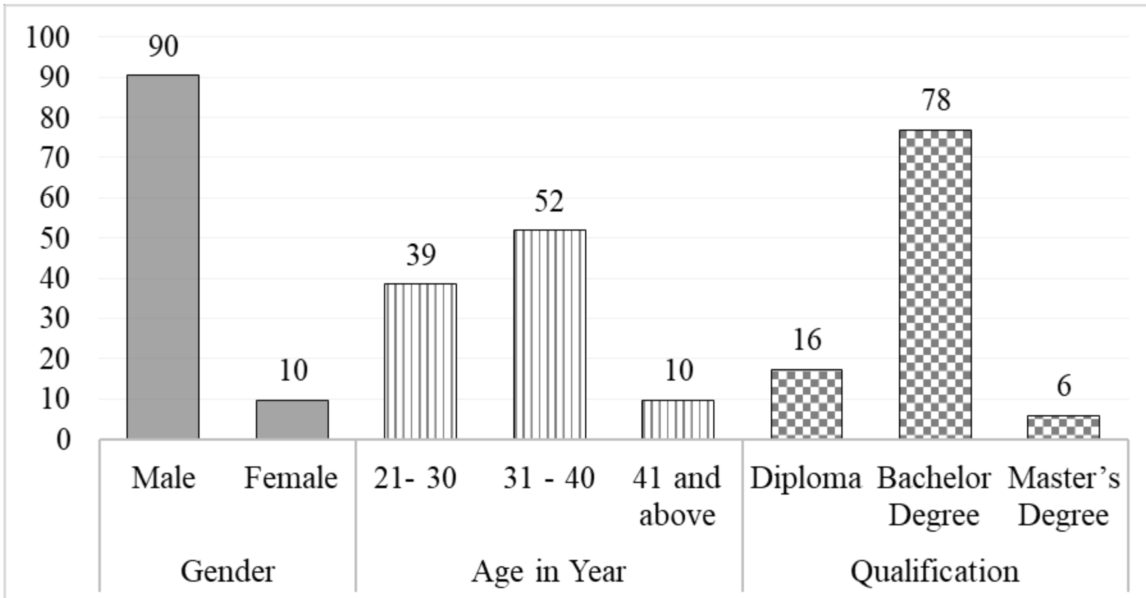


Figure 5 indicates the personal information of the school principals who participated in this study. Among 52 samples of school principals who completed the questionnaires, while most (90%) of them were males, few (10%) of them were females. Looking at their age category more than half (52%) of them were in the age range of 31 to 40 years old. The remaining 39 % and 10% of them were 21 to 30 and above 40 years old respectively. With regard to their highest educational qualification, the majority (77%) of them had bachelor's degree and the rest 17% and 10% of them had a master's degree and diploma respectively.

Table 2. School principals' work experience in percentage

Work experience	< 6 years	6-10 years	11 -15 years	> 15 years	Total
Total experience					
• Frequency	20	18	10	4	52
• Per cent	38	35	19	8	100
Experience at current school					
• Frequency	42	8	1	1	52

• Per cent	81	15	2	2	100

Table 2 presents the working experiences of the principals as school directors in general and their experiences at the current school in particular. Thus, about 73% of them had overall below 11 years of working experience and the rest (27%) of them had 11 and above years of working experience. Looking at their work experience at the current school as school principals, except 4% of them, the rest 81% and 15% of them had below six and six to ten years of work experience correspondingly.

3.4. Data Collection Instruments

Two kinds of data collection instruments were applied in this study. The first one was the mathematics achievement test of grade eight students and the second was the contextual questionnaires for both sampled students and school principals. These instruments were mainly adapted from the TIMSS 2011 data collection instruments. While mathematics achievement was used as the dependent variable, the variables from the responses of students and the school principals were considered as independent variables. The dependent variable, student learning achievement, was measured by raw percent scores as well as NLA and TIMSS converted scales scores of mathematics tests. The independent variables that refer to factors that affect the achievement of student learning in this study included students' background, the language of instruction, parental follow up and students' attitude to schools, students' attitude and confidence in learning mathematics, principals characteristics, classroom size, school resources, school emphasis on academic success and school discipline and safety. Table 3 shows the relation of variables and their descriptions.

Table 3. The description of student and school-related variables

Variables	Latent variables	Description of Observed variables
Dependent	Students' achievement	Mean score of mathematics in raw percent score and scale scores for NLA and TIMSS converted from correct score
Independent	Students' home background	Parents' education (mother and father or guardians), number of books at home, and family possessions (presence of bicycle, telephone, watch, table, radio, television…)
	Student personal characteristics	Students' gender, age, attitude to schools, attitude and confidence in mathematics, and student absenteeism
	Parental follow up in student learning	Parents asking what students learned at school, students talking with parents about schoolwork, parents make sure that students set aside time for their homework, and parents checking for students doing I do my homework
	School and principals' characteristics	Location, class size, school distance, principals' qualification, and work experience
	School resource	Lack of instructional materials (text books, teacher's guide, syllabus, concrete objects help to understand quantities, reference books in the library) Inadequate school facilities (school buildings and grounds, electricity, classrooms, computer technology for teaching and learning, pedagogical center, library, and laboratory) and lack of competent teachers in mathematics
	School emphasis on academic success	Teachers' understanding of the curricular objectives, their expectations for student achievement, and ability to motivate students Parental involvement in school activities, their commitment to ensure that students were ready to learn, and their pressure for the school to maintain high academic standards Students' desire to do well in school and their ability to reach the set academic goals
	School disciple and safety	Students and teachers unjustified absences, classroom disturbance, exam cheating, verbal abuse, and physical injury among the students

3.4.1. Mathematics Achievement Test

In the first place, the test items that aligned with Ethiopian primary school mathematics syllabi selected from the TIMSS released items in 2011. The test items of 2011 were used since the latest versions of the TIMSS mathematics item were not released yet. To check the validity of the test, subject specialists, i.e., two mathematics curriculum experts and four eighth grade mathematics teachers evaluated and selected 68 items that were more relevant to grade eight, based on the test blues print of the national learning assessment in 2016. Out of 68 items while 60 of them were multiple response items eight of them were constructed response items. In addition to that, 18 items were also added from the 2016 NLA mathematics test based on their quality of psychometric characteristics for the purpose equating. To reduce the students' overload or fatigue the items are assembled in two equivalent test forms or booklets; in which each of them consists of 40 items. This method is known as the "rotating booklet" method in the assessment world. All the adapted items were constructed in English language. However, in the case of Oromia as the language of instruction is Afan Oromo, the items were translated into Afan Oromo by a mathematics curriculum expert at the Oromia Education Bureau and a mathematics teacher who is teaching the subject in Afan Oromo. The intention was to create a high-quality test that properly adapted to the national context and at the same time is internationally comparable.

Before the main data collection took place, the achievement tests were piloted to check whether or not they are statistically fit for the purpose and to ensure that the assessment has a low bias. The piloted test was conducted in the fourth week of April 2018 at five primary schools selected from Addis Ababa, Benishagul Gumuz, and Oromia regions. In the pilot test, while three schools were selected from rural, two of them were from urban. After pilot testing, item and test analyzes carried out through a Classical Test Theory (CTT) and Item Response Theory (IRT) techniques using TAP and Item and Test Analysis (IATA) software respectively. The psychometric characteristics of each mathematics test items, such as, item difficulty, discrimination, and the quality of distractors as well as internal consistency reliability of the test, were thoroughly examined. The item analysis results indicated the mean item difficulty level (0.39), discrimination index (0.31), and reliability (0.81) were at an acceptable level. Moreover, the validity of the test items was also critically evaluated by expert judgments. Based on the analyses of the pilot data, the mathematics test items were further modified and improved in order to produce a valid and reliable final version of the test according to the existing contexts in Ethiopia (See Mathematics achievement test at appendix VI A and B).

3.4.2. Contextual Questionnaires

The primary purpose of the contextual questionnaires was to collect information about the home, classroom, and school contexts that could affect students' mathematics achievement. The questionnaire data when analyzed in relation to the achievement test yields insights into factors related to student achievement that can be helpful for policy-makers and educators. The contextual questionnaire is a self-completed survey instrument that can ensure a quick collection of a large amount of data both economically and conveniently (Bryman, 2008).

The student questionnaire, filled out by sampled grade eight students, mainly contained variables related to their personal characteristics, family background, their perceptions and attitudes towards the school, teachers, and mathematics lessons. Likewise, the school questionnaire, responded by the sampled school principals, also includes their background information and school situations such as class size, school feeding, school resources, emphasis on academic success, relationships with parents and community, school discipline and safety etc. Similar to the mathematics test items, the questionnaires were adapted from TIMSS contextual questionnaires of 2011. Then the students' questionnaire was translated into Afan Oromo and Amharic languages by language experts to assure their quality and appropriateness. Finally, the instruments were piloted together with the achievement test and necessary amendments and revisions, in turn, were made based on the feedback from students and school principals (see the English version of students and principals contextual questionnaires at appendix VII and VIII).

3.5. Validity and Reliability of Instruments

3.5.1. Validity

Validity is the extent to which research instruments measure what they are intended to measure (Borg & Gall, 1989). Validity is a matter of degree and no assessment instrument is completely valid or invalid. In particular, the assessment tool should measure both the appropriate content (content validity) and the skills and abilities that should be measured (construct validity) and; it should also be able to predict "success" as appropriate (predictive validity); it should entail few adverse consequences (consequential validity); and it should have an expected relationship with other measures of the same construct (external validity). Establishing both content and construct validity is an essential activity of test development (e.g., Lockheed, 2016; Messick, 1989). In this regard, assessing the alignment between the mathematics contents and objectives of Ethiopian primary schools across the TIMSS assessment framework for grade eight is thus very essential.

Content Domains: the thematic studying areas of mathematics in Ethiopia and the TIMSS mathematics assessment framework designed for about 60 countries in 2011 are similar. In both cases, the content domains include four thematic areas. These are numbers, algebra, geometry and data handling, and chance (probability). However, when these content domains are more decomposed into specific contents, few deviations are observed in Ethiopian Mathematics as compared to the TIMSS assessment framework, though, the majority of them are the same. These unique content areas that are found in the TIMSS assessment framework, but not incorporated in an Ethiopian primary school mathematics contents are: simultaneous of linear equations in two variables, generalize pattern relationships in a sequence and recognize and use of two-dimensional geometric transformations (translation, reflection, and rotation). On the contrary, specific content areas present in Ethiopian primary school mathematics but not included in the TIMSS assessment framework includes set theory, circles, absolute values, and an introduction to trigonometry.

Objectives/Cognitive Domains: a cognitive dimension specifying the domains or thinking processes to be assessed such as knowing, applying, and reasoning in the case of TIMSS framework, while in Ethiopian curriculum, they were formulated based on the Bloom's cognitive domain which includes knowledge, comprehension, application, analysis, synthesis and evaluation. These cognitive domains describe the sets of behaviors expected of students as they engage with mathematics content learning.

In both Ethiopian and TIMSS cases, the number content domain consists of understandings and skills related to whole numbers, fractions, decimals, integers, ratios, proportion, and percent. In this regard, in TIMSS framework, students are expected to extend their mathematical understanding of the whole numbers to integers, work with per cents and proportions and use proportional reasoning to solve problems related to everyday contexts and those where mathematics itself is the context (Mullis, Martin, Ruddock, O'Sullivan, & Preuschoff, 2009). Likewise, Ethiopian students are expected to extend their knowledge of the number system and operations on numbers to the set of rational numbers and operations using rational numbers, use the knowledge of ratio, proportions, and percentages to solve problems, apply mathematics to solve problems encountered in students' day-to-day life activities (MOE, 2015).

The content domains within algebra in TIMSS, include pattern recognition and extension, the use of algebraic symbols to describe mathematical contexts, and the development of fluency in the creation of similar expressions and the solution of linear equations. The contents in the mathematics of the Ethiopian curriculum also focused on working with variables in formulas, developing skills for solving linear equations and inequalities. A difference is that, in Ethiopian Mathematics contents, the concept of recognizing and extending patterns are missed.

Regarding geometry content domains students are expected to analyze the properties and characteristics of a variety of two and three-dimensional geometric figures, provide explanations based on geometric relationships, apply the Pythagorean Theorem to solve problems, using geometric properties and their relationships in TIMSS. In this case, the cognitive range extends from making drawings and constructions to mathematical reasoning about combinations of shapes and transformations. Similarly, Ehiopian curriculum noted specifically that students should identify, construct and describe properties of quadrilaterals, triangles, and circles, as the conditions for similarity to determine whether or not given triangles are similar and to solve related problems, identify congruent triangles by using the tests for congruency, identify pyramids and cones and their parts and calculate volumes of rectangular prisms. In geometry, the difference in transformation geometry that belongs only to TIMSS framework and in circles and solid figures (pyramids and cones) those only belong to Ethiopian Mathematics contents.

The data and chance content domain in TIMSS includes knowledge of how to arrange data collected by oneself or others, display data in graphs and charts, simple probability of designating the occurrence of familiar events as certain; as having greater, equal or lesser probability; or as unlikely and using data from experiments or knowledge of equally likely outcomes. In the objectives of the Ethiopian curriculum, students are also expected to collect data, construct and interpret simple graphs for giving data, calculate the mean, mode, and median of a given data and determine probabilities of simple events. Thus, in both cases the difference is negligible.

Moreover, to consider the relevance of the selected TIMSS test items with the Ethiopian primary school mathematics contents, the inter-rater reliability test was utilized. The inter-rater reliability is the degree of agreement among raters. It is a measure of how much consensus there is among the different judges' ratings. In this study, five grade eight mathematics teachers and two curriculum experts rated the selected test items of TIMSS against the Ethiopian primary school mathematics syllabus to evaluate to what extent they are relevant. A Krippendorff's alpha on reliability estimate test (κalpha) was run to determine the degree of agreement among the raters (i.e., teachers, curriculum experts). According to Krippendorff (1980), a κalpha is acceptable if it is 0.67 and above. The current study found a κalpha of 0.73, with 95% confidence interval between 0.56 and 0.88, indicating a substantial agreement among the raters. With regard to the contextual questionnaires, their validity was evaluated by two senior educational assessment experts who have long experience in designing questionnaires in NEAEA through expert judgment. This was checked on the clarity of the questionnaire items, thereby eliminating some ambiguous wording or misunderstood clauses. Thus, they were modified before the actual data collection. Items found to be inadequate for measuring the variables were modified to improve the quality of the contextual questionnaires and therefore improved their validity.

3.5.2. Reliability

Reliability is defined as the degree to which an instrument produces a constant result. It refers to the extent the instrument would be consistent in measuring what they are expected to measure (Mugenda & Mugenda, 1999). Testing for reliability is essential, as it refers to the consistency throughout the parts of a measuring tool (Huck, 2012). The most known measures of reliability comprise internal consistency, test-retest, and inter-rater reliabilities. Among these, the reliability of internal consistency looks at the accuracy of the score of individual-level items on an instrument, with the score of a group of items, or subscale, usually consisting of several items for measuring a single construct. Cronbach's alpha is one of the most familiar instruments that is used for measuring the internal consistency of reliability.

Though there are no strict rules for measuring internal consistencies, most agree upon a minimum coefficient of internal consistency of 0.70 (Cronbach, 1951; Whitley, 2002;

Tavakol & Dennick, 2011; Shiel & Cartwright, 2015). In addition, Hinton et al. (2004) suggested four reliability cut-off points that include excellent reliability (0.90 and above), high reliability (0.70-0.90), reasonable reliability (0.50-0.70), and poor reliability (0.50 and below). In this regard, the internal consistency reliability measured by Cronbach's alpha for mathematics achievement test in this study was 0.81 which is categorized under high reliability. Similarly, the coefficient of Cronbach's alpha for the students' contextual questionnaires in each of the subscales item ranged 0.73 to 0.91 on average, based on the pilot data (see Appendix IV). The reliability of the principal questionnaires was not discussed in this case, because it was administered only on five school principals' which was too small a sample size to apply a quantitative analysis. Overall, the reliability of the students' achievement mathematics test items and students' contextual questionnaires resulted in an acceptable value of internal consistency for this study.

3.6. Procedures of Data Collection and Management

Fifty-three samples of primary schools were selected and contacted by the researcher to get their consent for participating in the study ahead of time. Furthermore, a letter of cooperation written by the Center for International Comparative Education and Policy studies by Addis Ababa University was provided to the sample schools. Most of the data collectors were Zonal Education Bureau examination experts, who were directly responsible to manage assessment issues including national and regional examinations. They attended a fieldwork training session for one day on how to administer the achievement tests and questionnaires to the participants. Particularly, in the case of Benishangul and Oromia regions data collection was managed by these field workers with the close supervision of the researcher and colleagues from NEAEA. Nevertheless, the data collected from the Addis Ababa City Administration were fully handled by the researcher in collaboration with the sample of school principals and teachers. The data collecting fieldwork began on April 24/2019 and was completed on May 12/2019. In each school, a data collector stayed for a half-day to administer the instruments during the regular working time. In the data collection process, twenty-four Zonal Examination Unit experts and three NEAEA assessment experts have participated.

Followed the data collection, data capturing, and cleaning process was proceeding. The data capturing template has been designed in Microsoft Access 2016 that enables data encoders to enter data accurately and quickly. Since Microsoft access used a validation rule application that helps to ensure only valid responses are entered into the form field. In this regard, even though, more than half of the data were captured by the researcher to complete the data entry on time, two office assistants from NEAEA who had experience in data encoding tasks were involved in data capturing after they took hands-on training for a half day. To control the data quality, a double-entry method was applied using Microsoft excel 2016 that helps to minimize errors. Finally, the accuracy and consistency of the captured data were checked through the data cleaning process using Microsoft Excel and SPSS version 25.

3.7. Test Equating and IRT

Before proceeding to data analysis, discussing the concepts and procedures in test equating and IRT is very important since they are playing an important role in assessing the students' achievement. Equating adjusts for differences in difficulty among test forms that are built to be similar in difficulty and content in order to produce interchangeable and comparable scores (Kolen & Brennan, 2014). It is a statistical procedure to place two or more tests on a common scale. Test equating used to compare the results of national and international assessments, if they have common items that calibrated previously (Shiel & Cartwright, 2015).

The most popular way to perform equating is to link scores through the anchor-item non-equivalent groups design or known as fixed anchor calibration. In this design, the anchor or common test scores are used to adjust group differences in ability (Lu & Guo, 2018). Hence, when the two study sample groups vary inability, the use of the common test items information becomes critical, as it is the only means for differentiating differences in ability between the two groups of examinees from variances in difficulty level between the two equated tests. In such a process, one of the psychometric paradigms that play an essential role in designing, analyzing, and scoring tests is IRT (Sansivieri & Wiberg, 2017). IRT, unifies the concepts of item difficulty and discrimination and has also been described as latent trait theory. It is the most widely used theoretical approach in large-scale assessments.

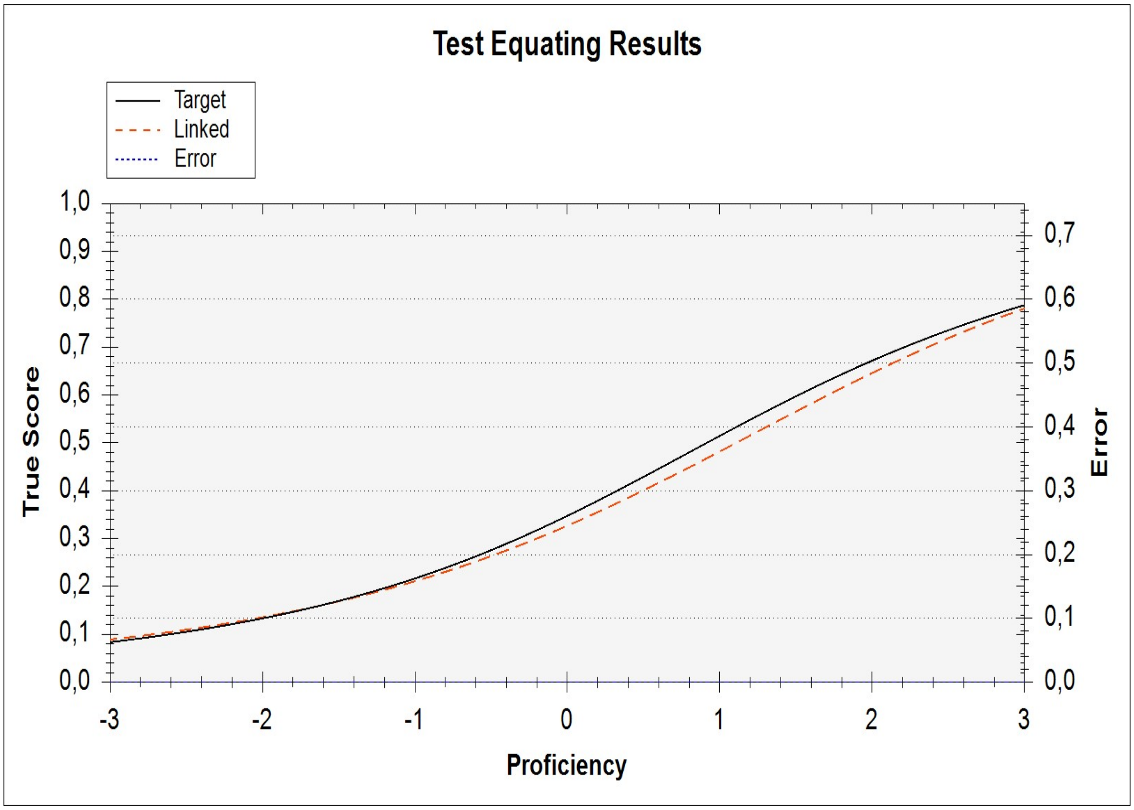
In this study, therefore, to compare the achievement mean score of students' in NLA across their achievement in TIMSS mathematics released test items, a non-equivalent group with anchor test design of test equating was involved, because both the test instruments and the participating groups are different. The most important property of the test with anchor items is its stability across the two tests equated. In selecting the common test items, content representativeness, psychometric properties (i.e., item difficulty and discrimination), and the total number of items were carefully considered. Since the international test items should be suitable to the national assessment test specifications with respect to their content, in selecting common items from both assessments that are deemed relevant to the grade eight mathematics contents and ensuring their IRT parameters were critical. In this case, only the items that were in the acceptable range of psychometric characteristics were selected as common items from the test items used in the NLA study 2016. Moreover, curriculum experts and the subject teachers evaluated the content representativeness of these items as already discussed.

Concerning the number of common items, Kolen and Brennan (1995) suggested a rule of thumb that the number of anchor items should be at least 20% of the length of a total test. Woldbeck (1998) recommended the minimum length of the anchor-item 20-25% of the total items. Angoff (1971) proposed 20% of the total length or 20 anchor items, whichever is greater. In this study, the number of common items included in the mathematics test achievement was about 23%. That means the number of common items utilized for this study was within the percentage ranges proposed by the scholars.

To produce IRT parametric estimate and perform test equating, the IATA software was employed. IRT models are employed to simultaneously estimate both the 'item' parameters (i.e., item difficulty and discrimination) and 'person' parameters (i.e., ability). Item difficulty describes the probability that the item is answered correctly based on the responses among the sample, while person ability depends on how many and which questions a student answers correctly. Thus, using IRT modeling, the scaled scores reflect not only the number of questions answered correctly by a student but also the difficulty levels of the questions answered correctly. The scaled scores are considered as directly comparable estimates of students' levels of the relevant proficiency (latent trait). The person ability parameter (latent trait) estimate or IRT score was transformed into a scale score. Similarly, in this study, the mathematics achievement test scores have been scaled using IRT and then transformed into a scale with a fixed mean score of 500 and a standard deviation of 100 to compare the student achievement with international and African countries' mean scores. The mean score of 500 and standard deviation of 100 are the standard scale scores for TIMSS. To compare the student mathematics achievement in TIMSS released test items with NLA achievement in 2016, the IRT scores were also transformed into scale scores using the fixed mean score of 300 and standard deviation of 50 which were the standards for Ethiopian NLA in 2016.

Figure 6 shows the result of two equated tests, namely, the targeted (adapted from released TIMSS items) and linked (NLA test items). The test characteristics curves of the two tests are almost similar after they are located on the same scale. This indicates how much the anchor or common items are strong in linking the two tests to equate. If the differences between the target and linked test items being consistent across a wide range of proficiency, then it is a problem to equate the two tests. However, as shown in this figure, the results of equating indicated a very stable linkage and the error line lies on the horizontal axis, which means it is almost equal to zero. Thus, the researcher is confident to compare students’ achievement using common test items from NLA and released test items of TIMSS.

Figure 6. Mathematics test equating results of on common items



3.8. Data Analysis

To analyze the collected and cleaned data, both descriptive and inferential statistics employed. The descriptive summary statistics such as frequencies, percentages, central tendencies, and dispersions were computed to summarize the overall finding of the study. Independent samples t-test used to compare the mean differences of mathematics scores between two groups such as gender, school locations, instructional language and assessment types (national versus international) etc. The one-sample t-test was also applied to compare the Ethiopian student achievement mean scores in mathematics against the African and international mean scores in TIMSS 2011. On the other hand, One-way Analysis of Variance (ANOVA) followed by the Post Hoc test was used to compute the mean differences among the regions and school principals' educational qualification category to identify the homogeneous subset groups. Here in all cases, the results were interpreted based on significance of statistical analysis, 95% confidence level with an acceptable level of probability significance value $p = 0.05$.

To find out the association of student and school-related factors with the students' mathematics achievement and their impacts on the students' achievement multilevel structural equation modeling or specifically two-level modeling was employed. Multilevel modeling is a model concerns with the interactions between variables described at different levels of a hierarchical data set that is frequently viewed from a hierarchically organized population of a multistage sampling. The important contribution of multilevel modeling is to avoid the problem of assigning the group-level mean to individual-level or treating subjects without reference to their groups (Silva, Bosancianu & Littvay, 2019). It means, avoiding the use of aggregated or disaggregated data to minimize the potential source of errors within the analysis. Particularly, Multilevel modeling with Structural Equation Modelling (SEM), which emphasizes latent variables has offered a number of benefits to the analysis of hierarchical data. It is useful in correcting model estimates for measurement error and offering the greater facility to define more complex sets of relationships among variables (Silva, Bosancianu & Littvay, 2019). SEM is also a statistical technique that involves a confirmatory approach or hypothesis-testing instead of an exploratory method to analyze the given data. It is a flexible model that allows estimation of relationships between observed and latent variables, direct and indirect effects, and assesses the fit of the overall model. Thus, to specify latent variables and their observed indicators, relationships between predictors and latent variables, and relationships between latent variables including possible intervening variables SEM plays an important role.

In this regard, as the sample of this study was a two-stage cluster, so that schools were selected randomly first and then from the selected schools, a representative number of students was randomly selected. Data sampled in such a way has a cluster structure and needs a multilevel structural equation modeling technique to adjust the underestimated standard error, so that valid inference on the population can be achieved. In this case, different factors that affect the achievement of students in mathematics both at student and school levels were identified in a set of two-level models for the data collected through the students' and school principals' questionnaires. This modeling is useful to look at the variability both at the student and school levels independently.

In multilevel level SEM, the hypothesized model can be tested statistically in a simultaneous analysis of the entire system of variables to determine the extent to which it is consistent with the data. In this case, if goodness-of-fit is adequate, the model argues for the plausibility of postulated relations among variables, otherwise, the model is rejected and further analysis is discontinued. In order to assess how well the proposed model fits the data, various statistical model fit tests can be used. However, in this study: the Chi-square test of model fit, Root Mean Square Error of Approximation (RMSEA), Comparative Fit Indices (CFI), and Standardized Root Mean Square Residual (SRMR) model fit tests were employed, as (Kline, 2016) recommended at least the indices of these four tests should be reported. In this case, the cutoff values to retain or reject the model were greater than .95 for CFI, less than .08 for both RMSEA and SRMR, and greater than 0.05 p- value for the Chi-square model fit, though it depends on the sample size of the study (e.g., Hooper, Couglan & Mullen, 2008).

To analyze the data, the Statistical Package for Social Science (SPSS) Version 25 and Mplus software (Mathén & Muthén, 1998-2017) used, and to apply the concept of IRT in item analysis, parametric estimation, test equating and comparing the Item and Test Analysis (IATA) software was applied. Finally, to store, manage, and cite references a Zotero software was employed.

3.9. Ethical Consideration in this Study

In the journey of the study, efforts were made to protect the research participants, to improve the validity of the data gathered, and to maintain the rigor of the overall research procedures and findings. Utmost efforts were made to protect the identities and moral values of the research participants. Equally, validity and procedural rigor have been taken care of as a part of maintaining ethical standards of the research work. Among others, efforts were made to give clear, understandable, and focused directions and criteria for data collection procedures and to maintain the anonymity of the data sources. Specifically, in conducting this research, the following ethical considerations had been considered.

- The participants received full consent prior to the study. That means consent and cooperation of data sources were secured through different mechanisms. For instance, the procedure to be followed in giving information based on specific data collection tools and the envisaged confidentiality of the responses were presented.
- The anonymity of individuals and schools participating in the study ensured. In this case, efforts were made to maintain the anonymity of the research participants, and the information, research participants provided by no means could reveal their identities.

- Respect for the dignity of the research participants is prioritized. This includes the use of derogatory, discriminatory, biased or other inappropriate language avoided in the formulation of test items and contextual questionnaires.
- Procedural rigor was one ethics, in which care was taken to maintain the rigor of the procedures in relation to developing and administering the instruments for data collection, in selecting appropriate participants, in collecting data, in meticulously sorting out and analyzing the data as well as in communicating research results with honesty and transparency.

  Overall, the utmost possible ethical roles have been considered to ensure that the research is conducted ethically, maintaining consent, anonymity, dignity, and procedural rigor.

CHAPTER FOUR FINDINGS OF THE STUDY

This chapter focuses on the findings of the study concerning the achievement of Ethiopian students' in mathematics and major students and school-related factors that affect their achievement. The first part discusses the overall descriptive summary statistics of the mathematics achievement. In the second part, the achievements of students' in mathematics across certain subgroups such as gender, location, instructional language, and region presented. In the third part, comparison of the achievement mean score of Ethiopian students in TIMSS released test items with that of TIMSS international and African countries mean scores discussed. Finally, the relationship between achievement and different contextual factors explored to gain some understandings of the determinants of students' achievement in mathematics.

4.1. The Overall Mathematics Achievement

The descriptive statistics of the mean scores of students' achievement in mathematics by percent raw scores and scale scores presented under this section. In this case, to transform the percent raw score to the scale scores the TIMSS mathematics achievement scales established in TIMSS 1995, where the scale center-point of 500 was set to correspond to the mean of the overall achievement distribution, and 100 points on the scale were set to correspond to the standard deviation was employed (Mullis, Martin, Foy, & Arora, 2012). Similarly, the conversion of the raw score into the NLA scale score was manipulated based on the standard set by NEAEA for grade eight NLA in 2016, where the mean and standard deviation were fixed to 300 and 50 respectively (NEAEA, 2016). In this regard, Table 4 shows the summary descriptive statistics of students' achievement in mathematics percent and scale scores.

Table 4. Descriptive statistics of students achievement in mathematics in percent and scale scores

Mathematics Score	N	Minimum	Maximum	Mean	SD
Raw score in per cent	2050	10	95	32.05	15.66
TIMSS scale score	2050	102.81	975.42	404.10	127.56
NLA scale score	2050	149.15	587.71	302.05	63.79

The overall achievement of students' raw mean scores in percent, NLA scale score, and TIMSS scale scores were about 32%, 302, and 404 points respectively. The results of the raw scores and NLA scale scores of this study are very close to the achievement mean scores (33%) and of NLA scale scores (300.36) points in 2016 findings. However, the TIMSS scale score was below the international center-point 500. Looking at the standard deviations (SD) for percent raw score (15.66), TIMSS scale score (127.56), and NLA scale score (63.79), the result indicated a wide variation among the students' achievement in mathematics average scores.

Table 5. Raw percent score of students' achievement by mathematics content areas

Content Area	N	Minimum	Maximum	Mean	SD
Number	2050	0	100	31.79	17.03
Algebra	2050	0	100	31.10	19.82
Geometry	2050	0	100	27.31	18.50
Data/Statistics	2050	0	100	38.02	25.19
Average	2050	5.77	85.90	32.05	14.27

Mathematics contents for grade eight in Ethiopia syllabus are very similar to that of what TIMSS content domains tested. As a result, the analysis of mathematics achievement in this study indicated that students relatively performed better in data handling or statistics with 38.02% mean scores, though the number of items in the test in this content area was fewer as compared to the other content areas. On the contrary, the achievement of students in geometry (27.31%) content domains was relatively the lowest. On the other hand, students' achievement in numbers and algebra content areas were almost similar about 32% and 31% respectively, which were in-between (Table 5). This finding is in contrast with the NLA studies in grade 10 students where students performed better in number (NEAEA, 2017). Similarly, it is also against the TIMSS 2011 findings, where 25 countries that participated in grade eight student assessment relatively achieve a high score in algebra, though, the achievement of their students in geometry content domain was lower and in line with this study. The largest values for the standard deviation (25.19) was referred to the data handling content domain, which indicated that the presence of a wide variation among the students' achievement in this content area.

4.2. Achievements of Students in Mathematics across Some Subgroups

4.2.1. Achievement by Gender

Table 6 presents differences in mathematics achievement across gender. In this regard, the raw mean score of achievement in mathematics for males (32.69%) was slightly higher than that of females (31.38%) with the mean difference of 1.3, which was not statistically significant. However, in NLA and TIMSS scale scores males achieve better than females in 2.17 and 12.25 mean difference points respectively. In this case, the independent samples t-test depicted that a statistically significant difference between males and females with regard to achievement in mathematics across gender in both NLA and TIMSS scale score at ($t_{(2048)} = 2.17$, $p < 0.05$) and ($t_{(2048)} = 2.18$, $p < 0.05$) respectively.

Table 6. Gender differences in mathematics achievement

Gender of students		N	Mean	SD	t	df	MD	*p*
Percent raw score	Male	1050	32.69	16.18	1.90	2048	1.31	0.058
	Female	1000	31.38	15.06				
NLA scale score	Male	1050	305.04	65.70	2.17	2048	2.17	0.030
	Female	1000	298.91	61.60				
TIMSS scale score	Male	1050	410.08	131.37	2.18	2048	12.25	0.030
	Female	1000	397.83	123.19				

The result of this study is supported by the findings of grade eight NLA in Ethiopia, in which consistently males significantly achieved better than females in mathematics (NEAEA, 2008, 2012, 2016). Similarly, it is also inlined with the findings of TIMSS 2011, wherein 22 countries males significantly achieved better than females, though, in ten countries females significantly achieved better than males in mathematics grade eight (Mullis et al., 2012). However, it is against the findings of Choudhury and Kumar (2009), Ajai and Imoko (2014), and Teketel (2014), which indicated that there is no statistically significant difference between males' and females achievement in mathematics. Interestingly, the variation of statistically insignificant p-value in percent raw score and its significant values in both scale scores between the subgroup of students was evidence for the importance of transforming raw scores into scale scores. However, the result showed that still, the gap between males and females achievement continued, even though the government had paid great attention to supporting girls' education differently due to gender stereotypes.

4.2.2. Achievement by School Location

Table 7. Mathematics achievement means scores by students' school location

School Location		N	Mean	SD	t	df	MD	*p*
Percent raw score	Urban	1163	31.13	15.76	-3.06	2048	-2.13	0.002
	Rural	887	33.26	15.45				
NLA scale score	Urban	1163	297.86	64.45	-3.41	2048	-9.68	0.001
	Rural	887	307.54	62.52				
TIMSS scale score	Urban	1163	395.73	128.89	-3.41	2048	-19.36	0.001
	Rural	887	415.08	125.03				

As shown in Table 7, the raw mean score (33.26%) of rural school students was better than their counterparts of urban school students whose raw mean score was 31.13% in mathematics. Likewise, with regard to the achievement in NLA and TIMSS scale scores, rural school students also achieved better than urban students with the mean difference (MD) of 9.68 and19.36 points respectively. In this case, the independent samples t-test revealed that the mean differences were statistically significant at ($t_{(2048)} = -3.06$, $p < 0.05$), ($t_{(2048)} = -3.41$, $p < 0.05$) and ($t_{(2048)} = -3.41$, $p < 0.05$) for the raw per cent score, NLA and TIMSS scale scores correspondingly. This finding is against the NLA studies in grade eight conducted in the last three consecutive assessment cycles (NEAE, 2008, 2012, 2016) where urban students achieved better, but similar to the findings of NLA 2004, where rural students achieve high (NEAEA, 2004). The reason for this could be, while the data were gathered in some urban areas of Oromia and Benishangule Gumuz, there were instability and conflicts, so that students may not be well prepared for the assessment. However, further study will be needed as students from urban schools have more privileges than rural schools in terms of various educational inputs.

4.2.3. Achievement by Students' Speaking Instructional Language at Home

Table 8 shows the mean differences of students in mathematics achievement by the frequency of speaking the instructional language at home. That means whether they speak the medium of instruction at home *never/sometimes* or *always*.

Table 8. Students' achievement by frequency of speaking instructional language at home

Speaking instructional language at home		N	Mean	SD	t	df	MD	P
Raw percent score	Never/sometimes	1097	29.07	13.85	-9.48	2043	-6.45	0.000
	Always	948	35.53	16.91				
NLA scale score	Never/sometimes	1097	289.70	58.40	-9.62	2043	-26.64	0.000
	Always	948	316.34	66.88				
TIMSS scale score as	Never/sometimes	1097	379.41	116.80	-9.62	2043	-53.29	0.000
	Always	948	432.69	133.76				

According to the finding, the raw mean score of achievement in mathematics for those who speak the language of instruction at home *always* (35.53%) was higher than that of students who *never or sometimes* speak the language at home (29.07%) with the mean score difference about 6.5 that was statistically significant. Similarly, the difference was also revealed both in NLA and TIMSS scale scores, which in favour of those students who always speak the language of instruction home. In this case, the independent samples t-test indicated a statistically significant difference both in NLA and TIMSS scale scores at ($t_{(2043)} = -9.62$, $p < 0.001$). This finding is inlined with several studies that indicated the impact of the similarity of language at home and school on students achievement in Ethiopia (Teshome, 2005; NEAEA, 2008, 2012; Habtamu, 2017; Abera, 2017). It also supported with the finding of Rahman, Mazlan, Kummin, Yasin, and Meerah (2010) indicated that the use of a second language as a medium of instruction can negatively affect the performance of students in learning the subjects. Thus, encouraging students to learn through their first language is very important and a human right aspect, particularly for those in primary schools.

4.2.4. Achievement of Schools by Students' Average Class Size

Several studies indicated the positive and negative effects of class size on students' academic achievement, although it is a debating issue among the researchers with regard to the proper number of average students per classroom. In this study, therefore, to look at the impact of class size on the students' achievement in mathematics, the data were analyzed using a t-test statistic based on two independent sample groups of schools. That is, the small class size (schools with 50 or less than50 students per classroom) and larger class size (schools with more than 50 students per classroom) on average.

Table 9. Achievement of schools by students' average class size

Average class size		N	Mean	SD	t	df	MD	*p*
Raw per cent score	≤ 50	32	33.63	10.16	1.79	48	4.66	0.079
	> 50	18	28.97	5.57				
NLA scale score	≤ 50	32	308.23	40.51	1.74	48	18.06	0.088
	> 50	18	290.17	22.30				
TIMSS scale score	≤ 50	32	416.46	81.01	1.74	48	36.13	0.088
	> 50	18	380.34	44.59				

As indicated in Table 9, the achievement mean scores in mathematics for smaller size schools were slightly higher than that of larger class size schools for all three types of scores (raw percent score, NLA and TIMSS scale scores). Nevertheless, the independent samples t-test indicated there are no statistically significant differences between the class sizes in raw percent score and both in NLA as well as TIMSS scale scores at ($t_{(48)} = 1.79$, $p = 0.079$) and ($t_{(48)} = 1.74$, $p = 0.088$) respectively. This result contradicts with studies of (Crosnoe, Johnson, & Elder, 2004; Ito et al., 2017) whose stated students in smaller class size performed better than students in larger class size and the finding of (Kupari & Nissinen, 2013) who indicated students in larger class size achieved better than students in smaller class size. However, this finding agreed with the justifications of Slate and Jones (2005), who underlined as both too small and too large class sizes had their own impact on students' achievement.

4.2.5. Students' Achievement by Free School Feeding

Table 10. Achievement of students' by providing free breakfast and lunch at schools

Number of students		N	Mean	SD	t	df	MD	p
Raw percent score	Some	21	29.27	10.22	-1.83	48	-4.63	0.073
	All	29	33.90	7.63				
NLA scale score	Some	21	289.93	39.90	-2.04	48	-20.34	0.047
	All	29	310.27	30.59				
TIMSS scale score	Some	21	379.86	79.81	-2.04	48	-40.68	0.047
	All	29	420.54	61.17				

Table 10 presents the difference in mathematics mean scores of the schools by the number of students provided with both free breakfast and lunch. Fortunately, all the sample school principal schools replied as they provided both free breakfast and lunch either to some or all students, except one school that was excluded from this analysis. In this regard, looking at the raw mean score, although the result of schools that provided free feeding for all students slightly better than those who feed some students, the difference in mean score was not statistically significant at p = 0.073. On the other hand, the finding indicates there is a statistically significant mean difference between the schools that provided free breakfast and lunch for *all students* and those who feed *some students* both in NLA and

TIMSS scale scores at ($t_{(2048)} = -2.04$, $p < 0.05$). This finding is in line with other several studies that indicated the positive effect of school feeding programs on students' achievement (Chepkwony et al., 2013, Afridi, Barooah & Somanathan, 2014; Ermias, 2015; Abiy, 2017). Thus, encouraging the school feeding program has a great contribution to the quality of students learning and better achievement.

4.2.6. Achievement by Region

Table 11. Mathematics achievement mean scores by region

	Region	N	Minimum	Maximum	Mean	SD
Raw percent score	Oromia	1015	10.00	92.50	32.99	15.44
	Benishangul Gumuz	445	10.00	70.00	29.51	11.33
	Addis Ababa	590	10.00	95.00	32.37	18.42
	Total	2050	10.00	95.00	32.05	15.66
NLA scale score	Oromia	1015	149.15	587.71	305.20	62.64
	Benishangul Gumuz	445	151.41	453.03	291.80	48.21
	Addis Ababa	590	149.73	587.71	304.37	74.47
	Total	2050	149.15	587.71	302.05	63.79
TIMSS scale score	Oromia	1015	102.81	975.42	410.40	125.27
	Benishangul Gumuz	445	102.81	706.06	383.59	96.42
	Addis Ababa	590	102.81	975.42	408.74	148.92
	Total	2050	102.81	975.42	404.10	127.56

The mathematics achievements of two regions (Oromia and Benishangul Gumuz) and a City of Addis Ababa administration analyzed as shown in Table 11 for all the students of sample schools. As a result, in percent raw scores, Oromia and Addis Ababa slightly scored above the total mean score of 32.05% with the mean scores of 32.99% and 32.37% respectively. Whereas, students from Benishangul Gumuz achieved the lowest mean score

(29.51%) that was below the total mean score. Similarly, with regard to the achievement of regions in scale scores, Oromia and Addis Ababa were achieved slightly above the total mean scores of both NLA and TIMSS. However, Benishangul Gumuz region was achieved both below the NLA and TIMSS total mean scores.

Table 12. Achievement of students by region using One-way ANOVA

Groups		Sum of Squares	df	Mean Square	F	*p*
Raw percent score	Between Groups	3830.16	2	1915.08	7.87	0.000
	Within Groups	498369.30	2047	243.46		
	Total	502199.47	2049			
NLA scale score	Between Groups	60014.18	2	30007.09	7.42	0.001
	Within Groups	8276855.11	2047	4043.41		
	Total	8336869.29	2049			
TIMSS scale score	Between Groups	240144.70	2	120072.35	7.43	0.001
	Within Groups	33102558.63	2047	16171.25		
	Total	33342703.33	2049			

As indicated in Table 12, One-way ANOVA carried out to look at the fit of the model and identify the differences among the regions in mathematics achievement taking regions as independent and the mathematics means scores as the dependent variable. As a result, F-test showed that there was a statistically significant difference at least between two participant groups or regions in mathematics achievement for percent raw mean scores, NLA and TIMSS scale scores at ($F_{(2, 2049)} = 7.87$, $p <. 001$), ($F_{(2,22049)} = 7.42$, $p <.05$) and ($F_{(2, 2049)} = 7.43$, $p <.05$) respectively. Thus, it is possible to find out the specific significant difference among the regions through the Post Hoc multiple comparison tests.

As mentioned earlier, to see the differences among regions, a Post Hoc multiple comparisons of ANOVA using the Tukey HSD test executed for the mathematics achievement of percent raw score as well as NLA and TIMSS scale scores (see Table 13).

Table 13. Post Hoc multiple comparisons among regions in mathematics mean scores

Dependent Variable	(I) Region	(J)Region	MD (I-J)	*p*
Raw percent score	Oromia	Benishangul Gumuz	3.48	0.000
		Addis Ababa	0.62	0.724
	Benishangul Gumuz	Oromia	-3.48	0.000
		Addis Ababa	-2.86	0.010
	Addis Ababa	Oromia	-0.62	0.724
		Benishangul Gumuz	2.86	0.010
NLA scale score	Oromia	Benishangul Gumuz	13.40	0.001
		Addis Ababa	0.83	0.966
	Benishangul Gumuz	Oromia	-13.40	0.001
		Addis Ababa	-12.57	0.005
	Addis Ababa	Oromia	-0.83	0.966
		Benishangul Gumuz	12.57	0.005
TIMSS scale score	Oromia	Benishangul Gumuz	26.81	0.001
		Addis Ababa	1.65	0.966
	Benishangul Gumuz	Oromia	-26.81	0.001
		Addis Ababa	-25.15	0.005
	Addis Ababa	Oromia	-1.65	0.966
		Benishangul Gumuz	25.15	0.005

As indicated in Table 13, in all mathematics achievement mean scores (i.e., percent and scale scores) there was a statistically significant difference between Oromia and Benishangul Gumuz regions as well as Addis Ababa and Benishangul Gumuz at $p < .05$, where Oromia and Addis Ababa were relatively performed higher than Benishangul Gumuz. Nevertheless, there is no statistically significant difference between Oromia and Addis Ababa at $p > .05$. This result is similar to the findings of NLA for the last two decades, where Benishangul Gumuz was categorized as one of the regions that performed least among the other regions.

Table 14. Homogenous subset groupings of regions using the Tukey HSD test by Mathematics percent and scale mean scores

Region	N	Raw score		NLA scale		TIMSS scale	
		1	2	1	2	1	2
Benishangul Gumuz	445	29.51		291.80		383.59	
Addis Ababa	590		32.37		304.37		408.74
Oromia	1015		32.99		305.20		410.40
Sig.		1.000	0.768	1.000	0.972	1.000	0.972

Means for groups in homogeneous subsets are displayed at *p* =.05).

a. Uses Harmonic Mean Sample Size = 608.85.

b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

Furthermore, as depicted in Table 14, regions categorized into two homogeneous subset groupings based on their students' achievement in mathematics percent raw scores as well as both in NLA and TIMSS scale scores. That means, students from Oromia and Addis Ababa categorized under the same group (group two) as they were relatively high achievers and students from Benishangul Gumuz categorized under group one since they were performed lower. This indicates that there was no statistically significant mean difference between Oromia and Addis Ababa at $p > .05$, because they were clustered into the same group. However, both of them have a statistically significant mean difference from that of the Benishangul Gumuz region at $p < .05$.

4.2.7. School Achievement Mean Scores by Principals' Characteristics

Table 15 indicates the achievement mean score of school based on the highest qualification of school principals. In this case, the highest qualification of the school principals was categorized as diploma, first degree, and master's degree.

Table 15 School achievement mean scores by principals' qualification

Principals Qualification		N	Minimum	Maximum	Mean	SD
Raw per cent score	Diploma	3	21.36	32.40	26.00	5.73
	First Degree	39	20.61	68.10	31.46	8.46
	Master's Degree	8	27.93	62.64	36.60	11.37
	Total	50	20.61	68.10	31.95	9.01
NLA scale score	Diploma	3	253.21	306.42	275.57	27.60
	First Degree	39	256.64	439.18	300.03	33.20
	Master's Degree	8	285.42	426.38	319.80	46.12
	Total	50	253.21	439.18	301.73	35.88
TIMSS scale score	Diploma	3	306.41	412.84	351.14	55.20
	First Degree	39	313.27	678.36	400.07	66.40
	Master's Degree	8	370.84	652.77	439.59	92.24
	Total	50	306.41	678.36	403.46	71.76

As a result, in all the three types of scores (raw percent score, NLA, and TIMSS), the achievement mean score of schools led by principals who held master's degree were higher than that of both schools led by diploma and first degree holders. Likewise, the mean score of schools led by first degree holder principals also better than that of schools led by diploma holders in scale scores, though it was lower in raw score.

Table 16. One-way ANOVA model for principals' qualification and achievement

Achievement by groups		Sum of Squares	df	Mean Square	F	*p*	Eta Squared
Raw per cent score	Between Groups	288.85	2	144.43	1.84	0.170	0.07
	Within Groups	3691.29	47	78.54			
	Total	3980.14	49				
NLA scale score	Between Groups	4777.00	2	2388.50	1.93	0.157	0.08
	Within Groups	58300.09	47	1240.43			
	Total	63077.09	49				
TIMSS scale score	Between Groups	19106.64	2	9553.32	1.93	0.157	0.08
	Within Groups	233186.21	47	4961.41			
	Total	252292.84	49				

However, as shown in Table 16 a One-way ANOVA indicates the model does not fit with regard to the differences among the school's achievement across the principals' qualification. The F-test also revealed that there is no statistically significant difference at least between the two categories (qualification of principals) in mathematics achievement at ($F_{(2, 49)} = 1.84$, p =0.07) and ($F_{(2,249)} = 1.93$, p =0.08) for percent raw mean scores as well as both NLA and TIMSS scale scores respectively. Furthermore, the values of eta squared 0.07 and 0.08 for raw score and scale scores respectively indicated, the moderate difference in effect size among the students' achievement across principals' qualification. According to the interpretation of Cohen (1988), Eta squared values 0.01 = small effect size, 0.06 = moderate effect size and 0.14 = large effect size.

Table 17. School achievement mean scores by principals' working experience

Achievement by working experience		N	Minimum	Maximum	Mean	SD
Raw per cent score	≤ 5 years	19	20.61	42.53	30.49	6.36
	6-10 years	17	22.09	62.64	31.36	10.04
	> 10 years	14	25.08	68.10	34.65	10.70
	Total	50	20.61	68.10	31.95	9.01
NLA scale score	≤ 5 years	19	253.21	341.95	295.08	25.47
	6-10 years	17	266.67	426.38	300.45	40.86
	> 10 years	14	267.75	439.18	312.29	41.38
	Total	50	253.21	439.18	301.73	35.88
TIMSS scale score	≤ 5 years	19	306.41	483.90	390.17	50.94
	6-10 years	17	333.34	652.77	400.90	81.72
	> 10 years	14	335.49	678.36	424.59	82.75
	Total	50	306.41	678.36	403.46	71.76

Table 17 shows the achievement mean score of schools by principals' total working experience. As a result, in all the three types (raw percent score, NLA and TIMSS scale scores) average scores, schools led by principals who had more than 10 years of working experience were performed better than the schools who led by the principals with the working experience of six to ten and less than five years. Similarly, schools in which their principals' working experiences were six to ten achieved higher than those schools led by less than 5 years working experience.

Table 18. One-way ANOVA model for principals' working experience and achievement

Groups		Sum of Squares	df	Mean Square	F	p	Eta Square
Raw percent score	Between Groups	148.14	2	74.07	0.91	0.410	0.04
	Within Groups	3832.00	47	81.53			
	Total	3980.14	49				
NLA scale score	Between Groups	2429.01	2	1214.50	0.94	0.397	0.04
	Within Groups	60648.08	47	1290.38			
	Total	63077.09	49				
TIMSS scale score	Between Groups	9713.20	2	4856.60	0.94	0.397	0.04
	Within Groups	242579.65	47	5161.27			
	Total	252292.84	49				

Nevertheless, as shown in Table 18 a One-way Analysis result indicates that the model does not fit with regard to the differences among the schools' achievement across the principals' working experience. The F-test also revealed that, there is no statistically significant difference at least between the two categories of principals working experience in mathematics achievement at ($F_{(2, 49)} = 0.91$, p =0.037) for per cent raw mean scores and ($F_{(2,249)} = 0.94$, p =0.039) as well as ($F_{(2,249)} = 0.94$, p =0.038) for NLA and TIMSS scale scores respectively. The effect size of eta square among the students' achievement by the working experience of the principals was also small at 0.04, which confirmed that the difference was insignificant.

4.3. Students' Achievement as Compared to International and African Mean Scores

To compare the achievement of Ethiopian students in mathematics achievement with the international TIMSS mean score or center-point (500), it requires to equate the tests and transform the raw scores into scale scores as discussed earlier under section 3.8. TIMSS usually uses the center-point (500) of the scale as a point of reference that remains constant from assessment to assessment (Mullis et al., 2012). In this case, before the comparison has taken place, the curriculum analysis results of the five African countries in TIMSS 2011 were checked, whether they aligned with the Ethiopian mathematics syllabus or not and then mathematics achievement of students in raw scores of this study was transformed into the scale scores using the IRT model.

Table 19. Ethiopian students' achievement as compared to TIMSS international Center-point and African countries mean scores

Comparison	Test Value = 500						
	N	Mean	SD	t	df	MD	*p*
Ethiopian Vs International mean	2050	404.10	127.56	-34.04	2049	-95.90	0.000
	Test Value = 397						
Ethiopian Vs Botswana	2050	404.10	127.56	2.52	2049	7.10	0.012
	Test Value = 331						
Ethiopian Vs Ghana	2050	404.10	127.56	25.95	2049	73.10	0.000
	Test Value = 371						
Ethiopian Vs Morocco	2050	404.10	127.56	11.75	2049	33.10	0.000
	Test Value = 351						
Ethiopian Vs South Africa	2050	404.10	127.56	18.85	2049	53.10	0.000
	Test Value = 425						
Ethiopian Vs Tunisia	2050	404.10	127.56	-7.42	2049	-20.90	0.000

As shown in Table 19, the achievement mean scale score (404.10) of the Ethiopian students was found lower than that of the international center-point or mean score (500) of TIMSS with the mean difference point of 95.9. In this case, the one-sample t-test confirmed that, the difference between the achievement mean scores of Ethiopian students in this study and the international mean score 500 or center-point for TIMSS study in 2011 was statistically significant at ($t_{(2050)} = -34.04$, $p < .001$).

On the other hand, the achievement of Ethiopian students in mathematics also compared with the average mean scores of five African countries that were participated in TIMSS 2011. The mathematics achievement means score points of the countries were 397, 331, 371, 351, and 425 for Botswana, Ghana, Morocco, South Africa, and Tunisia respectively. Thus, the achievement mean score of Ethiopian students in this study was compared against the mean score points of each of those African countries. As a result, the analysis of the finding indicates that the mean scale score (404.10) of Ethiopian students' achievement in mathematics was better than the mean scores of four African countries (i.e., Botswana, Ghana, Morocco, South Africa) students in TIMSS grade eight assessment. A one-sample t- test also showed that the mean differences were statistically significant at ($t_{(2049)} = 2.52$, $p < 0.05$), ($t_{(2049)} = 25.95$, $p < 0.001$), ($t_{(2049)} = 11.75$, $p < 0.001$) and ($t_{(2049)} = 18.85$, $p < 0.001$) for Botswana, Ghana, Morocco, South Africa respectively, which was in favour for Ethiopia. However, with regard to the comparison between Ethiopia and Tunisia, the achievement mean score of Tunisian students was higher than that of Ethiopian students, which was statistically significant at ($t_{(2049)} = -7.42$, $p < 0.001$). Nevertheless, in this case, one should have to be noticed that, the reason for the better achievement of the Ethiopian student higher than those four countries could be due to many of the test items (about

80%+) administered to the students were in a multiple-choice formats in which much more guessing could be possible to get the correct answer.

4.4. Mathematics Achievement in TIMSS Released Items Compared to NLA

Table 20 displays the comparison of students' achievement of percent raw score and scale score in TIMSS released mathematics test items with the NLA percent and scale mean scores of the fifth cycle study of 2016. That means it shows the comparison of students' achievement by the types of assessment (i.e., NLA result of 2016 and NLA result transformed from TIMSS released test items of this study or TIMSS-NLA).

Table 20. Mathematics scale scores by types of assessment across years

Assessment type by year		N	Mean	SD	t	df	MD	*p*
Raw scores (%)	NLA 2016	13376	33.95	15.92	5.08	2739	1.89	0.000
	TIMSS-NLA 2019	2050	32.05	15.66				
Scale scores	NLA 2016	13376	300.36	56.47	-1.13	2565	-1.69	0.257
	TIMSS-NLA 2019	2050	302.05	63.79				

When comparing the mathematics achievement of students in percent raw scores across the two types of assessment, the result of students in NLA 2016 was higher than that of TIMSS- NLA 2019 result of this study. In this case, the mean difference was statistically significant at ($t_{(2739)} = 5.08$, $p < 0.001$) according to the independent sample t-test, where Levene's test for equal variances was not assumed. However, comparing the achievement result through the raw score is not appropriate, as it is only depends on item characteristics rather than both on students' ability and test parameters.

In this regard, therefore, to compare the students' achievement in NLA and TIMSS-NLA released test items using a scale score, first the raw score of the lator converted to scale score using the standard set by NEAEA for grade eight NLA in 2016, where the mean and standard deviation were 300 and 50 respectively as mentioned earlier. Then the two tests were equated through IRT estimating parameters and students' ability using IATA software. As a result, the independent sample t-test indicated that the achievement of students in mathematics was slightly higher in TIMSS released test items than NLA 2016 in contrast to the achievement in raw score. However, the difference in the mean score (1.69) was not statistically significant at ($t_{(2565)} = -1.13$, $p = 0.257$), where Levene's test for equal variances was not assumed again. In this case, the inconsistency of the significant values in raw scores and scale scores was due to the result of equating and transforming percent raw score into scale score using IRT. Thus, it is possible to conclude that, the achievements of students both in NLA and TIMSS are similar.

4.5. Factors Affecting Students' Achievement in Mathematics

In this section, the results from a multilevel analysis of the influence of student and school-related factors on students' achievement in mathematics were presented. The information on model fit evaluation, Intraclass Correlation Coefficient (ICC), and standardized parameter estimates were scrutinized. Moreover, the hierarchal modeling analysis was also employed to look at the overall impact of student and school-related factors on students' achievement in mathematics.

4.5.1. Model Fit Evaluation and Analysis of Student-Related Factors

The findings of the student-related factors that affect students' achievements in mathematics were mainly contained about three thematic areas. These were factors related to student's family SES, students' attitude towards their school, and the mathematics subject. The student's attitudinal factors related to the subject matter include students' liking mathematics, interest in mathematics, and confidence in learning mathematics.

4.5.1.1. Model Evaluation for student-related factors

The multilevel structural equation modeling can be evaluated through various model fit indices. As mentioned under the data analysis section, the Chi-square test of model fit, RMSEA, CFI, and SRMR model fit tests were inspected for each student-related factor. In this case, the suggested cutoff values to accept or reject the model for tests of statistics both RMSEA and SRMR were less than 0.08 and greater than 0.95 for CFI as well as greater than 0.05 p- value for the Chi-square model fit, though it depends on the sample size of the study (e.g., Hooper, Couglan and Mullen, 2008). However, in the case of Chi-square, due to its' sensitiveness to the larger sample size, sometimes it is recommended that to consider the ratio of χ^2 value to its degree of freedom. In this case, although there is no consensus regarding an acceptable ratio for these statistic values, Tabachnick and Fidell (2007) and Wheaton, Muthen, Alwin, and Summers (1977) suggested the acceptance range from 0.5 to 2.0 lowest and highest respectively. In this regard, the model fitness of each of the student-related variables was evaluated as shown in Table 21.

Table 21.The model of fit test statistics for student-related variables

Variables	Chi-Square Test			RMSEA	CFI	SRMR	
	χ^2	df	p			Within	Between
Socio-economic status	0.15	4	0.997	0.000	1.000	0.001	0.007
Parental follow up	67.705	10	0.000	0.053	0.97	0.017	0.078
Attitude to school	2.9	4	0.575	0.000	1.000	0.003	0.029
Attitude to Maths	21.600	11	0.028	0.022	0.980	0.020	0.010
Confidence in Maths	55.561	36	0.02	0.016	0.991	0.015	0.077

Concerning the model fit for the socio-economic status variable, the finding showed that the test statistics for χ^2= 0.150, df= 4, p = 0.997, RMSEA (0.000), CFI (1.000) and SRMR (within=0.001, Between= 0.007) the data were well fit with the model. Similarly, with regard to the model fit statistics for the variables related to the students' attitude towards their school, the result reveals a good fitting of a two-level model in accordance with the χ^2= 0.290, df= 4, p = 0.575, RMSEA (0.000), CFI (1.000), and SRMR (Within = 0.003, Between = 0.029). In both cases, this model evaluation implies that the model inferred population variance-covariance matrix does not differ significantly from the observed variance-covariance matrix of the selected sample. In other words, there is no significant discrepancy between the data and the hypothesized model.

The model fits for other student-related variables such as parental follow-up in student learning, students' attitude toward mathematics, and their confidence in learning mathematics were also evaluated. As a result, the test statistics of CFI (0.970), RMSEA (0.053), SRMR (Within = 0.017, Between = 0.078) for the variable parental follow up; CFI (0.980), RMSEA (0.022), SRMR (Within = 0.020, Between = 0.010) for the latent variable attitude of towards mathematics and CFI (0.991), RMSEA (0.016) and SRMR (Within= 0.015, Between = 0.077) for the variable students self-confidence in learning mathematics were indicated as the model fits the data better at both the student and the school levels. In addition, the Chi-square test for each of these variables was: χ^2 = 67.705, df = 10, p <0.000 for parental follow up variable, χ^2 = 21.317, df = 10, p = 0.019 for students' attitude towards mathematics and χ^2 = 55.561, df = 36, p = 0.020 for the variable students self-confidence in learning mathematics. In this case, the findings for these three variables indicated statistically significant values for $p<0.05$, which violated the criteria stated above due to its' sensitivity to a larger sample size (N = 2050) in this study. Nevertheless, the ratio of χ^2 to its degree of freedom for each of the variables was within the acceptable range indicated above. Thus, it is possible to conclude that, the sample data was better to fit with the hypothesized model. That means the population variance-covariance matrix does not differ significantly from the observed variance-covariance matrix of the sample.

4.5.1.2. Students' family socioeconomic status

Intraclass Correlation Coefficient: Intraclass correlation (ICC) assesses the level of variance in the observed variable that is attributable to membership in its cluster. In this case, ICC captures the number of differences in a construct that is due to the students belonging to different schools. That means ICC describes the proportion of the total variance in each indicator that can be attributed to students attending different schools.

Table 22. Estimated Intraclass Correlation Coefficient of home background variables

Variables	ICC
Parent highest education level	0.22
Family possessions	0.16
Number of books at home	0.13
Mathematics achievement	0.28

Table 22 shows the ICCs pertinent for each of the observed variables, i.e., the indicators to home SES factor and achievement in mathematics. The ICC values for the observed variables ranged from 0.13 for the variable *the number of books at home* to 0.22 for the *parent highest education level*. These ICC measures for the SES indicators can be thought of as school segregation measures of the socio-economic status of school intake. The higher ICCs are the more intensive the school segregation is in terms of these indications of

SES. As we know that, different schools may attract students from different socio-economic backgrounds with their special profiles, very often, schools with high achievement levels and a good reputation will attract many students from highly educated and economically better families. Teachers too are more likely to find schools with students from a higher socio-economic background, and in tune ensure the highest teaching quality. This, in turn, leads to the students' difference in mathematic achievement between the schools. The between school difference captured by the ICC for Mathematics achievement, being about 0.28, indicated that, the probability for the students taught in the same school achieve similar in their mathematics is about 0.72. However, albeit the small size of the ICCs, they are substantial to indicate that students in the same school are much more similar in both their family SES background and mathematics achievement than those of students from a different school. In other words, the substantial amount between school differences deserve special attention since this part of the variance can reveal some important mechanisms at the school-level that are highly relevant for policy-makers.

Factor loadings and structural relations: Table 23 presents standardized estimates for factor loadings of observed indicators on the latent variable socioeconomic status at the student and school levels. SES is measured by *parental highest education level*, *family possessions,* and *the number of books at home*. In this regard, standardized estimates can be useful for comparing the size of the relationship between each indicator and a latent variable SES. Thus, the standardized factor loadings can be interpreted as the amount of variance of each SES indicator explained by the latent factor SES. It can also be noticed that the factor loadings vary across the SES indicators. At the individual-level, family SES is mainly represented by parental educational level while for the school SES composition of its students' intakes, the economic aspect of the family SES defines the concept. The school SES composition can be interpreted as neighbourhoods demographic circumstance or representing residential segregation if the students go to the nearby school where they live. However, if school choice is allowed in the school system, the school SES is the school segregation measure. Whatever the school policy is, school SES is a combined segregation measure of residential area and school, which has a potential effect on student achievement. In this finding, at both student and school levels, all the observed variables have significant and positive factor loadings, though the factor loadings at between school-level higher than those at the individual level. That means each of the indicators or observed variables was related to the latent variable SES that they assumed to measure it. Moreover, the R-square estimate which refers to the proportion of variance explained in the model indicated that for each observed variable there was were statistically significant values at both student and school levels, which implied that the latent variables explained a substantial amount of the variation in each SES indicator (see Table A in Appendix V for these R-square).

Table 23 Standardized estimates for observed variables and SES latent variable at student and school-level

Parameters	Estimate	S.E.	z	*p*
Student-Level (Within)				
SES indicated by				
Parent highest education level	0.550	0.096	5.742	0.000
Family possessions	0.342	0.052	6.577	0.000
Number of books at home	0.365	0.051	7.205	0.000
Regression coefficient of Math Achievement ON SES	-0.024	0.031	-0.758	0.448
School-Level (Between)				
School SES composition indicated by				
Parent highest education level	0.776	0.085	9.112	0.000
Number of books at home	0.707	0.111	6.388	0.000

Family possessions	0.895	0.094	9.564	0.000
Regression coefficient of school mean Math achievement ON school SES	0.453	0.180	2.516	0.012

Looking at the structural relations between family SES and mathematics achievement, a significant and positive relationship (0.453) was found at the school-level. However, it is insignificant at the student-level. The finding implies that schools with a higher average SES composition of their student intake achieve higher average mathematics scores and vice versa. However, for each individual student, his or her mathematics achievement is not significantly related to the family's socio-economic status. This is interesting because it is the school that the students attend that mainly makes a difference in their academic achievement than their family background at the individual-level. This result may indicate that schools are segregated with respect to student intake's socio-economic composition as well as school academic achievement.

4.5.1.3. Parental support and follow up in student learning

Intraclass Correlation Coefficient: Table 24 shows the ICCs related to each of the observed variables of parental support and follow-up in student learning. The latent variable parental support and follow up was represented by the observed variables that students responded to the items: *My parents ask me what I am learning in school, I talk about my schoolwork with my parents, My parents make sure that I set aside time for my homework, and My parents check if I do my homework* based on the Likert scale consists of four degree levels.

Table 24. Estimated Intraclass Correlation Coefficient (ICC) for parental follow up in student learning

Variables	ICC
My parents ask me what I am learning in school	0.09
I talk about my schoolwork with my parents	0.10
My parents make sure that I set aside time for my homework	0.11
My parents check if I do my homework	0.12

As indicated in Table 24, the ICC coefficient values for the observed variables of parental support and follow-up were in the range of 0.09 least and 0.12 which is relatively high for *My parents ask me what I am learning in school* and *My parents check if I do my homework* respectively. Overall, the values of ICC for all the variables were small. That means the similarity between the values of the variables from the same schools was not large. This indicates that students in the same school are more similar both in obtaining parental support or follow up and achieving in mathematics than those of students from different schools.

Factor loadings and structural relations: Table 25 presents the standardized estimates for factor loadings of observed indicators to the latent variable of parental follow-up for the students' learning both at student and school levels.

Table 25. Standardized estimates for observed and latent variables of parental follow up in student learning

Parameters	Estimate	S.E.	z	*p*
Student-Level (Within)				
Parental follow up BY				
My parents ask me what I am learning in school	0.703	0.021	34.188	0.000
I talk about my schoolwork with my parents	0.701	0.020	35.822	0.000
My parents make sure that I set aside time for my homework	0.707	0.022	31.960	0.000
My parents check if I do my homework	0.733	0.023	31.598	0.000
Mathematics Achievement ON				

Parental follow up	-0.010	0.027	-0.364	0.716
School-Level (Between)				
Parental follow up BY				
My parents ask me what I am learning in school	0.845	0.077	10.922	0.000
I talk about my schoolwork with my parents	0.800	0.085	9.431	0.000
My parents make sure that I set aside time for my homework	0.993	0.024	41.933	0.000
My parents check if I do my homework	0.936	0.050	18.677	0.000
Mathematics Achievement ON				
Parental follow up	-0.017	0.126	-0.134	0.893

From the result of this study, the model showed that, all the observed variables: My parents ask me what I am learning in school, I talk about my schoolwork with my parents, My parents make sure that I set aside time for my homework have a positive statistically significant factor loadings with the latent variable parental follow up for students learning at both student and school levels where $p<0.001$. That means each of the observed variable indicators was related to the latent variable parental follow-up of students' learning as they assumed. However, regarding the relationship between parental follow-up in students learning and achievement in mathematics, the findings indicated that there is no statistically significant association between them both at the student and school levels.

On the other hand, the R^2 estimate for each of the observed variables was statistically significant both at student and school levels. Thus, each of the indicators was related to the latent variable parental follow-up in student learning in such a way that they assumed to measure it. In this case, the model explains about 53% highest and 49% lowest for the observed variables: *My parents check if I do my homework and I talk about my schoolwork with my parents* respectively at school-level. Similarly, it explains about 97% highest and 64% lowest for the observed variables: *My parents make sure that I set aside time for my* *homework and I talk about my schoolwork with my parents* respectively at school-level (See Appendix V in Table B for details).

4.5.1.4. Students' Attitude towards their School

Intraclass Correlation Coefficient: Table 26 shows the ICC related to each of the observed variables of students' attitude towards their school. The latent variable of the attitude of students towards their school was represented by the observed variables related to the degree of students like their teachers, proud with their schools and eager to see their classmates at school that were measured by Likert scale.

Table 26. Estimated Intraclass Correlation Coefficient for the students' attitude towards their school

Variables	ICC
Liking their teachers	0.16
Proudness with their school	0.12
Eagerness to see their classmates at school	0.11

In this regard, the ICC coefficient values for the observed variables of the students' attitude towards *liking their teachers, proud of their school, and eagerness to see their classmates* were about 0.16, 0.12 and 0.11 respectively. The between-school difference specified by the ICC for Mathematics achievement was 0.28 which indicated that the probability for the students taught in the same school achieve similar in their mathematics is about 0.72. In this case, the values of ICC for all the variables were small. That means the similarity between the values of the variables from the same schools was not large. This indicates that students in the same school are much more similar in both their attitude towards their school and mathematics achievement than those of students from different schools.

Factor loadings and structural relations: Table 27 presents standardized estimates for observed and latent variables related to students' attitude towards their school at the student and school-level. The attitude of students towards their school was measured by the statements related to their schools, teachers, and classmates. These included: students' proudness of their school, liking their teachers in the school, and their eagerness to see their classmates at school.

Table 27. Standardized estimates for observed and latent variables related to students' attitude towards their school

Parameters	Estimate	S.E.	z	*p*
Student-Level (Within)				
Attitude to school BY				
Proud with their school	0.685	0.028	24.762	0.000
Eager to see their classmates at school	0.820	0.027	29.929	0.000
Liking teachers in their school	0.638	0.027	24.049	0.000
Mathematics Achievement ON				
Attitude to school	-0.032	0.027	-1.176	0.239
School-Level (Between)				
Average Attitude to school BY				
Proud with their school	0.946	0.041	22.921	0.000
Eager to see their classmates at school	0.981	0.037	26.540	0.000
Liking teachers in their school	0.794	0.068	11.646	0.000
Mathematics Achievement ON				
Attitude to school	0.360	0.112	3.215	0.001

According to Table 27, the results indicated that at both student and school levels all the observed variables have positive factor loadings that were statistically significant at p = 0.000. That means each of the observed variable indicators was related to the latent variable students' attitude towards their school. In this case, the attitudes of students towards their school were positively affecting their achievement in mathematics only at school levels with the factor loading of 0.360 and p = 0.00. That means, at the school-level, as students' attitude positively increased by one unit, their achievement in mathematics also increased by the rate of 0.360 estimate coefficient and vice versa. However, at the individual-level the impact of students' attitude towards the school on mathematics achievement was not statistically significant. Furthermore, the R^2 estimate for each observed variable related to the attitude of students towards their school also showed statistically significant values both at student and school levels (see Appendix V Table C for the detail).

4.5.1.5. Students' Attitude towards mathematics

Intraclass Correlation Coefficients: Table 28 shows the intraclass correlation coefficients related to each of the observed variables of students' attitude towards mathematics as well as achievement in Mathematics.

Table 28. Estimated Intraclass Correlation Coefficient for students liking Maths

Variable	ICC
I wish to study mathematics	0.17
I enjoy learning mathematics	0.07
I learn many interesting things in mathematics	0.06
It is important to do well in mathematics	0.06

The ICC values for the observed indicators ranged from the lowest value 0.04 for the statements: *It is important to do well in mathematics and I learn many interesting things in mathematics* to the highest value of 0.17 for the statement *I wish to study mathematics*. Each of the variables has a low ICC indicates a low level of dissimilarity of students' attitude towards liking to learn mathematics across different schools, except for the indicator *I wish to study mathematics*. Thus, it can be argued that attitude is more an individual trait than an aggregate level construct, and the variation mainly among students not across schools.

Factor loadings and structural relations: Table 29 presents standardized estimates for factor loadings of observed indicators to the latent variable students' attitude towards liking in learning mathematics at the student and school-level. The construct of liking learning mathematics was measured by a four Likert scaled statements *I enjoy learning mathematics, I wish to study mathematics, I learn interesting things in mathematics and It is important to do well in mathematics*.

Table 29. Standardized estimates for observed and latent variables in attitude towards liking Mathematics

Parameters	Estimate	S.E.	z	*p*
Student-Level (Within)				
Liking mathematics BY				
– I enjoy learning mathematics	0.450	0.038	11.688	0.000
– I wish to study mathematics	0.101	0.045	2.263	0.024
– I learn interesting things in mathematics	0.723	0.037	19.575	0.000
– It is important to do well in mathematics	0.549	0.042	13.184	0.000
Mathematics Achievement ON				
– Attitude toward Mathematics	0.049	0.022	2.248	0.025
School-Level (Between)				
Liking mathematics BY				
– I enjoy learning mathematics	0.698	0.147	4.744	0.000
– I wish to study mathematics	0.145	0.245	0.592	0.554
– I learn interesting things in mathematics	0.989	0.004	65.061	0.000
– It is important to do well in mathematics	0.756	0.138	5.49	0.000
Mathematics Achievement ON				
– Attitude toward Mathematics	-0.002	0.166	-0.007	0.994

The model results of the study showed that at both student and school levels all the observed variables have positive and significant factor loadings on the latent variable attitude towards learning mathematics except for the statement *I wish to study mathematics* at the school-level which was insignificant. This is odd, as ICC of this variable was highest than other variables which means there is more between-school differences in this variable. However, each of the remaining observed variable indicators was related to the latent variable students' liking mathematics as they assumed to measure it.

Regarding the relationship between students' attitude towards liking mathematics and achievement in mathematics a statistically significant and positive relationship (0.049) was found at the student-level. The finding implies that students who like mathematics more achieve better in average mathematics scores and vice versa at the individual-level. At the school level, the latent variable may be interpreted as the general school ethos towards mathematics, since if a great majority of students are interested and motivated in learning mathematics, it will form a positive school culture towards learning in general and mathematics learning in particular. However, for each school, mathematics achievement was not significantly related to their students' attitude towards liking mathematics. This result may indicate that it is an individual student's motivation and interests that makes difference in mathematics achievement than school.

Moreover, the R^2 estimate for each observed variable was statistically significant at both student and school levels except for the variable *I wish to study mathematics* both at student and school-level. In this case, the model explains about 52%, 30% and 20% variances for the statements: *I learn interesting things in mathematics, It is important to do well in mathematics and I enjoy learning mathematics* respectively at student-level. Likewise, at the school-level 97%, 57% and 49% of the variance explained by the statements: *I learn interesting things in mathematics*, *It is important to do well in mathematics and I enjoy learning mathematics and* in using this model respectively (For details see Appendix V of Table D).

4.5.1.6. Students' confidence in learning mathematics

Intraclass Correlation Coefficient: Table 30 indicates the intraclass correlation coefficient related to each of the observed variables of students' confidence in learning mathematics. In this case, the observed variables were include: *Mathematics is more difficult for me than other my classmates, I learn things quickly in mathematics, I would like a job that involves using mathematics, Mathematics makes me nervous, I usually do well in mathematics and Mathematics is harder for me than any other subject*.

Table 30. Estimated Intraclass Correlation Coefficient for the students' confidence in learning Mathematics

Variables	ICC
I usually do well in mathematics	0.13
Mathematics is more difficult for me than my classmates	0.09
I learn things quickly in mathematics	0.09
Mathematics makes me nervous	0.12
My teacher tells me that I am good at mathematics	0.13
Mathematics is harder for me than any other subject	0.10
I would like a job that involves using mathematics	0.09

From the finding, the ICC values for the observed variables were the lowest (0.09) for the statements: *Mathematics is more difficult for me than my classmates, I learn things quickly in mathematics and I would like a job that involves using mathematics.* However, the ICC value for the statements said: *I usually do well in Mathematics and My teacher tells me that I am good at Mathematics* was relatively high (0.13). Overall, each of the observed variables represented confidence of students in learning mathematics had a small size of ICC. This indicates the similarity in student's confidence in mathematics from the same schools is rather low. This might be due to the fact that students' mathematics self-confidence is a personal trait that is not affected by the shared teacher and school-related factors.

Factor loadings and structural relations: Student's mathematics self-confidence showed up to two-dimensional. One dimension is related to students' internal self-perception of their mathematics skills, where four of the items were related to students' confidence or motivation to learn mathematics. These items were: *I usually do well in mathematics, I* *learn things quickly in mathematics, My teacher tells me that I am good at mathematics and I would like a job that involves using mathematics*. The other dimension is about math anxiety and self-doubting of their mathematics skills in reference to their classmates, and it was measured by three items. Namely, *mathematics is more difficult for me than for many of my classmates, mathematics is harder for me than any other subject and mathematics make me nervous* were used to measure the lack of students' confidence in learning mathematics. The two dimensions are correlated in an oblique model.

Table 31. Standardized estimates for observed and latent variables for students' confidence in learning Mathematics and their achievement

Parameters	Estimate	S.E.	z	*p*
Student-Level (Within)				
Confidence in mathematics BY				
— My teacher tells me I am good at mathematics	0.572	0.024	24.195	0.000
— I learn things quickly in mathematics	0.671	0.029	23.201	0.000
— I usually do well in mathematics	0.707	0.028	25.548	0.000
— I would like a job that involves using mathematics	0.569	0.027	21.475	0.000
Lack of Confidence in mathematics BY				
— Mathematics is harder for me than any other subject	0.660	0.029	23.093	0.000
— Mathematics makes me nervous	0.545	0.035	15.783	0.000
— Mathematics is more difficult for me than my classmates	0.629	0.030	20.756	0.000
Mathematics Achievement ON				
— Confidence in mathematics	0.019	0.026	0.723	0.470
— Lack of confidence in mathematics	-0.033	0.028	-1.192	0.233
School-Level (Between)				
Confidence in mathematics BY				
— My teacher tells me I am good at mathematics	0.878	0.046	19.023	0.000
— I learn things quickly in mathematics	0.933	0.036	25.721	0.000
— I usually do well in mathematics	0.983	0.025	39.713	0.000
— I would like a job that involves using mathematics	0.867	0.062	13.890	0.000
Lack of Confidence in mathematics BY				
— Mathematics makes me nervous	0.933	0.036	25.621	0.000
— Mathematics is harder for me than any other subject	0.983	0.032	30.846	0.000
— Mathematics is more difficult for me than for many of my classmate	0.981	0.028	35.352	0.000

Mathematics Achievement ON				
− Confidence in mathematics	0.312	0.141	2.216	0.027
− Lack of confidence in mathematics	-0.093	0.174	-0.534	0.593

Table 31 shows the standardized estimates for factor loadings of observed indicators of students' mathematics self-confidence (latent variable) at the student and school-level. The attitude of students' confidence or lack of confidence in learning mathematics was measured by seven items with four Likert scaled.

With regard to the structural relations between students' confidence in learning mathematics and their achievement, the findings revealed that, at both student and school levels all the observed variables related to the positive or negative confidence of students in learning mathematics have positive factor loadings that were statistically significant at $p<0.01$. That means, each of the observed variable indicators was associated with the latent variables represented by confidence in learning mathematics and lack of confidence in learning mathematics. Moreover, both students' confidence and lack of confidence in learning mathematics were not significantly affecting the achievement of students in mathematics at the individual-level. Similarly, a student's lack of confidence does not significantly affect their achievement at school-level too. However, the confidence and motivation of the students in learning mathematics were positively affecting their achievement in mathematics at the school level with the factor loading 0.312 which was statistically significant at p = 0.027. That means, as the students' confidence positively increased by one unit in learning mathematics, their achievement in mathematics also increased nearly by 0.3 rates. In this case, the school-level average self-confidence in mathematics, analogous to student's attitude towards mathematics, represent a specific culture in relation to student's intakes overall self-confidence. Such a culture will positively affect school average mathematics achievement while the negative schools culture do in reverse

Furthermore, the R^2 estimate for each observed variable related to confidence and lack of confidence in learning mathematics were statistically significant at both student and school levels. In this case, at the student-level, the model explains about 50% (highest) and 30% (lowest) of variance for the statements 'I usually do well in mathematics' and 'mathematics makes me nervous' respectively. At the school level, the model also explains about 98% (highest) and 87% (lowest) of the variance for the statements 'I usually do well in mathematics' and 'I would like a job that involves using mathematics' variables in using this model respectively (see Appendix V in Table E).

4.5.2. Model Fit Evaluation and Analysis of School-Related Factors

This analysis was carried out based on the data from the school principals' responses at the school level. In this part, the analysis includes the variables related to school resources, school emphasis on academic success as well as school discipline and safety. Under this section, results of the model fit evaluation and standardized parameter estimates (i.e., factor loadings and structural relation) for each of the variables will be presented.

4.5.2.1. Model Fit Evaluation for school-related factors

Table 32.The model of fit test statistics for school related variables

Variables	Chi-Square Test			RMSEA	CFI	SRMR	
	χ^2	df	p			Within	Between
Lack of school resource	1.479	2	0.477	0.000	1.000	0.000	0.026
School emphasis on academic success	1.455	2	0.483	0.000	1.000	0.000	0.028

School discipline and safety	3.807	2	0.149	0.021	0.962	0.000	0.041

As shown in Table 32, the responses of school principals concerning the variables related to the school factors were evaluated by different model fit indices in the two-level model. As a result, the model fit statistics for the variable lack of school resources were found χ^2 = 1.479, df = 2, p = 0.477; CFI (1.000), RMSEA (0.000) and SRMR (Within= 0.000, Between = 0.026). Similarly, the model fit indices for school emphasis on academic success as well as school discipline and safety variables were χ^2 = 1.455, df = 2, p = 0.483, CFI (1.000), RMSEA (0.000) and SRMR (within= 0.000, between= 0.028) and of χ^2 = 3.807, df = 2, p = 0.149, CFI (0.962), RMSEA (0.021) and SRMR (Within= 0.000, Between= 0.041). This finding showed, there is a very good fit of the model with the sample data for all the three school-related variables. That means, the model fit statistics for all tests confirmed the best fit of the sample data with the hypothesized model.

4.5.2.2. Factor loadings and structural relations for lack of school resource variables

Table 33 shows the standardized estimates of the loading coefficients for the observed indicators related to the shortage of school resource factors (latent variables). The school resource observed indicators were represented by three constructs that measured on the four Likert scale. These observed variables were included in the measurement of *lack of instruction materials* (textbooks, teacher's guide, syllabus, concrete objects help to understand quantities, reference books in the library), *inadequate school facilities* (adequate school buildings and grounds, availability of electricity, adequacy of classrooms, presence of computer technology for teaching and learning) and *lack of competent teachers* like teachers specialized in mathematics.

Table 33. Standardized estimates for observed and latent variables related to school resources

Parameters	Estimate	S.E.	z-value	p-value
School-level				
Shortage of school resources BY				
– Lack of instructional Materials	0.941	0.107	8.789	0.000
– Inadequate school Facilities	0.809	0.114	7.080	0.000
– Lack of competent teachers	0.538	0.101	5.314	0.000
Mathematics achievement ON				
Shortage of school resources	-0.124	0.164	-0.756	0.450

Regarding the structural relationships between the observed variables and a latent variable or lack of school resources, the finding indicated that all the observed variables were positively associated with the latent variable. In this case, the factor loadings for *lack of instructional materials*, *inadequate school facilities, and lack of competent teachers* by the latent variable (lack of school resource) were 0.941, 0.809 and 0.538 respectively. This shows, there is a statistically significant relationship between these variables at $p < 0.01$. However, even though factors related to the shortage of school resources were negatively affecting the achievement of students in mathematics to some extent with the factor loading of -0.124 at school-level, the finding indicates that the influence was not statistically significant at $p = 0.450$. Thus, there is no evidence that indicates variability between the schools' achievement in mathematics is due to the lack of school resources as of this study. Furthermore, the R^2 estimate for each observed variable related to the shortage of school resources based on the response of the school principals was shown in Appendix V in Table G. As a result, the model explains about 89%, 65% and 29% for the lack of instructional materials, inadequate school facilities and lack of competent teachers respectively. In this case, the values of all the observed variables were statistically significant. The value for R^2 in Mathematics achievement was also explained by the model with about 2% which was not statistically significant.

4.5.2.3. Factor loadings and structural relations for school emphasis on academic success variables

Table 34 presents the standardized estimates of the factor loadings and structural relations for the observed indicators or factors related to the schools' emphasis on academic success as a latent variable and their impact on students' mathematics achievement. The schools' emphasis on academic success was represented by three construct observed indicators. The first construct was *teachers' emphasis on the academic success* that was measured by *teachers' understanding of the curricular objectives*, *their expectations for student achievement, and their ability to motivate students*. The second construct was *parents' emphasis on the academic success* that was represented by *parental involvement in school activities, their commitment to ensure that students were ready to learn, and their pressure for the school to maintain high academic standards*. The third construct was *students' emphasis on academic success* which was measured by *students' desire to do well in school and their ability to reach the set academic goals*.

Table 34. Standardized estimates for observed and latent variables related to schools' emphasis on academic success

Parameters	Estimate	S.E.	z	*p*
School-level				
Emphasis on academic success BY				
– Teachers emphasis	0.702	0.118	5.936	0.000
– Parents emphasis	0.741	0.118	6.272	0.000
– Students emphasis	0.548	0.169	3.250	0.001
Mathematics achievement ON				
Emphasis on academic success	0.339	0.150	2.263	0.024

In this regard, the findings indicated that as all the observed variables have positive factor loadings that were statistically significant at $p < 0.05$. That means, each of the observed variables allied to teachers, parents, and students were related to the latent variable (schools' emphasis on academic success) as they assumed to measure it.

Looking at the structural relations between schools' emphasis on academic success and students' achievement in mathematics, the finding indicated that, factors related to the schools' emphasis on academic success were positively affecting the achievement of students in mathematics with the factor loading value of 0.339, that was statistically significant at $p = 0.024$. This implies that, schools with a higher emphasis on academic success by their teachers, parents, and students achieve a higher score in mathematics or as the emphasis of teachers, parents, and students for academic success increased at school, the achievement of students in mathematics also increased and vice versa.

On the other hand, Table H in Appendix V indicates the R^2 estimate for each observed variable related to the emphasis for academic success based on the response of the school principals. As a result, the model explains about 49%, 55% and 30% for the teachers', parents' and students' emphasis on academic success at school respectively. In this case, while the estimated values of the observed variables for teachers and parents were statistically significant, the value for students' emphasis on academic success was statistically insignificant. In addition, the value for R^2 in Mathematics achievement was also insignificantly explained by the model with about 12%.

4.5.2.4. Factor loadings and structural relations for school disciple and safety variables

Table 35 presents the standardized estimates of factor loadings for the observed indicators related to the school discipline and safety (latent variable) and the structural relation between the latent variable and students' achievement in mathematics. The latent variables, school discipline, and safety were represented by three sub-construct observed indicators. The first construct was *school absenteeism* that was measured by *students' and teachers' unjustified absences.* The second was *classroom disturbance* that was represented by *students' classroom destruction and exam cheating* and the third was *verbal abuse and physical injury* which was measured by the presence of verbal abuse and physical injury among students at school.

Table 35. Standardized estimates for observed and latent variables related to school discipline and safety

Parameters	Estimate	S.E.	Z	*p*
School-level				
School discipline and safety BY				
– School absenteeism	0.600	0.124	4.823	0.000
– Classroom disturbance	0.826	0.109	7.568	0.000
– Verbal abuse and physical injury	0.801	0.114	7.029	0.000
Mathematics achievement ON				
– School discipline and safety	-0.070	0.171	-0.409	0.682

In this regard, the findings indicated that all the observed variables have positive factor loadings that were statistically significant at $p < 0.05$ with the latent variable. That means school absenteeism, classroom disturbance as well as verbal abuse and physical injury among the students were significantly measured school discipline and safety as they assumed.

With regard to structural relations between the latent variable and students' achievement in mathematics, the finding indicated that factors related to the school discipline and safety were negatively associated to the achievement of students in mathematics with the factor loading value of -0.07 which was not statistically significant at $p = 0.682$. That means the problems of discipline and safety are not the contributing factors for the variance of students' achievement in mathematics between school mean scores.

On the other hand, the R^2 estimate for each observed variable was statistically significant at $p < 0.05$, as indicated in Table I of Appendix V. In this case, the model explains about 36%, 62%, and 64% for the observed variables students' school absenteeism, verbal abuse, and physical injury and students' classroom disturbance respectively. However, the value for R^2 in mathematics achievement about 1% was not statistically significant at $p = 0.838$.

4.5.3. Analysis of Hierarchical Modelling for Factors Affecting Students' Achievement

The overall impact of student and school-related factors on students' achievement in mathematics as well as their interaction with mathematics achievement was found using the analysis of hierarchical modeling at various levels such as model 0, model 1, model 2 and model 3 as shown in Table 36. As a result, model 0 estimates the mean of the means of mathematics achievement for each school's (404.3) that was statistically significant. In this case, the estimates of covariance parameters indicated, there was more variation within the schools than among the different schools. In model 1, the finding from the hierarchical modeling indicated that all student-related factors such as students' family background variables (SES), family support (PSUPRT), liking mathematics (MATHATT), attitude of the students towards the school (SCHLATT) and confidence in learning mathematics

(MATCONF) that added to model 0 were not significantly affecting the achievement of students in mathematics at $p < 0.05$. Likewise, at model 2 when the three factors related to the schools such as lack of school resources (LSCHLRES), school emphasis on academic success (SEAS), school discipline and safety (DISPSAF) were also added to student-related factor their effects were also not statistically significant on students' achievement in mathematics at $p < 0.05$. In model 3, the interaction between some factors like student family background with student attitude to learning mathematics, family support, students' attitude towards the school, school emphasis on academic success, and lack of school resources (SES * MATHATT, SES * PSUPRT, SES * SCHLATT, SES * SEAS, SES * LSCHLRES); liking learning mathematics with confidence in learning mathematics (MATHATT * MATCONF) and parental support with school emphasis on academic success (PSUPRT * SEAS) were also revealed a statistically insignificant impact on students achievement in mathematics.

Table 36. The Estimates of Fixed Effects for student and school-related factors as their intersections on mathematics achievement

Parameter	Model 0	Model 1	Model 2	Model 3
Intercept	404.37	405.30	382.40	382.99
SES		1.04	1.96	1.49
PSUPRT		-0.05	0.62	0.75
MATHATT		1.97	3.25	3.08
SCHLATT		0.74	0.69	1.14
MATCONF		-2.53	-1.06	-0.93
LSCHLRES			-2.31	-2.06
SEAS			13.99	15.11
DISPSAF			-1.99	-2.35
PRIEXP			4.11	3.90
SES * MATHATT				-3.24
SES * PSUPRT				3.35
SES * SCHLATT				-4.72
MATHATT * MATCONF				-0.81
SES * SEAS				2.39
PSUPRT * SEAS				-2.65
SES * LSCHLRES				2.10
Within variance	11705.72	11382.07	11303.11	11303.94

Between variance	4624.53	4351.70	4579.21	4602.03
ICC	0.28	-	-	-
Individual-level explained variance	-	2.8%	0.7%	-
School-level explained variance	-	5.9%	-	-

Furthermore, the intraclass correlation coefficient for the Null model was calculated using a ratio of between school-level variance over the total variance (where the total variance is the sum between variance and within variance). The ICC for the outcome variable mathematics achievement was 0.28. This indicated that about 30% of the variation in mathematics achievement could be attributed to the students attending different schools. When adding individual-level predictors in the model (Model 1), about 3 % individual-level variance in math achievement and about 6 % school-level variation is explained by the model. For model 2 and model 3 - including school-level predictors and interactions - did not contribute much of the explanatory power of the model. In sum, the hierarchical linear model did not found important significance that significantly affects mathematics achievement. However, it should be cautious with the significant testing, since the number of schools is not high, we may have less statistical power to detect the effects that actually exist in the population from where our data were drawn. In other words, we should be cautious for conducting Type II errors.

CHAPTER FIVE: CONCLUSIONS AND RECOMMENDATIONS

This chapter consists of two sections. The first one presents the conclusions of the research works undertaken, and the second section forwarded some of the recommendations based on the results of this study.

5.1. Conclusions

The main objective of this study was to assess the achievement of grade eight students in mathematics through the lens of TIMSS released test items mathematics and find out factors that affect the students' achievement. The result of this study help policy-makers, concerned stakeholders, and scholars to have adequate information about the achievement of students, and factors affecting their achievement in mathematics for a better discussion and planning. It was a quantitative survey research design, in which the released TIMSS mathematics achievement test items and contextual questionnaires were employed. The participants were 53 primary schools, 53 school principals, and 2,120 students selected through two-stage cluster sampling techniques from the three regions. Descriptive and inferential statistics used to summarize the findings and generalize about a population from the samples.

The findings of this study were analyzed the overall achievement of students in mathematics as measured by the percent raw score and scale score that was low. The students' achievement raw mean score and NLA scale score were nearly close to the achievement raw and scale scores in NLA results in 2016. Nevertheless, the raw score was by far below the minimum standard of the passing mark (50%) set in the Ethiopian Education and Training Policy. At the same time, the TIMSS mean scale score 404.10 was also below the international center-point 500. The analysis of the raw score of students' achievements by mathematics content also revealed that as they faced difficulties in geometry content areas, while they relatively performed better in data or statistics content domains. This is similar results with that of various TIMSS studies, where students were more challenged in geometry content than others contents.

The achievements of students were also analyzed across gender, school location and region. The results of the study indicate that the mean difference of males was significantly better than females. This shows that still there is a gap between girls' and boys' in mathematics achievement. Thus, it needs more attention from all concerned stakeholders. Regarding the achievement of students by location, rural students achieved better than urban students and the mean difference was statistically significant. This contradicts with various latest findings in NLA where urban students perform better. The reason for this could be, while the data were gathered in some urban areas of Oromia and Benishangule Gumuz, there were instability and conflicts, so that students may not be well prepared for the assessment. However, it needs more investigation. Concerning the impact of instructional language, students were asked how frequently they speak it at home. The mean of mathematics achievement scores for those who speak the language of instruction at home *always* was higher than that of those who *never or sometimes* speak it at home, and the difference was statistically significant. This showed that the more the students learn through their first language, the more they achieved better in learning as several studies agreed upon it. Thus, while some regions provide the teaching process according to the education and training policy through the mother tongue of students up to grade eight, others are limited to primary schools only in the first cycle. Hence, the ministry of education and the rest of regional education bureaus should promote learning up to grade eight through the home language of students.

The mean differences in mathematics achievement for schools with small and large classroom sizes were analyzed. The finding indicates that the mean achievement scores for smaller size schools were slightly better than that of larger class size, though, the differences were not statistically significant. In this case, as the number of students per classroom is more than 50 in most urban areas of Ethiopia, it is important to reduce the number average class size to facilitate the instructional process and enhance students' learning. The achievement means a difference of schools across free school feeding programs was also considered in this study. The finding showed that the result of schools that provided free feeding for all students was better than those who feed some students with a significant mean score difference. Thus, the initiation of the government with regard to free school feeding should be continued and scaled up to all other schools.

At the regional level, Oromia was a relatively high achiever and Benishangul Gumuz was less achieved, while Addis Ababa was at the intermediate level. The mean score of Oromia was slightly above the total mean point (404.10) and the mean scores of the rest were below the total mean point. There is no statistically significant difference between Oromia and Addis Ababa, although their achievements were significantly different from Benishangul Gumuz region. This indicates that Benishangul Gumuz region will need more support than the two regions as it is one of the disadvantaged areas in the country.

The achievement mean score differences of schools by their principals' qualification and working experience category were also analyzed using a one-way ANOVA. The finding revealed that the achievement of schools those led by the principals who had the highest qualification with more working experiences performed better than those schools led by both lowest qualification and working experiences. Thus, the quality of the school principals can determine the quality of learning and achievement of schools. Hence, the government should pay attention in producing the qualified and experienced school leaders through the capacity-building programs.

In addition, the findings of this research also provide insights for the achievement of students in mathematics as compared to the mean score of African countries that participated in TIMSS 2011 and the international mean score or center-point (500). As a result, the mean achievement score of the Ethiopian students was found significantly lower than the international mean score like other African countries that participated in ILSAs. However, the mean achievement score of Ethiopian students was better than four African countries (Botswana, Ghana, Morocco, and South Africa) that participated in the study, where the difference was statistically significant. However, the mean achievement score of Ethiopian students was significantly below the mean score of Tunisian students, even though its performance was also below the international mean score of 500 points. In the same way, the result of Ethiopia is somewhat promising to participate in the TIMSS study like other African countries, though; its achievement is below the international mean scores as it is important to obtain the evidence that is useful to improve the education quality.

In comparing the Ethiopian students' achievement in TIMSS released mathematics test items that were converted to NLA scale score and the achievement scale score of the NLA study in 2016, the difference in mean score was not statistically significant. This revealed that the achievement of Ethiopian students in mathematics was almost similar in both national and international assessments. However, the study did not include a large enough number of constructed response items in the mathematics achievement test instrument due to most of the released items were not aligned with the syllabus.

Furthermore, using a multilevel modeling or a two-level model, this study was examined the influence of student and school-related factors on student achievement in mathematics based on the response of the respondents. Accordingly, among the students' related factors: home background variables or SES, such as parent educational level, family possessions or property and the number of books at home as well as gender and instructional language were significantly affected the achievement of students in mathematics. The positive attitude of students towards learning mathematics has also a positive impact on mathematics achievement at the student-level and vice versa. Similarly, students' attitude towards their school and their confidence in learning mathematics were the main student-related variables that affect the achievement of students at the school level. In this case, the ICC in each of these variables and mathematics achievement was small. In other words, the similarity between the values of the variables from the same schools was not large. This showed that students in the same school are much more similar in student-related factors and mathematics achievement than those of students from different schools.

The influence of school or classroom related variables to mathematics achievement were the other findings obtained from the response of school principals. In this regard, although various factors had their impact on the students' achievement, in this study, the major factor that significantly affects the achievement of students in mathematics was the emphasis of schools on academic success. Specifically, the main observed variables identified under these latent variables were related to teachers', parents', and the students' emphasis on academic success. Some of them were: teachers' understanding of the curricular objectives, expectations for student achievement and their ability to motivate students; parents' involvement in school activities, the commitment to ensure that students were ready to learn, and their pressure on the school to maintain high academic standards as well as students' desire to do well in school and ability to reach the set academic goals. Therefore, as the emphasis of various stakeholders increased on academic success, the achievement of students also increased and vice versa at the school level. In general, looking at the impact of all students and school-related factors, the analysis of the hierarchical model at various levels revealed as their influence was not substantial on the students' achievement both at individual and school levels.

5.2. Recommendations

The research findings of this study are vital for policy-makers and all concerned stakeholders, teachers, parents, and students. Therefore, the following recommendations stemming from this study may contribute to the improvement of students' learning in mathematics.

The overall findings of this study indicate that the achievement of students' in mathematics is below the expected standard similar to several previous studies in Ethiopia. Thus, to improve the achievement of students in learning mathematics and make them internationally competent, all the concerned governmental institutions, non-governmental organizations and concerned stakeholders are required to make a rigorous effort. Particularly, the government structures in the education sector from the ministry of education to school levels, teachers, parents, and students should play an important role. In this regard, the following specific points should be taken into considerations:

Support learners from economically disadvantaged families: variables related to the family socio-economic status were one of the problems that affect the achievement of students' in mathematics. Thus, to enhance the success of students in learning mathematics, the government should assist at least those students who came from economically disadvantaged families by expanding school feeding program that has already started in a few schools and providing necessary instructional materials in collaboration with different stakeholders and social agents.

Enhance learners' mathematics attitudes and self-concept: From this study, it was found that, high achiever students' have positive attitudes towards their schools and mathematics. In this regard, government officials at each level of education system should coordinate efforts to increase interest in educational achievement in mathematics and encourage learners to engage in mathematics and science activities at schools. Moreover, parents, teachers, and school guidance, and counselors need to provide more convincing evidence that there is an economic payoff when one follows mathematics-related careers and enable them to realize the usefulness of mathematics as well attending the school in general. Moreover, teachers need to use cooperative learning strategies in their lessons. Since, these learning strategies allow problem-solving to continue, when an individual member might have encountered an impasse.

Minimize the disparity in mathematics achievement across certain subgroups: There was a statistically significant mean difference in mathematics achievement across various subgroups such as gender, location, and regions. So, bridging the gaps and give special attention to those disadvantaged and low performed subgroups should be considered. Particularly, in this study, the achievement of girls in mathematics was found lower than that of boys, even though, there have been efforts made to promote girls' education at all grade levels in Ethiopia to change the prevailing situation where gender stereotypes are rife. Thus, to alleviate this problem, the Ministry of Education and Regional Educations Bureaus should strengthen the special supports for girls that already implemented and search for additional measures that could be helpful to enhance their learning. In this case, the intervention programs which aim to facilitate girls' achievement should be also tailored to include the involvement of different social agents. In addition, the implementation of different policies, initiatives, and programs relating to the affirmative actions of females should be regularly monitored and evaluated.

Regarding the achievement of students' by location, rural students achieved better than urban, this contradicts with several studies undertaken in Ethiopia. To find out the reasons for the low performance of urban students, there is a need to carry out detailed investigations to come up with specific recommendations, as they are in a better position of access to various educational resources compared to rural students. Furthermore, policy-makers should focus their attention on the diverse performance in mathematics among the regions and differentiated support mechanism for the least performed region (Benishagul Gumiz) required, as it is one of the socially and economically disadvantaged emerging regions. In general, the Ministry of Education and/or the Regional Education Bureaus should track differences regularly and take preventive measures which can minimize the gap among these various subgroups.

Promote learning through first languages: There is a need to encourage learning through local languages than English in order to attain better achievements of students. In this study, the mean score of achievement in mathematics for those students who speak the language of instruction at their home *always* was higher than that of students who *never or sometimes* speak the language at home, which was statistically significant. Thus, the finding of this study coincides with the present-day pedagogical conclusions and with the Ethiopian Education and Training Policy which promoting learning through the mother tongue. Thus, encouraging students to learn through their first language as much as possible is vital and a human right aspect, particularly for those in Benishangul region and Addis Ababa city Administration primary schools.

Increase school emphasis on academic success: In the present study, the emphasis of schools on academic success was positively affecting the achievement of students in mathematics significantly. That means schools with a higher emphasis on academic success by their teachers, parents, and students were achieved higher in mathematics. Therefore, the Ministry of Education, Regional Education Bureaus, and schools should motivate and encourage teachers and students to put the whole of their attention on academic success, in addition, to build a partnership with the parents to develop mutual responsibility for the attainment of students' in mathematics.

Improve the quality of school principals: The highest qualification and working experience of the school principals was found affecting the achievement of schools in mathematics, though they were statistically insignificant. Therefore, the government should try to recruit, train and assign professionally qualified and experienced school principals, as they are the key actors to provide quality education. Moreover, it is suggested that less qualified principals should be encouraged to attend further high education levels, through various short-term and long-term programs. For those less experienced school principals, the experience sharing programs with the best-performed schools should be arranged.

Encourage the participation of Ethiopia in ILSAs: In this study, Ethiopian students were performed better than some African countries that participated in the TIMSS study, though; the data gathering instruments were mainly limited to multiple response items. Besides, the achievements of Ethiopian students both in the NLA and TIMSS studies were also similar. Therefore, to revise policies, adjust curriculum, and target teacher training as well as to increase international competitiveness, it is promising for Ethiopia to participate in ILSAs. However, before participating in ILSAs, it is suggested to conduct a curriculum or content analysis for each subject areas across ILSA assessment frameworks and revise it accordingly, since some variations were shown in the mathematics curriculum in this study. Moreover, the researcher recommended, if Ethiopia will be participating in the ILSAs to benchmark and /or select appropriate programs that were designed for developing countries such as SACMEQ, LaNA or PISA-D, rather than joining international-large scale assessments that were developed primarily for upper-middle-income and high-income countries..

Further researches will be conducted in this study area: This study has certain limitations and left with many unanswered questions. Thus, further research is needed to explore the problem on a large sample from more scattered geographical regions in the countrywide. In this study, since the model explained moderate amounts of variance, it is recommended that the effects of student and/or school-level factors on mathematics achievement to be investigated to determine whether these factors influence Ethiopian students' mathematics achievement to a greater extent. Finally, since this study had used questionnaires as the only data collection instruments, it is recommended to use some qualitative techniques such as interviews, observation, and policy document analysis, to validate some of the findings suggested here.

References

Abaineh Ergago. (2012). *Factors influencing students' Learning of Mathematics: The case of First Cycle Secondary Schools of Hadiya Zone* (Thesis, Addis Ababa University). Retrieved from http://etd.aau.edu.et/bitstream/handle/123456789/282/Abaineh%20%20Ergogo.pdf?sequence=1&isAllowed=y

Abdul-Hamid, H., Abu-Lebdeh, K. & Patrinos, H. (2011). Assessment testing can be used to inform policy decisions: The case of Jordan. Policy Research Working Paper; The World Bank Human Development Network Education Team, Washington, DC.

Abera Wirtu. (2017). *Factors affecting the academic achievement of lower secondary school students: At Gata High School in Sigimo woreda of Jimma zone* (Thesis, Addis Ababa University). Retrieved from http://etd.aau.edu.et/bitstream/handle/123456789/17257/Abera%20Wirtu.pdf?sequence=1&isAllowed=y

Abiy Yohannes. (2017). *The effect of School Feeding Program on the school performance of primary public-school children in Arada Sub City, Addis Ababa* (Thesis, Addis Ababa University). Retrieved from http://etd.aau.edu.et/bitstream/handle/123456789/1176/Abiy%20Yohannes.pdf?sequence=1&isAllowed=y

Addey, C. (2015). What drives participation in international educational assessments? Laboratory of International Assessment Studies. Retrieved November 19, 2018, from http://international-assessments.org/what-drives-participation-in-international-educational-assessments/.

Addey, C., & Sellar, S. (2019). Is it worth it? Rationales for (Non) participation in international large-scale learning assessments. Education Research and Foresight Working Papers Series, No. 24. Paris, UNESCO. https://en.unesco.org/node/268820

Adewuyi, T., Taiwo, O., & Olley, B. (2012). Influence of Examination Anxiety and Self-Efficacy on Academic Performance among Secondary School Students. *IFE PsychologIA*, *20*(2), 60–68.

Ado, T. (2015). Influence of Learning Environment on Students' Academic Achievement in Mathematics: A Case Study of Some Selected Secondary Schools in Yobe State – Nigeria. *Journal of Education and Practice*, *6*(34), 5.

Adrogué, C., & Orlicki, M. (2013). *Do In-School Feeding Programs Have an Impact on Academic Performance? The Case of Public Schools in Argentina*. *21*(50), 23.

Afana, Y., Lietz, P., & Tobin, M. (2013). The relationship between school resources and grade 8 mathematics achievement: A comparison of Palestinian Authority, Israeli Hebrew and Israeli Arab schools in TIMSS 2007. *Journal for Educational Research Online*, *5*(1), 59–89.

Afridi, F., Barooah, B., & Somanathan, R. (2014). *School Meals and Classroom Attention: Evidence from India*. 28.

Ahmed, A. (2004). *Impact of feeding children in school: evidence from Bangladesh*.

Aggarwala, N. (2004). Quality Assessment of Primary and Middle Education in Mathematics

and Science (TIMSS). Takoma Park, MD: Eaton-Batson International.

Ajayi, K., Lawani, A., & Salomi, M. (2012). The influences of self-concept and academic motivation on students' attitude to mathematics in selected secondary schools in Ogun State, Nigeria. *European Journal of Scientific Research*, *67*(3), 444–455. Retrieved from Scopus.

Aliyu, A., Bello, M., Kasim, R., & Martin, D. (2014). Positivist and Non-Positivist Paradigm in Social Science Research: Conflicting Paradigms or Perfect Partners? *Journal of Management and Sustainability*, *4*(3). Retrieved from, http://www.ccsenet.org/journal/index.php/jms/article/view/39893

Allen, M. (2004). Assessing Academic Programs in Higher Education. San Francisco: J Ossey-Bass.

Allison, H. & Hobbs, R. (2006). Science and policy in natural resource management: Understanding system complexity. Cambridge University Press.

Amogne, A. (2018). *Does School Type Matter in Academic Achievement in Regional Examination? Comparative Study of Private and Government Schools in Dessie Administrative Town, North Central Ethiopia*.

Anna, C, &, Elizabeth, D. (2016). The Analysis of International Large-Scale Assessments to Address Causal Questions in Education Policy. Retrieved November 5, 2018, from https://www.utsc.utoronto.ca/people/dhuey/wp-content/uploads/sites/30/2017/05/ChmielewskiDhuey_2017.pdf

Appiah, K. (2001). The significance of international comparison of student achievement. West African Exams Council, Accra-Ghana.

Arslan, H., Çanlı, M., & Sabo, H. (2012). *A research of the effect of attitude, achievement, and gender on mathematic education*. *Acta Didactica Napocensia* 5(1), 8.

Arulmoly, C., & Branavan, A. (2017). The relationship between parental involvement and students' achievement in mathematics and science in Batticaloa educational zone, Sri Lanka - An International Multidisciplinary Research Journal. *South Asian Academic Research Journals*, *7*(4). Retrieved from https://www.researchgate.net/deref/http%3A%2F%2Fdx.doi.org%2F10.5958%2F2249-7137.2017.00046.5

Asrat Dagnew (2017). The relationship between students' attitudes towards school, values of education, achievement motivation, and academic achievement in Gondar secondary schools, Ethiopia. *Research in Pedagogy*, *7*(1), 30–42. https://www.researchgate.net/deref/http%3A%2F%2Fdx.doi.org%2F10.17810%2F2015.46

Azubuike, O. (2015). *Impact of Different Types of Schooling on Achievement in the School System: Evidence from Ethiopia* (Dissertation, University of Sussex). Retrieved from https://www.younglives.org.uk/sites/www.younglives.org.uk/files/ESP-WP-No%2067-09-08-2015-Azubuike.pdf

Backhoff, E. (2013), "Validity issues in international large-scale assessment (ILSA) programs:thoughts for developing countries", in Chatterji, M. (Ed.), Validity and Test Use:An International Dialogue on Educational Assessment, Accountability, and Equity, Emerald Group Publishing, Bingley, pp. 233-249.

Badri, M. (2018). School emphasis on academic success and TIMSS science/math Achievements. *International Journal of Research in Education and Science* (IJRES), 5(1), 176-189

Beaton, A., Mullis, I., Martin, M., Gonzalez, E., Kelly, D. & Smith, T. (1996).

Mathematics Achievement in the Middle School Years: IEA'S TIMSS, Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.

Beatty, A. & Pritchett, L. (2012). "From Schooling Goals to Learning Goals: How Fast Can Student Learning Improve?" CGD Policy Paper 012. Washington DC: Center for Global Development. https://www.cgdev.org/publication/schooling-goals-learning-goals-how-fast-can-student-learning-improve

Beller, M. (2013). Technologies in Large-Scale Assessments: New Directions, Challenges, and opportunities. In M. Davier, E. Gonzalez, I. Kirsch& K. Yamamoto (eds.). The Role of International Large-Scale Assessments: Perspectives from Technology, Economy, and Educational Research (pp.), DOI 10.1007/978-94-007-4629-9.

Bernard, J., & Michaelowa, K. (n.d.). *Managing the Impact of PASEC Projects in Francophone Sub-Saharan Africa*. 13, CONFEME, https://www.ipz.uzh.ch/dam/jcr:00000000-2029-c978-0000-000056a73686/25-paper.pdf

Berova, M. & Matusova, S. (2000). Slovak Republic. In: Robitaille, D.F., Beaton, A.E., and Plomp, T. (Eds.), The Impact of TIMSS on the Teaching & Learning of Mathematics & Science (pp. 133–138). Vancouver, Canada: Pacific Educational Press

Bofah, A., & Hannula, M. (2015). TIMSS data in an African comparative perspective: Investigating the factors influencing achievement in Mathematics and their psychometric properties. Helsinki, Finland.

Bolotov, V., Kovaleva, G., Pinskaya, M. & Valdman, I. (2013). Developing the Enabling

Context for Student Assessment in Russia. Systems Approach for Better Education Results (SABER) country report; 2013, SABER student assessment working paper; no. 9. World Bank, Washington, DC.

Borg, W., & Gall, M. (1989). Educational research: An introduction (5th Ed.). New York, NY: Longman. ISBN: 0-801-0334-6 [LB1028.B6 1989]

Bos, K. (2002). Benefits and limitations of large-scale international comparative achievement studies. The case of IEA's TIMSS study. Enschede: Unpublished Thesis, University of Twente.

Braun, H., Kanjee, A., Bettinger, E., & Kremer, M. (2006). *Improving education through assessment, innovation, and evaluation*. http://hdl.handle.net/20.500.11910/6609

Braun, H., & Kanjee, A. (2006). Using assessment to improve education in developing nations. In J. E. Cohen, D. E. Bloom, & M. B. Malin (Eds.), Educating all children: A global Agenda (pp. 303–353). Cambridge, MA: MIT Press.

Bray, M., Adamson, B., & Mason, M. (Eds.). (2007). *Comparative education research: Approaches and methods* (1. ed). Hong Kong: Comparative Education Research Center.

Brese, F., Mirazchiyski, P., Educational Testing Service (ETS), & International Association for the Evaluation of Educational Achievement (IEA). (2013). *Issues and methodologies in large-scale assessments: Special issue 2 measuring students family background in large-scale international education studies.*

Bryman, A. & Bell, E., 2011. Business research methods. Oxford University Press, USA

Bush, W. (2005). Improving research on Mathematics learning and teaching in rural contexts. *Journal of Research in Rural Education*, 20(8):1-11, (Retrieved May 20, 2013, retrieved from, http://jrre.psu.edu/articles/20-8.pdf.

Çakır, L. (2014). The Relationship between Underachievement of Gifted Students and their Attitudes toward School Environment. *Procedia - Social and Behavioral Sciences*, *152*, 1034–1038. https://doi.org/10.1016/j.sbspro.2014.09.269

Card, D., & Krueger, A. B. (1996). School Resources and Student Outcomes: An Overview of the Literature and New Evidence from North and South Carolina. *Journal of Economic Perspectives*, *10*(4), 31–50.

Cariola, L., Covacevich, C., Gubler, J., Lagos, E & Ortiz, M. (2011) Chilean participation in IEA studies. In: Papanastasiou C, Plomp T, Papanastasiou E (eds) IEA 1958–2008: 50 years of Experiences and Memories. Volume 1. Nicosia: Cultural Center of the Kykkos Monastery, pages 373–387.

Carr, H. D. (2016). *An Analysis of Varying Instructional Time and the Association With Third Grade Reading and Mathematics Proficiency on the Pennsylvania System of School Assessment* (Thesis, Indiana University of Pennsylvania). Retrieved from https://pdfs.semanticscholar.org/4e05/81dce7f99c18e0f5432642f94c268b414ba.pdf

Carson, D., Gilmore, A., Perry, C., & Gronhaug, K. (2001). *Qualitative Marketing Research*. https://doi.org/10.4135/9781849209625.\

Cattaneo, M. A., Oggenfuss, C., & Wolter, S. C. (2017). The more, the better? The impact of instructional time on student performance. *Journal of Education Economics*, *25*(5), 433–445. https://doi.org/10.1080/09645292.2017.1315055

Chamberlain, M. & Aalst, I. (2000). New Zealand. In: Robitaille, D.F., Beaton, A.E., and

Plomp, T. (Eds.), The Impact of TIMSS on the Teaching & Learning of Mathematics & Science (pp. 98–103). Vancouver, Canada: Pacific Educational Press.

Chepete, P. (2008). Modeling of the Factors Affecting Mathematical Achievement of Form 1 Students in Botswana Based on the 2003 Trends In International Mathematics and Science Study. Dissertation, Indiana University, Bloomington.

Cherinet Tuge (2009). Mathematics Curriculum, the Philosophy of Mathematics and its Implications on Ethiopian Schools Mathematics Curriculum. *Ethiopian Journal of Education and Sciences*, *4*(1). https://www.ajol.info//index.php/ejesc/article/view/42996

Chevalier, A., Gibbons, S., Thorpe, A., Snell, M., & Hoskins, S. (2009). Students' academic self-perception. *Economics of Education Review*, *28*(6), 716–727. https://doi.org/10.1016/j.econedurev.2009.06.007

Chinapah, V., H'ddigui, E. M., Kanjee, A., Falayajo, W., Fomba, C. O., Hamissou, O., Rafalimanana, A., & Byomugisha, A. (2000). *With Africa for Africa. Towards Quality Education for All. 1999 MLA Project*. Human Sciences Research Council, Private Bag X41, Pretoria, South Africa, 0001; Web site: http://www. https://eric.ed.gov/?id=ED444931.

Chinapah, V. (2003). Monitoring Learning Achievement (MLA) Project in Africa. Paris: ADEA.

Chowa, G., Masa, R., Ramos, Y., & Ansong, D. (2015). How do student and school characteristics influence youth academic achievement in Ghana? Hierarchical linear modeling of Ghana YouthSave baseline data. *International Journal of Educational*

Development, *45*, 129–140. https://www.sciencedirect.com/science/article/abs/pii/S0738059315001157

Christensen, L., Johnson, B., & Turner, L. (2013). *Research methods, design, and analysis* (12th ed). Boston, Munich: Pearson.

Chu, J. H., Loyalka, P., Chu, J., Qu, Q., Shi, Y., & Li, G. (2015). The impact of teacher credentials on student achievement in China. *China Economic Review*, *36*, 14–24. https://doi.org/10.1016/j.chieco.2015.08.006

Creemers, B. & Kyriakides, L. (2008). The dynamics of educational effectiveness: A contribution to policy, practice, and theory in contemporary schools. London: Routledge.

Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334. https: doi.org/10.1007/BF02310555.

Crosnoe, R., Johnson, M., & Elder, G. (2004). School Size and the Interpersonal Side of Education: An Examination of Race/Ethnicity and Organizational Context. *Social Science Quarterly*, *85*(5), 1259–1274. https://doi.org/10.1111/j.0038-4941.2004.00275

Dağlı, Ü. (2019). Effect of increased instructional time on student achievement. *Educational Review*, *71*(4), 501–517. https://www.tandfonline.com/doi/full/10.1080/00131911.2018.1441808

Daniel Basazen & Bamlaku Alamirew (2017). *Analysis of Factors Influencing Academic Achievement of Students in Some Selected Schools in Addis Ababa*. Retrieved from https://core.ac.uk/download/pdf/234641175.pdf

Deno, S. L. (2003). Developments in Curriculum-Based Measurement. *Journal of Special Education*, *37*(3), 184–192.

Department of Basic Education (2017). The SACMEQ IV Project in South Africa: A Study of the Conditions of Schooling and the Quality of Education, Short report,http://www.sacmeq.org/sites/default/files/sacmeq/publications/sacmeq_iv_project_in_south_africa_report.pdf, Republic of South Africa.

Di Giacomo, T., Fishbein, B. & Buckley, V. (2013). International Comparative Assessments: Broadening the Interpretability, Application, and Relevance to the United States. Research in Review 2012-5, College Board.

Diop, M. (2015). Internationally comparable learning tests pave the way for education reform inAfrica.https://blogs.worldbank.org/africacan/internationally-comparable-learning-tests-pave-the-way-for-education-reform-in-africa.

Dolata, S., Ikeda, M., Murimba, S. (2004). Different pathways to EFA for different school systems. IIEP newsletter (Paris) January–March 2004, vol. 22, no. 1, pp. 8–9.

Dwyer, C. (2008). The future of assessment: Shaping teaching and learning. New York: Lawrence Erlbaum Associates.

Eklöf, H. (2010). Skill and will: Test-taking motivation and assessment quality. *Assessment in Education: Principles, Policy & Practice*, *17*(4), 345–356. https://doi.org/10.1080/0969594X.2010.516569

Elley, W. (2005). How TIMSS-R contributed to education in 18 developing countries. Prospects 35(2): 199–212.

Ending the Torment: Tackling Bullying from the Schoolyard to Cyberspace. (2016).

https://doi.org/10.18356/27a397d3-en.

Engin-Demir, C. (2009). Factors influencing the academic achievement of the Turkish urban poor. *International Journal of Educational Development*, *29*(1), 17–29. https://doi.org/10.1016/j.ijedudev.2008.03.003

Ernest, P., Skovsmose, O., van Bendegem, J., Bicudo, M., Miarka, R., Kvasz, L., & Moeller, R. (2016). The Philosophy of Mathematics Education. In P. Ernest, O. Skovsmose, J. P. van Bendegem, M. Bicudo, R. Miarka, L. Kvasz, & R. Moeller (Eds.), *The Philosophy of Mathematics Education* (pp. 1–26). https://doi.org/10.1007/978-3-319-40569-8_1

Fardin, D., Alamolhodaei, H., & Radmehr, F. (2011). *A Meta -Analyze on Mathematical Beliefs and Mathematical Performance of Iranian Students*. *International Research Journals Vol. 2(4) pp. 1051-1058*. Retrieved from https://profdoc.um.ac.ir/articles/a/1022014.pdf

Finlayson, M. (2009). *The Impact of Teacher Absenteeism on Student Performance: The Case of the Cobb County School District* (Thesis, Kennesaw State University). Retrieved from https://digitalcommons.kennesaw.edu/etd/4

Gil-Flores, J., Padilla-Carmona, T. & Suárez-Ortega, M. (2011). Influence of gender, educational attainment, and family environment on the educational aspirations of secondary school students. *Educational Review*, 63(3), 345–363.Retrieved from https://www.tandfonline.com/doi/abs/10.1080/00131911.2011.571763

Goddard, R., Sweetland, S., & Hoy, W. (2000). Academic Emphasis of Urban Elementary

Schools and Student Achievement in Reading and Mathematics: A Multilevel Analysis. *Educational Administration Quarterly*, *36*(5), 683–702. https://journals.sagepub.com/doi/10.1177/00131610021969164

Greaney, V. & Kellaghan, T. (2008). Assessing national achievement levels in education. National assessments of educational achievement. Vol. 1. Washington, DC: World Bank.

Guba, E. & Lincoln, Y. (1994). Competing paradigms in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), Handbook of qualitative research (pp. 105-117). Thousand Oaks, CA: Sage.

Haan, D. (1992). Measuring curriculum-test overlap. Doctoral Dissertation, Enschede, University of Twente.

Habtamu Shankut. (2017). Factors Affecting Learning Mathematics in the Case Assosa University College of Natural Science. *International Journal of Education, Culture and Society*. Vol. 2, No. 1, pp. 6-12. DOI: 10.11648/j.ijecs.20170201.12

Hailu Muleta & Habtamu Garoma. (2016). The assessment of grade10 students' performance in Mathematics: The case of Jimma Zone. *International Journal of Current Research*, 8(9), 38620-38627. https://www.journalcra.com/sites/default/files/issue-pdf/17636.pdf

Hailu Nigus Weldeana. (2016). Ethno-mathematics in Ethiopia: Futile or Fertile for Mathematics Education? *Momona Ethiopian Journal of Science*, *8*(2), 146. https://doi.org/10.4314/mejs.v8i2.4.

Halai, A., & Tennant, G. (Eds.). (2016). *Mathematics Education in East Africa: Towards*

Harmonization and Enhancement of Education Quality. https://doi.org/10.1007/978-3-319-27258-0

Hanushek, E. & Woessmann, L. (2012). "Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation," *Journal of Economic Growth,* Springer, vol. 17(4), pp.267-321.

Hanushek, E. & Luque, J. (2003). Efficiency and equity in schools around the world. *Economics of Education Review*, *22*(5), 481–502. https://doi.org/10.1016/S0272-7757(03)00038-4

Hanushek, E. & Woessmann, L. (2017). School Resources and Student Achievement: A Review of Cross-Country Economic Research. In M. Rosén, K. Yang Hansen, & U. Wolff (Eds.), *Cognitive Abilities and Educational Outcomes* (pp. 149–171). https://doi.org/10.1007/978-3-319-43473-5_8

Heyneman, S., & Lee, B. (2014). The impact of international studies of academic achievement on policy and research. In L. Rutkowski, M. von Davier, & D. Rutkowski (Eds.), Handbook of international large-scale assessment: Background, technical issues, and Methods of Data Analysis (pp. 37–72). Boca Raton, Florida: Taylor and Francis Group.

Ho, H., Chen, W., Tran, C., & Ko, C. (2010). Parental Involvement in Taiwanese Families: Father-Mother Differences. *Journal of Childhood Education*, *86*(6), 376–381. https://doi.org/10.1080/00094056.2010.10523173

Holmlund, H., McNally, S., & Viarengo, M. (2008). Does Money Matter for Schools? *Institute for the Study of Labor (IZA), IZA Discussion Papers*, *29*. http://ftp.iza.org/dp3769.pdf

Howie S. (2006). Multilevel factors affecting the performance of South African pupils in Mathematics. In SJ Howie & T Plomp (eds). Contexts of Learning Mathematics and Science: Lessons learned from TIMSS. London: Routledge.

Howie, S. (2011). The involvement of African countries. In: IEA 1958–2008:50 Years of Experiences and Memories, (eds.), In C. Papanastasiou, T. Plomp, and E. Papanstasiou, 289–316. Nicosia: Cultural Center of the Kykkos Monastery.

Huang, F. L. (n.d.). The role of socioeconomic status and school quality in the Philippines: Revisiting the Heyneman–Loxley effect. *International Journal of Educational Development*, *30*(3), 288–296.

Huck, S. W. (2012). *Reading statistics and research* (6th ed). Boston: Pearson.

Hudson, A., & Ozanne, J. (1988). Alternative Ways of Seeking Knowledge in Consumer Research | Journal of Consumer Research | Oxford Academic. *Journal of Consumer Research*, *14*(4), 508–521.

Hungi, N., Makuwa, D., Ross, K., Saito, M., Dolata, S., van Cappelle, F., Vellien, J. (2011).

SACMEQ III project results: Levels and trends in school resources among SACMEQ school systems SACMEQ Working Document. Paris: SACMEQ.

Igboegwu, E., & Okonkwo, I. (2012). Influence of gender and location of school on students' achievement in chemistry. *Journal of Research in Education,* 1(1), 1-14.

Isaacs, T., Zara, C, Herbert, G., Coombs, S. & Smith, C. (2013). Key Concepts in Educational Assessment, SAGE Publications Ltd.

Ito, H., Hamada, M., Murayama, Y., Takayanagi, N., Nomura, K., Myogan, M., & Tsujii, M. (2017). Effects of Class Size on Academic Achievement and Emotional and

Behavioral Problems: *The Japanese Journal of Educational Psychology*, *65*(4), 451–465. https://doi.org/10.5926/jjep.65.451

Izmirli, I. (2011). Pedagogy on the Ethno-Mathematics-Epistemology Nexus: A Manifesto. *Journal of Humanistic Mathematics*, 1(2): 27-50.

Jack, R., Bridget, A., & Caine, R. (2017). *Young Lives School Survey, 2016–17: Evidence from Ethiopia*. Retrieved from https://www.younglives-ethiopia.org/node/806

Jones, A., & Buntting, C. (2013). International, national, and classroom assessment: Potent factors in shaping what counts in school science. (Eds.), In D. Corrigan, R. Gunstone, & A. Jones, Valuing assessment in science education: Pedagogy, curriculum, policy (pp.33-53). Netherlands.

Joshi, R. & Verspoo, A. (2013). Secondary education in Ethiopia: Supporting Growth and Transformation, the World Bank 1818 H Street NW, Washington DC.

Kaahwa, J. (2012). The experience of Ugandan females in Mathematics. *Science Journal of Psychology*, 2012, Article ID sjpsych-103, 12 pages. https://www.sjpub.org/sjpsych/sjpsych-103.pdf.

Kassahun, T., & Kedir, B. (2006). Girls' Performance in Mathematics in Upper Primary Schools of Addis Ababa. *Indian Journal of Gender Studies*, *13*(3), 401–424. https://doi.org/10.1177/097152150601300304

Kaur, B., & Wong, K. Y. (2011). *Assessment in the Mathematics Classroom: Yearbook 2011, Association of Mathematics Educators*. World Scientific.

Kemboi, N., Githua, B. & Changeiywo, J. (2014). The Relationship between Students'

Attitudes towards mathematics and their Mathematics Self-Concept and Achievement in the Military Science Programme of Egerton University, *Journal of Education and Practice*, 5(39), 56-63, Kenya.

Kiamanesh, A. R. (2005). The Role of Students' Characteristics and Family Background in Iranian Students' Mathematics Achievement. *PROSPECTS*, *35*(2), 161–174. https://doi.org/10.1007/s11125-005-1819-1

Kissau, S. (2006). Gender Differences in Second Language Motivation: An Investigation of Micro-and Macro-Level Influences. *Canadian Journal of Applied Linguistics*, 9, 73-96.

Kiwanuka, H., Damme, J., Noortgate,V., Anumendem, D. & Namusisi, S. (2015). Factors Affecting Mathematics achievement of first-year secondary school students in Central Uganda, *South African Journal of Education*, Volume 35, Number 3, https://www.ajol.info/index.php/saje/article/view/121853

Klemencic, E. (2010). The impact of international achievement studies on national education policy-making: The case of Slovenia. How many watches do we need? (eds.), In: T. Wiseman, The Impact of International Achievement Studies on National Education Policy-making, 239–268.

Kline, R. (2016). *Principles and Practice of Structural Equation Modelling* (4th ed.). The Guilford Press, http://gen.lib.rus.ec/search.php?req=Principles+and+Practice+of+Structural+Equation+Modelling&lg_topic=libgen&open=0&view=simple&res=25&phrase=1&column=def.

Klinger, D., McDivitt, P., Howard, B., Munoz, M., Rogers, W. & Wylie, E. (2015). The classroom assessment standards for PreK-12 teachers. Kindle Direct Press.

Kocakaya, S., & Kocakaya, F. (2014). A Structural Equation Modelling on Factors of How Experienced Teachers Affect the Students' Science and Mathematics Achievements. *Education Research International*, 1–8.

Kolen, M. & Brennan, R. (2014). *Test equating, scaling, and linking: Methods and practices* (Third Edition). New York: Springer.

Kpolovie, P., Joe, A, & Okoto, T. (2014). Academic Achievement Prediction: Role of Interest in Learning and Attitude towards School. *International Journal of Humanities Social Sciences and Education*, *I*(II), 28.

Kovaleva, G. (2011). Use and Impact of TIMSS and PIRLS data in the Russian Federation, IEA Newsletter, Sept 2011.

Kovalyova, G. (2000). Russia. In: Robitaille, D.F., Beaton, A.E., and Plomp, T. (Eds.), The Impact of TIMSS on the Teaching & Learning of Mathematics & Science (pp.120–124). Vancouver, Canada: Pacific Educational Press. Bingley: Emerald.

Krippendorff, K. (1980). Content analysis: An introduction to its methodology. Sage Publications; Beverly Hills, CA.

Kulpoo, D. and Coustere, p. (1999). Developing National Capacities for Assessment and Monitoring through Effective Partnerships. In Partnership for Capacity Building and Quality Improvements in Education: Papers from ADEA 1997 Biennial Meeting. Dakar, Senegal. Paris.

Kupari, P., & Nissinen, K. (2013). *Background factors behind mathematics achievement in Finnish education context: Explanatory models based on TIMSS 1999 and TIMSS 2011 data*.

Kutsyuruba, B., Klinger, D., & Hussain, A. (2015). Relationships among school climate, school safety, and student achievement and well-being: A review of the literature. *Review of Education*, *3*, 103–135. https://doi.org/10.1002/rev3.3043.

Kyei, K.& Nemaorani T.(2014). Establishing factors that affect performance of grade ten students in high school: A case study of Vhembe district in South Africa. *Journal of Emerging Trends in Educational Research and Policy Studies*, 5(7):83-87.

Lavy, V. (2015). Do Differences in Schools' Instruction Time Explain International Achievement Gaps? Evidence from Developed and Developing Countries. *The Economic Journal*, *125*(588), F397–F424. https://doi.org/10.1111/ecoj.12233.

Lawrence, B. (2013). What is the purpose of classroom assessments? Retrieved from https://classroom.synonym.com/purpose-classroom-assessments-7651703.html

Leavy, P. (2017). *Research design: Quantitative, qualitative, mixed methods, arts-based and community-based participatory research approaches*. New York ; London: Guildord Press.

Lee, V., & Linda Zuze, T. (2011). School Resources and Academic Performance in Sub-Saharan Africa. *Comparative Education Review*, *55*, 369–397.Rertieved from https://www.journals.uchicago.edu/doi/full/10.1086/660157

Lee, V., Linda Zuze, T., & N. Ross, K. (2005). School effectiveness in 14 sub-Saharan African countries: Links with 6 th Graders' reading achievement. *Studies in Educational Evaluation*, *31*, 207–246. https://doi.org/10.1016/j.stueduc.2005.05.011

Leste, A. (2005). Streaming in Seychelles: From SACMEQ research to policy reform. Paper

Presented at SACMEQ Research Conference.

Lietz, P. & Tobin, M. (2016). The impact of large-scale assessments in education on education policy: evidence from around the world, *Journal of Research Papers in Education*, 31:5, 499-501, DOI: 10.1080/02671522.2016.1225918.

Lin, M., Bumgarner, E., & Chatterji, M. (2014). Understanding validity issues in international large-scale assessments. *Quality Assurance in Education*, *22*(1), 31–41. https://doi.org/10.1108/QAE-12-2013-0050

Lindahl, M. (2005). Home versus School Learning: A New Approach to Estimating the Effect of Class Size on Achievement. *Scandinavian Journal of Economics*, *107*(2), 375–394. https://doi.org/10.1111/j.1467-9442.2005.00413

Lockheed, M. (2010). The Craft of Education Assessment: Does Participating in International and Regional Assessments Build Assessment Capacity in Developing Countries. Amsterdam: International Association for the Evaluation of Educational Achievement.

Lockheed, M. (2011, December). Assessing the assessors: Who benefits from international large-scale assessments in developing countries. Paper presented at the SUNY Albany Conference "The Role of Benchmarking in Global Education: Who Succeeds in PISA and Why?" Albany, New York.

Lockheed, M. (2013), "Causes and Consequences of International Assessments in Developing Countries" in H.-D. Meyer & A. Benavot (eds.), PISA, Power and Policy: The Emergence of Global Educational Governance, pp. 163-185, Symposium Books, Oxford.

Lockheed, M. (2015). Why Do Countries Participate in International Large-Scale

Assessments? The Case of PISA. Policy Research Working Paper, Education Global Practice Group.

Lockheed, M. (2016). Measures that Matter: Learning Outcome Targets for Sustainable Development Goal 4 - An examination of national, regional and international learning assessments, Background paper prepared for the 2016 Global Education Monitoring Report Education for people and planet: Creating sustainable futures for all.

Lockheed, M. & Wagemaker, H. (2013). International Large-Scale Assessments: Thermometers, Whips or useful policy tools? J*ournal of Research in Comparative and International Education,* Volume 8, Number 3, PP. 297-306.

Loveless, T., (2007). *Lessons learned: what international assessments tell us about math achievement*. Brookings Institution Press.

Lu, R., & Guo, H. (2018). A Simulation Study to Compare Non-equivalent Groups with Anchor Test Equating and Pseudo-Equivalent Group Linking. *ETS Research Report Series*, *2018*(1), 1–16. https://doi.org/10.1002/ets2.12196

Luse, A., Mennecke, B., & Townsend, A. (2012). Selecting a research topic: A framework for Doctoral students. *International Journal of Doctoral Studies*, 7, 143-152.

Lynn, R., Cheng, H., & Meisenberg, G. (2019). Cognitive Abilities in Young Lives: An Overview of Results from Ethiopia, India, Peru and Vietnam. *MANKIND QUARTERLY*, 17.

Margot, F. (2013) Achievement gaps in Australia: What NAPLAN reveals about education inequality in Australia. *Journal of Race, Ethnicity and Education*, 16:1, 80-102, DOI: 10.1080/13613324.2011.645570.

List, M., Robitzsch, A., Lüdtke, O., Köller, O., & Nagy, G. (2017). Performance decline in low-stakes educational assessments: Different mixture modeling approaches. *Large-Scale Assessments in Education*, *5*(1), 15. https://doi.org/10.1186/s40536-017-0049-3.

Marzano, R. J. (2007). *The Art and Science of Teaching: A Comprehensive Framework for Effective Instruction*. ASCD.

Matang, R.2002. The Role of Ethno-Mathematics in Mathematics Education in Papau New Guinea: Implications for Mathematics Curriculum. *Journal of Educational Studies*, 24(1):27-37.

Mbugua, Z., Kibet, K., Muthaa, G., & Nkonke, G. (2012). *Factors Contributing To Students' Poor Performance in Mathematics at Kenya Certificate of Secondary Education in Kenya: A Case of Baringo County, Kenya*.

McDonnell, L. (1995). Opportunity to learn as a research concept and a policy instrument. *Educational Evaluation and Policy Analysis, 17*(3), 305–322. https://doi.org/10.2307/1164509

Medrich, E., & Griffith, J. (1992). International Mathematics and Science Assessment: What Have We Learned? *US Department of Education, Office of Educational Research and Improvement*, 145.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), Educational measurement (3rd ed., pp. 13-103). New York: Macmillan.

Metzler, J., & Woessmann, L. (2012). The impact of teacher subject knowledge on student achievement: Evidence from within-teacher within-student variation. *Journal of*

Development Economics, *99*(2), 486–496. https://doi.org/10.1016/j.jdeveco.2012.06.002

Mevavech, Z. (2000). Israel. In: Robitaille, D.F., Beaton, A.E., and Plomp, T. (Eds.), The Impact of TIMSS on the Teaching & Learning of Mathematics & Science (pp. 66–70). Vancouver, Canada: Pacific Educational Press.

Miles, M. & Huberman, A. (1994). Qualitative data analysis: An expanded sourcebook (2nd ed.), Newbury Park, CA: Sage.

MOE (2008). Mathematics Syllabus Grades 5 - 8., Federal Democratic Republic of Ethiopia Ministry of Education, Addis Ababa.

MOE (2009). Curriculum Framework for Ethiopian Education (KG – Grade 12), Federal Democratic Republic of Ethiopia Ministry of Education, Addis Ababa.

MOE (2015). Education Sector Development Programme V (ESDP V), 2015/16 - 2019/20: Programme Action Plan, Addis Ababa, Ethiopia.

MOE (2018). Education Statistics Annual Abstract 2016/17. Federal Democratic Republic of Ethiopia Ministry of Education, Addis Ababa.

Moè, A., Pazzaglia, F., Tressoldi, P., & Toso, C. (2009). Attitude towards school, motivation, emotions, and academic achievement. *Educational Psychology: Cognition and Learning, Individual Differences and Motivation*, 261–276.

Mohammadpour, E. (2012). Factors accounting for Mathematics achievement of Singaporean eighth graders. The Asia-Pacific Education Researcher, 21(3):507-518.

Moreira, D.2007. Filling the gap between global and local Mathematics. CERME, 5: 1587-1596. NCTM. 1989. Curriculum and Evaluation Standards for School Mathematics. Reston, VA: Author

Mugenda, O. & Mugenda, A. (1999). *Research Methods: Quantitative and Qualitative Approaches*. Nairobi: Acts Press.

Mullis, I. & Martin, M. (2013). TIMSS 2015 Assessment Frameworks. International Study Center,Lynch School of Education, Boston College.Mullis, I. & Martin, M. (2007). Lessons learned: What international assessments tell us about math achievement. (ed.), In: T. Lovelace TIMSS in Perspective: Lessons learned from IEA's Four Decades of International Mathematics Assessments, 9–36. Washington, DC: Brooking Institution Press.

Mullis, I., Martin, M., P., Foy, P. & Hooper, M. (2016). TIMSS 2015 International Results in Mathematics. Retrieved October 26, 2018, from https://www.iea.nl/publications/timss-2015-international-results-mathematics.

Mullis, I.., Martin, M., Foy, P., & Arora, A. (2012). *TIMSS 2011 international results in mathematics*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center.

Mullis, I., & Martin, M. (2009). *TIMSS advanced 2011 assessment frameworks*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center.

Murimba, S. (2005). The Impact of the Southern and Eastern Africa Consortium for Monitoring Educational Quality (Sacmeq). *PROSPECTS*, *35*(1), 91–108. https://doi.org/10.1007/s11125-005-6822-z.

Muthén, L. & Muthén, B. (1998-2017). Mplus User's Guide. Eighth Edition. Los Angeles, CA: Muthén & Muthén.

Nakajima, H. (2010). Polities of externalization in reflexive time: In Takayama, K., Reinventing Japanese education reform discourses through "Finnish PISA Success", *Comparative Education Review* Vol. 54, No. 1.

NCTM (2000). Principles and standards for school Mathematics. Reston, VA.

Ndlovu, M., & Mji, A. (2012). Alignment between South African Mathematics Assessments

standards and the TIMSS assessment frameworks. *Pythagoras*, *33*, 1–9. https://doi.org/10.4102/pythagoras.v33i3.182

NAE (2008). Ethiopian Second National Learning Assessment of eight Students: National Assessment for Examinations, Addis Ababa, Ethiopia.

NEAEA (2012). Ethiopian Second National Learning Assessment of Grades 4 and 8 Students: NEAEA, Addis Ababa, Ethiopia.

NEAEA (2016). Ethiopian Second National Learning Assessment of Grades 4 and 8 Students: NEAEA, Addis Ababa, Ethiopia.

Newton, P. (2007). Clarifying the purposes of educational assessment. Assessment in Education: Principles, Policy, and Practice, 14(2), 149–170.

Nelson, R. & Dawson, P. (2014). A contribution to the history of assessment: How a Conversation Simulator Redeems Socratic Method: *Journal of Assessment and Evaluation in Higher Education,* Vol. 39, 195-204. Taylor and Francis.

Noveanu, G. & Noveanu, D. (2000). Romania. (Eds.), In Robitaille, D.F., Beaton, A.E., and Plomp, T. The Impact of TIMSS on the Teaching & Learning of Mathematics & Science (pp. 117–119). Vancouver, Canada: Pacific Educational Press.

Nyabuto, B. (2007. December 1). Grade Repetition in Kenyan Primary Schools: Issues of Learning Disabilities. Retrieved July 25, 2019, from Adlibris website: https://www.adlibris.com/se/bok/grade-repetition-in-kenyan-primary-schools-issues-of-learning-disabilities-9783825808778

Obeng-Denteh, W., Yeboah, A., Sam, C., & Monkah, E. (2011). *The impact of student and teacher absenteeism on student performance at the junior high school: the case of the Kumasi-metro school district*. *4*(1), 7–17.

Ochwo, P. (2013). Pupil, teacher, and school factors that influence student achievement on the primary leaving examination in Uganda: Measure development and multilevel modeling. Unpublished doctoral dissertation. Ohio, USA.

Ogunsola, O., Osuolale, K., & Ojo, A. (2014). Parental and Related Factors Affecting Students Academic Achievement in Oyo State Nigeria. *International Journal of Social, Education, Economics and Management Engineering*, *8*(9). Retrieved from https://www.academia.edu/36065610/Parental_and_Related_Factors_Affecting_Students_Academic_Achievement_in_Oyo_State_Nigeria.

Ojimba, D. (2013). *Home Background and Senior Secondary Students' Achievement in Mathematics in Rivers State, Nigeria*. https://doi.org/10.5901/jesr.2013.v3n2p283

Oliveira, F., De Menezes, T., Irffi, G., & Oliveira, G. (2018). Bullying effect on student's performance. *ECONOMIA*, *19*(1), 57–73. https://doi.org/10.1016/j.econ.2017.10.001.

PASEC (2015). PASEC 2014 Education System Performance in Francophone Sub-Saharan Africa: Competencies and Learning Factors in Primary Education, CONFEMEN, Dakar (Senegal)

Palomba, C. & Banta, T. (1999). Assessment Essentials: Planning, Implementing, and Improving Assessment in Higher Education. Jossey-Bass, Inc., San Francisco.

Parcel, T., Dufur, M., & Zito, R. (2010). Capital at home and at school: A Review and

Synthesis. *Journal of Marriage and Family, 72*(4), 828–846. https://doi.org/10.1111/j.1741-3737.2010.00733.x

Patrick, H., Anderman, L., & Ryan, A. (2002). Social motivation and the classroom social environment. In *Goals, Goal Structures, and Patterns of Adaptive Learning* (pp. 85–108).

Pelgrum, H., Blahova, V., Dukynaite, R. (2011). IEA experiences from Latvia, Lithuania, the Slovak Republic, and Slovenia. (eds.), In C. Papanastasiou, T. Plomp, and E. Papanstasiou, IEA 1958–2008:50 Years of Experiences and Memories, 447–468. Nicosia: Cultural Center of the Kykkos Monastery.

PISA (2003). OECD PISA. Retrieved on 29 June 2011 fromhttp://www.pisa.oecd.org/.

Poppe, R., Frölich, M., & Haile, G. (2019). School Meals and Educational Outcomes in Rural Ethiopia. *The Journal of Development Studies*, *55*(8), 1741–1756. https://doi.org/10.1080/00220388.2017.1311405.

Postlethwaite, T & Kellaghan, T. (2008) National assessments of educational achievement. Education Policy Series 9. Paris: UNESCO, International Institute for Educational Planning (IIEP) and Brussels: International Academy of Education (IAE).

Rajoo, M. (2013). Students' perceptions of Mathematics classroom environment and Mathematics achievement: A study in Spitang, Sabah, Malaysia. Paper presented at the International Conference on Social Science Research, Penang, Malaysia.

Rahman, S., Mazlan, M., Kummin, S., Yasin, R., & Meerah, T. (2010). Examining the role of language on students' achievement: A study on the use of second language as a medium of instruction in teaching science subject in Malaysia. Procedia - Social and Behavioral Sciences, 9,1261–1265. https://doi.org/10.1016/j.sbspro.2010.12.317

Ravela, P., Arregui, P., Valverde, G., Wolfe, R., Ferrer, G. Rizo, F., Aylwin, M. & Wolff, L. (2009). "The Educational Assessments that Latin America Needs." Washington, DC: PREAL.

Ritzen, J. (2013). International Large-Scale Assessments as Change Agents. In M. von Davies, E. Gonzalez, I. Kirsch, & K. Yamamoto (Eds.), *The Role of International Large-Scale Assessments: Perspectives from Technology, Economy, and Educational Research* (pp. 13–24). https://doi.org/10.1007/978-94-007-4629-9_2

Robitaille, D. (2000). The Impact of TIMSS on the Teaching and Learning of Mathematics and Science. Vancouver: Pacific Educational Press.

Rosa, M., & Orey, D. C. (2010). Culturally relevant pedagogy: an thno-Mathematical approach. Horizontes, 28(1): 19-31.

Ross, K. (2007). Sample design for educational survey research, UNESCO International Institute for Educational Planning, Module 3.

Rumberger, R., & Palardy, G. (2004). Multilevel Models for School Effectiveness Research. In D. Kaplan, *The SAGE Handbook of Quantitative Methodology for the Social Sciences* (pp. 236–259). https://doi.org/10.4135/9781412986311.n13.

Rutkowski, D., & Prusinski, E. (2011). *The Limits and Possibilities of International Large-Scale Assessments*. *9*(2), 4.

Rutkowski, L., Davier, M., & Rutkowski, D. (2014). *Handbook of International Large-Scale Assessment: Background, Technical Issues, and Methods of Data Analysis* (1st ed.). https://doi.org/10.1201/b16061.

Salvia, J., & Ysseldyke, J. E. (1995). Assessment is special and remedial education. Boston, MA: Houghton Mifflin.

Säljö, R., & Radišić, J. (2018). Promoting teaching, learning, and informed decision-making

through the lenses of international large-scale assessment: Looking beyond numbers. *European Journal of Psychology of Education - EJPE (Springer Science & Business Media B.V.)*, *33*(3), 419–426. https://doi.org/10.1007/s10212-01-0388-7.

Sansivieri, V., & Wiberg, M. (2017, June 7). *IRT Observed-Score Equating with the Non-equivalent Groups with Covariates Design*. 275–285. DOI: 10.1007/978-3-319-56294-0_25

Saritas, T. & Akdemir, O. (2009). Identifying Factors Affecting the Mathematics Achievement of Students for Better Instructional Design, Turkey.

Schmidt, W. & Cogan, L. (1996). In: M.O. Martin & D.L. Kelly. TIMSS Technical Report Volume I: Design and development. Boston: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.

Schmidt, W., Raizen, S., Britton, E., Bianchi, L., & Wolfe, R. (1997). Many Visions, Many Aims Volume 2: A Cross-National Investigation of Curricular Intentions in School Science. Dordrecht: Springer Netherlands: Dordrecht.

Shepherd, E. & Godwin, J. (2004). Assessments through the Learning Process. Question Mark White Paper, Question Mark Corporation.

Shiel, G., & Cartwright, F. (2015). *Analyzing data from a national assessment of educational achievement*. Washington, DC: World Bank Group.

Silva, B., Bosancianu, C., & Littvay, L. (2019). *Multilevel Structural Equation Modelling*. Sage.

Singh, P. & Ellerton, N. (2013). International Collaborative Studies in Mathematics Education. In M. Clements, A. Bishop, C. Keitel, J. Kilpatrick & F. Leung (eds.), Third International Handbook of Mathematics Education (pp.827- 860). DOI-10.1007/978-1-4614-4684-2.

Slate, J. & Jones C. (2005). Effects of school size: A review of the literature with recommendations. Essays in Education, 13 (1):1-24. http://edresearch.yolasite.com/resources/slate.pdf.

Slevitch, L. (2011). Qualitative and quantitative methodologies compared ontological and epistemological perspectives. *Journal of Quality Assurance in Hospitality & Tourism*, 12(1), pp.73-81.

Stewart, E. (2008). School Structural Characteristics, Student Effort, Peer Associations, and Parental InvolvementThe Influence of School and Individual-Level Factors on Academic Achievement. *Education and Urban Society - EDUC URBAN SOC*, *40*, 179–204. https://doi.org/10.1177/0013124507304167.

Strakova, J., Paleckova, J., &Tomasek, V. (2000). Czech Republic. In Robitaille, D.F. Beaton, A.E., and Plomp, T. (Eds.), The Impact of TIMSS on the Teaching & Learning of Mathematics & Science (pp. 41–44). Vancouver, Canada: Pacific Educational Press.

Strietholt, R., & Scherer, R. (2018). The Contribution of International Large-Scale Assessments to Educational Research: Combining Individual and Institutional Data

Sources. *Scandinavian Journal of Educational Research*, *62*(3), 368–385. https://doi.org/10.1080/00313831.2016.1258729.

Suurtamm, C., Thompson, D., Kim, R., Moreno, L., Sayac, N., Schukajlow, S., Silver, E., Ufer, S. & Vos, P. (2016)' Assessment in Mathematics Education: Large-Scale Assessment and Classroom Assessment, ICME-13 Topical Surveys, http://www.springer.com/series/14352.

Tassew Woldehanna & Aregawi Gebremedhin (2016). *Learning Outcomes of Children Aged 12 in Ethiopia: A Comparison of Two Cohorts*. https://www.younglives.org.uk/sites/www.younglives.org.uk/files/YL-WP152-Learning%20outcomes%20in%20Ethiopia.pdf

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53–55. DOI: 10.5116/ijme.4dfb.8dfd

Taylor, P., & Medina, M. (2013). *Educational research paradigms: From positivism to Multi-paradigmatic*. https://doi.org/10.13140/2.1.3542.0805.

Taylor, A. (2000). Canada. In: Robitaille, D.F., Beaton, A.E., and Plomp, T. (Eds.), The Impact of TIMSS on the Teaching & Learning of Mathematics & Science (pp. 50–55). Vancouver, Canada: Pacific Educational Press.

Tesfaledet, M., Beyene, Z., & Tsegaye, T. (2019). Multilevel analysis of high school student's achievement in physical science in Kafta, Sheka, and Bench Maji zones, Southwest Ethiopia. *Asian Journal of Computer Science and Information Technology*, *10*(1), 9151–91560.

Teshome, Nekatibeb. (2005). The impact of learning with the mother tongue on academic achievement: A case study of Grade 8 students in Ethiopia. *7th International*

Language and Development Conference, 11. Retrieved from https://www.teachingenglish.org.uk/sites/teacheng/files/Language%20%26%20Development%20Africa%20and%20Beyond.pdf.

Torney-Purta, J., & Amadeo, J. (2013). International Large-Scale Assessments: Challenges in Reporting and Potentials for Secondary Analysis. *Research in Comparative and International Education*, *8*(3), 248–258. https://doi.org/10.2304/rcie.2013.8.3.248

Travers, K. & Westbury, I. (1989). The IEA study of Mathematics I: International analysis of Mathematics curricula. Oxford; Pergamon Press.

Tyumeneva, Y. (2013). Disseminating and using student assessment information in Russia. SABER student assessment working paper; no.10; SABER student assessment working paper; no. 10. Washington, DC.

UNESCO (2000). Education for All: Status and Trends 2000, Assessing Learning Achievement. A Report Published for the International Consultative Forum on Education For All: France.

UNESCO (2000). The Dakar Framework for Action-Education for All: Meeting Our Collective Commitments. Paris

UNESCO (2000a) Status and Trends. Assessing Learning Achievement. UNESCO, Paris

UNESCO (2003a). Education in a multilingual world. UNESCO Education Position Paper. Paris: UNESCO. Retrieved September 15, 2020, from: http://unesdoc.unesco.org/images/0012/001297/129728e.pdf.

UNESCO (2012). Primary School Curricula on Reading and Mathematics in Developing Countries.Montreal, Canada.

UNESCO (2018). *The Impact of Large-Scale Learning Assessments*. Retrieved from http://uis.unesco.org/sites/default/files/documents/impact-large-scale-assessments-2018-en.pdf, UNESCO Institution for Statistics.

UNESCO (2018). *Large scale learning assessment regional workshop*. Retrieved October 20, 2020, from http://www.education2030-africa.org/index.php/en/lien-talent-blog-en/876-large-scale-assessment-workshop.

Ushie, M., Emeka, J., Ononga, G. & Owolabi, E. (2012). Influence of Family Structure on Students' Academic Performance in Agege Local Government Area, Lagos State, Nigeria. *European Journal of Educational Studies*, 4(2), 177 – 187.

Van der Werf, C. (2014). The Effects of Bullying on Academic Achievement. *Desarrollo y Sociedad*, (74), 275–308. https://doi.org/10.13043/DYS.74.6.

Veresova, M., & Mala, D. (2016, November 22). *Attitude toward School and Learning and Academic Achievement of Adolescents*. 870–876. DOI:10.15405/epsbs.2016.11.90

Wagemaker, H. (2014). International Large-Scale Assessments: From Research to Policy. In L. Rutkowski, M. Davier & D. Rutkowski (eds.), Handbook of International Large-Scale Assessment: Background, Technical Issues, and Methods of Data Analysis (pp. 11-36). CRC Press, USA.

Wagemaker, H. (2011). IEA: International studies, impact, and transition. In: IEA 1958–2008:50 Years of Experiences and Memories, eds. C. Papanastasiou, T. Plomp and E. Papanstasiou, 253–273. Nicosia: Cultural Center of the Kykkos Monastery.

Wagemaker, H. (2010). IEA: Globalization and assessment. In: International Encyclopedia of Education, ed. E. B. Penelope Peterson, Vol. 4, 663–668. Oxford: Elsevier.

Walvoord, B. (2004). Assessment Clear and simple: A practical guide for institutions, departments, and general education. San Francisco: Jossey-Bass.

Wheaton, B., Muthen, B., Alwin, D & Summers, G. (1977), "Assessing Reliability and Stability in Panel Models," *Sociological Methodology*, 8 (1), 84-136.

Whitley, B. (2002). Principals of Research and Behavioral Science, Boston, McGraw-Hill

Willms, J. D., & Somer, M.-A. (2001). Family, Classroom, and School Effects on Children's Educational Outcomes in Latin America. *School Effectiveness and School Improvement*, *12*(4), 409–445. https://doi.org/10.1076/sesi.12.4.409.3445

Wößmann, L. (2003). Schooling Resources, Educational Institutions, and Student Performance: The International Evidence. *Oxford Bulletin of Economics and Statistics*, *65*(2), 117–170. https://doi.org/10.1111/1468-0084.00045.

Woessmann, L., & Fuchs, T. (2004). *What Accounts for International Differences in Student Performance? A Re-Examination Using PISA Data* (SSRN Scholarly Paper No. ID 572802). Retrieved from Social Science Research Network website: https://papers.ssrn.com/abstract=572802.

Woodhouse, D. (1973). Modern Mathematics in East Africa: A Survey, 1961-1971. *Australian Mathematics Teacher*. Retrieved from https://eric.ed.gov/?id=EJ106457

Yayan, B., & Berberoglu, G. (2004). A re-analysis of the TIMSS 1999 mathematics assessment data of the Turkish students. *Studies in Educational Evaluation*, *30*(1), 87–104.

Yohannes, Bisa, Bereket Merkine, & Hailu Chare. (2017). The relationship between parental involvement and children’s academic achievement in alternative basic education centers and regular primary first cycle school of Hamer Woreda, South Omo zone, Ethiopia. *International Journal of Current Research*, *9*(7), 9.

Yusuf, M. & Adigun, J. (2010). The influence of school sex, location, and type on students' Academic performance. *International Journal of Educational Sciences*, 2(2):81-85. Education: Principles, Policy & Practice, 17, 345–356. https://doi.org/10.1080/09751122.2010.11889992

Zakharov, A., Carnoy, M., & Loyalka, P. (214). Which teaching practices improve student performance on high-stakes exams? Evidence from Russia. *International Journal of Educational Development*, *36*, 13–21. https://doi.org/10.1016/j.ijedudev.2014.01.003.

Zaslavsky, G. (1993). Multicultural Mathematics: One road to the goal of Mathematics for all. In: G.J. Cuevas, M. J. Driscoll (eds.) Reaching All Students with Mathematics (pp.45-55). Reston, VA: The NCTM, Inc.

Appendices

Appendix I: TIMSS Grade eight Participating Countries/Systems since 1995-2015

Countries/Education system	1995	1999	2003	2007	2011	2015
Algeria				●		
Argentina	●		●			
Armenia			●	●	●	●
Australia	●	●	●	●	●	●
Austria	●					
Bahrain			●	●	●	●
Belgium	●	●	●			
Bosnia and Herzegovina				●		
Botswana			●	●		
Bulgaria	●	●	●	●		
Canada	●	●				●
Chile		●	●		●	●
Chinese Taipei		●	●	●	●	●
Colombia	●			●		
Cyprus	●	●	●	●		
Czech Republic	●	●		●		
Denmark	●					
Egypt			●	●		●
El Salvador				●		
England-GBR	●	●	●	●	●	●
Estonia			●			
Finland		●			●	
France	●					
Georgia				●	●	●
Germany	●					
Ghana			●	●	●	
Greece	●					
Hong Kong-CHN	●	●	●	●	●	●
Hungary	●	●	●	●	●	●
Iceland	●					
Indonesia	●	●	●	●	●	
Iran, Islamic Republic of	●	●	●	●	●	●
Ireland	●					●
Israel	●	●	●	●	●	●
Italy	●	●	●	●	●	●
Japan	●	●	●	●	●	●

Countries/Education system	1995	1999	2003	2007	2011	2015
Jordan		●	●	●	●	●
Kazakhstan					●	●
Korea, Republic of	●	●	●	●	●	●
Kuwait	●			●		●
Latvia	●	●	●			
Lebanon			●	●	●	●
Lithuania	●	●	●	●	●	●
Macedonia, Republic of		●	●		●	
Malaysia		●	●	●	●	●
Malta				●		●
Mexico	●					
Moldova, Republic of		●	●			
Mongolia				●		
Morocco		●	●	●	●	●
Netherlands	●	●	●			
New Zealand	●	●	●		●	●
Northern Ireland-GBR						
Norway1	●		●	●	●	●
Oman				●	●	●
Palestinian Nat'l Auth.			●	●	●	
Philippines	●	●	●			
Portugal	●					
Qatar				●	●	●
Romania	●	●	●	●	●	
Russian Federation	●	●	●	●	●	●
Saudi Arabia			●	●	●	●
Scotland-GBR	●		●	●		
Serbia			●	●		
Singapore	●	●	●	●	●	●
Slovak Republic	●	●	●			
Slovenia	●	●	●	●	●	●
South Africa	●	●	●			
Spain	●					
Sweden	●		●	●	●	●
Switzerland	●					
Syrian Arab Republic				●	●	
Thailand	●	●		●	●	●
Tunisia		●	●	●	●	
Turkey		●		●	●	●
Ukraine				●	●	
United Arab Emirates					●	●
United States	●	●	●	●	●	●
Total	45	38	47	50	42	38
US benchmarking states						

Countries/Education system	1995	1999	2003	2007	2011	2015
Alabama-USA					●	
California-USA					●	
Colorado-USA					●	
Connecticut-USA		●			●	
Florida-USA					●	●
Idaho-USA		●				
Illinois-USA	●	●				
Indiana-USA		●	●		●	
Maryland-USA		●				
Massachusetts-USA		●		●	●	
Michigan-USA		●				
Minnesota-USA	●			●	●	
Missouri-USA2	●	●				
North Carolina-USA		●			●	
Oregon-USA2	●	●				
Pennsylvania-USA		●				
South Carolina-USA		●				
Texas-USA		●				
Total	4	13	1	2	9	1
Other Benchmarking participants						
Abu Dhabi-UAE					●	●
Alberta-CAN	●	●			●	
Basque Country-ESP			●	●		
British Columbia-CAN		●		●		
Buenos Aires-ARG						●
Dubai-UAE				●	●	●
Ontario-CAN	●	●	●	●	●	●
Quebec-CAN	●	●	●	●	●	●
Total	3	4	3	5	5	5
Off-grade participants						
Botswana					●	●
Honduras					●	
Norway1						●
South Africa					●	●
Total	0	0	0	0	3	3
Grand total	52	55	51	57	59	47

Note that ● indicates countries participation in assessment cycles

Appendix II. List of 10 High and Low Performer Countries in Grade 8 Mathematics TIMSS

Table A. Top 10 high-performed countries

Rank	TIMSS（1995）		TIMSS-R（1999）		TIMSS (2003)		TIMSS (2007)		TIMSS (2011)		TIMSS (2015)	
	Country	Score	Country	Score	Country	Score	Country	Score	Country	Score	Country	Score
1	Singapore	643	Singapore	604	Singapore	605	Taiwan	598	South Korea	613	Singapore	621
2	South Korea	607	South Korea	587	South Korea	589	South Korea	597	Singapore	611	South Korea	606
3	Japan	605	Taiwan	585	Hong Kong	586	Singapore	593	Taiwan	609	Taiwan	599
4	Hong Kong	588	Hong Kong	582	Taiwan	585	Hong Kong	572	Hong Kong	586	Hong Kong	594
5	Belgium	565	Japan	579	Japan	570	Japan	570	Japan	570	Japan	586
6	Czech Republic	564	Belgium	558	Belgium	537	Hungary	517	Russia	539	Russia	538
7	Slovakia	547	Netherlands	540	Netherlands	536	England	513	Israel	516	Kazakhstan	528
8	Switzerland	545	Slovakia	534	Estonia	531	Russia	512	Finland	514	Canada	527
9	Netherlands	541	Hungary	532	Hungary	529	US	508	US	509	Ireland	523
10	Slovenia	541	Canada	531	Malaysia	508	Lithuania	506	England	507	US & England	518

Table B. 10 Least perfumed countries

Rank	TIMSS（1995）		TIMSS-R（1999）		TIMSS (2003)		TIMSS (2007)		TIMSS (2011)		TIMSS (2015)	
	Country	Score	Country	Score	Country	Score	Country	Score	Country	Score	Country	Score
1	South Africa	354	South Africa	275	South Africa	264	Qatar	307	Ghana	331	Kuwait	353
2	Colombia	385	Morocco	337	Ghana	276	Ghana	309	Oman	366	South Africa	376
3	Kuwait	392	Philippines	345	Saudi Arabia	332	Saudi Arabia	329	Morocco	371	Morocco	377
4	Iran	428	Chile	392	Botswana	366	El Salvador	340	Syrian	380	Saudi Arabia	383
5	Portugal	454	Indonesia	403	Philippines	378	Kuwait	354	Indonesia	386	Jordan	388
6	Cyprus	474	Iran	422	Morocco	387	Botswana	364	Saudi Arabia	394	Indonesia	397
7	Lithuania	477	Jordan	428	Chile	387	Palestinian	367	Palestinian	404	Abu Dhabi	419
8	Romania	482	Turkey	429	Palestinian	390	Oman	372	Jordan	406	Oman	425
9	Greece	484	Macedonia	447	Bahrain	401	Colombia	380	Bahrain	409	Iran	431
10	Iceland	487	Tunisia	448	Egypt	406	Algeria	387	Qatar	410	Argentina	432

Appendix III. Alignment of Contents and Objectives in Ethiopian Mathematics and TIMSS Assessment Framework

Content Domain	Specific Content	TIMSS Objectives (Grade 7-8)	Ethiopian Objectives (Grade 5-8)
1. Number	1. Whole numbers	1. Demonstrate understanding of whole numbers and operations (e.g., the four arithmetic operations; place value; and the commutative, associative, and distributive properties).	Whole Numbers and Operations on numbers (Grade 5) • Determine the place value of each digit in a given whole number • Add, subtract, multiply and divide whole numbers • Identify the commutative and associative properties of addition • Identify the distributive property of multiplication over addition
		2. Compute with whole numbers in problem situations.	Under Operations on numbers Grade 7 students' textbook
		3. Find and use multiples or factors of numbers, identify prime numbers, and 4. Evaluate powers of numbers and square roots of perfect squares up to 144.	Identify prime and composite whole numbers (grade 6) Grade 8 • Find the squares of rational numbers • Find the square root of perfect squares • Find cubes of rational numbers • Find cube root of perfect cubes
	2. Fractions, decimals, and integers	1. Identify, compare, or order rational numbers (fractions, decimals, and integers) using various models and representations (e.g., number line); and know that there are numbers that are not rational.	Rational Numbers (Grade 5 & 6) • Explain the meaning of proper and improper fractions and mixed numbers • Compare and order simple fractions by using equivalent fractions • Express a percentage as a fraction and convert a percentage to fractions and decimals and vice versa\ • Write a fraction in its lowest terms • Change a fraction to a decimal and percentage • Convert terminating decimals to fractions and decimals • Convert percentages to fractions and decimals • Compare and order fractions
		2. Compute with rational numbers (fractions, decimals, and integers) including those set in	Grade 5 • Add fractions with the same and with different denominator • Add decimals up to three decimal places

Content Domain	Specific Content	TIMSS Objectives (Grade 7-8)	Ethiopian Objectives (Grade 5-8)
		problem situations.	• Subtract fractions with the same and different denominator • Subtract decimals up to three decimal places • Multiply decimals up to three decimal places by one-digit whole number • Multiply decimals up to three decimal places by 10, 100, 1000 • Divide decimals up to three decimal places by one-digit natural number • Divide decimals up to three decimal places by 10, 100, and 1000 (quotient not more than three decimal places) Add, subtract, multiply and divide fractions and decimals (grade 6)
	3. Ratio, proportion, and percent	1. Identify and find equivalent ratios; 2. Model a given situation by using a ratio and 3. Divide a quantity in a given ratio. 4. Convert among percent, proportions, and fractions. 5. Solve problems involving percent or proportions.	Ratio, proportion, and percentage (Grade 6 & 7) • Explain and use a direct proportion to solve problems • Explain and use an inverse proportion to solve determine the factors of proportionality • Draw graphs to illustrate direct and inverse proportion • Solve problems using the ratio • Solve problems on profit and loss • Solve problems on simple interest
2. Algebra	1. Expressions and operations	1. Find the value of an expression given values of the variables.	Available in Grade 7 and 8 under Unit two of the student's text but not as an objective Variables, equations and inequalities (MLC of Grade 5) • Use letters to represent numbers in simple terms and find the value of the term • Express simple statements as Mathematical expressions • Find the values of simple Mathematical expressions by substituting numbers for letters • Identify algebraic expressions based on terms
		2. Simplify algebraic expressions involving sums, products, and powers of expressions; and compare expressions to determine if they are equivalent.	Available in Grade 7 and Grade 8 Unit two of the student's text but not as an objective

Content Domain	Specific Content	TIMSS Objectives (Grade 7-8)	Ethiopian Objectives (Grade 5-8)
		3. Use expressions to represent problem situations.	Available in Grade 7 and Grade 8 Unit two of the student's text but not as an objective
	2. Equations and inequalities	1. Write equations or inequalities to represent situations.	Variables, equations, and inequalities (Grade 5 & 6) • Solve simple linear equations in one unknown • Solve simple linear inequalities in one unknown • Solve one-step linear equations • Solve one-step linear inequalities and show the solution on a number line
		2. Solve linear equations, linear inequalities, and	Variables, equations, and inequalities (Grade 7 & 8) • Solve any linear equation with positive coefficients by using equivalent transformations • Solve real-life problems using linear equations • Solve linear inequalities with positive coefficients by using equivalent transformations • Solve real-life problems using inequalities • Solve linear equations involving brackets • Solve linear equations involving fractions • Solve real-life word problems using linear equations • Recognize and use the symbols ≥ and ≤ • Solve linear inequalities
		Solve simultaneous linear equations in two variables.	**Not available**
	3. Relationships and functions	1. Generalize pattern relationships in a sequence, or between adjacent terms, or between the sequence number of the term and the term, using numbers, words, or algebraic expressions.	**Not available**
		2. Interpret, relate, and generate representations of functions in	Grade 8 *Graphs, relations, functions, and the Cartesian coordinate plane*

Content Domain	Specific Content	TIMSS Objectives (Grade 7-8)	Ethiopian Objectives (Grade 5-8)
		tables, graphs, or words.	• Plot the coordinates from a table of ordered pairs corresponding to a linear equation in two variables and realize that it is always a straight line
		3. Identify functions as linear or non-linear; contrast properties of functions from tables, graphs, or equations; and interpret the meanings of slope and y-intercept in linear functions.	Graphs, relations, functions, and Cartesian coordinate plane (Grade 8) • Draw the four quadrants of the Cartesian plane and mark the origin and the X and Y axes • Plot points on the Cartesian coordinate plane giving their coordinates point on the Cartesian coordinate plane • Read the coordinates of any point on the Cartesian coordinate plane • Draw the lines of Y = 0 and recognize that it is the X axis, if x = 0 and recognize it is the Y axis • Y = mx • Determine simple equations that relate to the coordinates of a given list of ordered pairs • Plot the coordinates from a table of ordered pairs corresponding to a linear equation in two variables and realize that it is always a straight line
3. Geometry	1. Geometric shapes	1. Identify different types of angles and use the relationships between angles on lines and in geometric figures.	Lines, angles, and plane figures (Grade 5 & 6) • Define an angle and degrees and identify elements of angles • Classify angles according to their magnitude • Measure angles using a protractor • Identify vertically opposite, alternate interior, alternate exterior, and corresponding angles • Solve problems related to angles formed by two parallel lines and a transversal
		2. Identify geometric properties of two- and three-dimensional shapes, including line and Rotational symmetry.	Quadrilaterals and polygons (Grade 7) • Describe the properties of the trapezium, parallelogram, rectangle, square, and rhombus • Identify and name polygons with up to 10 sides • Determine the sum of the interior angles of a given polygon **Not available**

Content Domain	Specific Content	TIMSS Objectives (Grade 7-8)	Ethiopian Objectives (Grade 5-8)
		3. Identify congruent triangles and quadrilaterals and their corresponding measures; and identify similar triangles and use their properties.	Triangles (Grade 6 & 8) • Deduce triangle properties from constructing triangles and state them • Identify congruent triangles by using the tests for congruency (SSS, SAS, ASA) • Determine the facts about similar triangles' • Apply the similarity facts to determine the similarity of two triangles • Determine the relationship between the perimeters of similar triangles • Determine the relationship between the areas of similar triangles • Derive Euclid's theorem using similar triangles • Derive Pythagoras Theorem • Apply Euclid's Theorem and its converse to solve problems
		4. Relate three-dimensional shapes with their two-dimensional representations (e.g., nets, two-dimensional views of three-dimensional objects).	Nets of cubes and rectangular prisms (Grade 5 syllabus)
		5. Use geometric properties, including the Pythagorean Theorem, to solve problems.	Triangles Grade 8 • Derive Pythagoras Theorem
	2. Geometric measurement	1. Draw and estimate the size of given angles, line segments, and perimeters; and estimate areas and volumes. 2. Select and use appropriate measurement formulas for perimeters, circumferences, areas, surface areas, and volumes; and find measures of compound areas.	Lines, angles, and plane figures (Grade 5) • Define an angle and degrees and identify elements of angles • Classify angles according to their magnitude • Measure angles using a protractor Perimeter, Area, and Volume (grade 7) • Compute the area of a triangle • Calculate the area and perimeter of a trapezium and a parallelogram • Compute the circumference and area of a circle • Calculate the surface area of prisms and cylinders

Content Domain	Specific Content	TIMSS Objectives (Grade 7-8)	Ethiopian Objectives (Grade 5-8)
			• Calculate the volume of prisms and cylinders Volume (grade 8) • Identify the radius and surface of a sphere
	3. Location and movement	1. Locate points and solve problems involving points in the Cartesian plane.	The Cartesian coordinate plane (Grade 8) • Draw the four quadrants of the Cartesian plane and mark the origin and the X and Y axes • Plot points on the Cartesian coordinate plane giving their coordinates point on the Cartesian coordinate plane • Read the coordinates of any point on the Cartesian coordinate plane
		2. Recognize and use geometric transformations (translation, reflection, and rotation) of two-dimensional shapes.	**Not available**
4. Data and Chance	1. Characteristics of data sets	1. Identify and compare characteristics of data sets including mean, median, mode, range, and shape of distributions (in general terms).	Data handling and probability (Grade 7) • Use simple examples to show when to use the mean, mode or median
		2. Calculate, use, or interpret mean, median, mode, or range to solve problems.	• Calculate the mean, mode, median, and range of given data • Use simple examples to show when to use the mean, mode or median
	2. Data interpretation	1. Read data from a variety of visual data displays.	• Construct and interpret simple line graphs • Construct and interpret pie charts
		2. Use and interpret data sets to solve problems (e.g., make inferences, draw conclusions, and estimate values between and beyond given data points).	• Construct and interpret simple line graphs • Construct and interpret pie charts
		3. Identify and describe approaches to organizing and displaying data that could lead to misinterpretation (e.g.,	• Collect simple data using tally marks • Construct and interpret simple line graphs

Content Domain	Specific Content	TIMSS Objectives (Grade 7-8)	Ethiopian Objectives (Grade 5-8)
		inappropriate grouping, and misleading or distorted scales).	• Construct and interpret pie charts • Show understanding of the terms mean, mode, media
	3. Chance	1. Judge chances of outcomes as certain, more likely, equally likely, less likely, or impossible in general terms.	Data handling and probability (Grade 8) • Identify events from every day life that demonstrate outcomes that are certain, uncertain, and impossible • Determine the probability of simple events e.g tossing a coin • Identify events from every day life that demonstrate outcomes that are certain, uncertain, and impossible • Determine the probability of simple events e.g tossing a coin • Express probability using fractions, decimals or percentages
		2. Use data, including experimental data, to estimate the chances of future outcomes.	
		3. Given a process designed to be random, determine the chances of possible outcomes.	

Appendix IV. Summary description of student-related variables and their reliabilities

Student-related variables	N	Mean	SD	Reliability
Student family SES				
Parental education level	200	3.12	1.18	0.90
Number of books at home	200	1.74	1.37	
Family possessions	200	1,77	1.61	
Parental follow up				
My parents ask me what I am learning in school	199	2.17	1.45	0.86
I talk about my schoolwork with my parents	199	2.53	1.64	
My parents make sure that I set aside time for my homework	199	2.23	1.45	
My parents check if I do my homework	199	2.38	1.71	
Attitude of students towards the schools				
I like being in school	200	1.64	1.51	0.91
I feel safe when I am at school	200	1.96	1.86	
I like to see my classmates at school	200	1.52	1.39	
Teachers at my school are fair to me	200	1.69	1.51	
I am proud to go to this school	200	1.62	1.32	
I learn a lot in school	200	1.54	1.38	
My teachers understand student's problem	200	1.9	1.39	
I am happy with the ethics of my teachers	200	1.68	1.24	
Attitude of students towards learning Mathematics				
I enjoy learning mathematics	200	1.66	1.35	0.74
I wish I did not have to study mathematics	200	3.04	1.66	
I learn many interesting things in mathematics	200	1.70	1.43	
It is important to do well in mathematics	200	1.43	1.24	
Confidence of students in learning Mathematics				
I usually do well in mathematics	200	1.8	1.02	0.73
Mathematics is more difficult for me than for many of my classmate	200	2.61	1.42	
I learn things quickly in mathematics	200	2.37	1.7	
Mathematics makes me nervous	200	2.83	1.67	
My teacher tells me I am good at mathematics	200	2.13	1.47	
Mathematics is harder for me than any other subject	200	2.42	1.52	
I would like a job that involves using mathematics	200	1.73	1.21	

Appendix V. The values of R- squared for student and school-related variables that explained Achievement in Mathematics

Table A. The values of R^2 for the variables related students' home background

Observed Variables	Estimate	S.E.	z	*p*
Student-Level (Within)				
Parent highest education level	0.302	0.105	2.871	0.004
Number of books at home	0.133	0.037	3.603	0.000
Family possessions	0.117	0.036	3.288	0.001
School-Level (Between)				
Parent highest education level	0.602	0.132	4.556	0.000
Number of books at home	0.500	0.157	3.194	0.001
Family possessions	0.801	0.168	4.782	0.000

Table B. The values of R^2 for the variables related to parental follow up of students learning

Observed Variables	Estimate	S.E.	z	*p*
Student-Level (Within)				
My parents ask me what I am learning in school	0.496	0.029	17.094	0.000
I talk about my schoolwork with my parents	0.491	0.027	17.911	0.000
My parents make sure that I set aside time for my homework	0.500	0.031	15.980	0.000
My parents check if I do my homework	0.537	0.034	15.799	0.000
School-Level (Between)				
My parents ask me what I am learning in school	0.714	0.131	5.461	0.000
I talk about my schoolwork with my parents	0.640	0.136	4.715	0.000
My parents make sure that I set aside time for my homework	0.986	0.047	20.967	0.000
My parents check if I do my homework	0.876	0.094	9.339	0.000

Table C. The values of R^2 for the variables related to students' attitude towards their school

Observed Variables	Estimate	S.E.	z	*p*
Student-Level (Within)				
Proud with their school	0.469	0.038	12.381	0.000
Eager to see their classmates at school	0.673	0.045	14.965	0.000
Liking teachers in their school	0.408	0.034	12.025	0.000
School-Level (Between)				
Proud with their school	0.895	0.078	11.461	0.000
Eager to see their classmates at school	0.962	0.073	13.270	0.000
Liking teachers in their school	0.631	0.108	5.823	0.000

Table D. The values of R^2 for the variables related to students' attitude towards learning Mathematics

Observed Variables	Estimate	S.E.	z	*p*
Student-Level (Within)				
I enjoy learning mathematics	0.202	0.035	5.844	0.000
I wish to study mathematics	0.010	0.009	1.132	0.258
I learn interesting things in mathematics	0.522	0.053	9.788	0.000
It is important to do well in mathematics	0.301	0.046	6.592	0.000
School-Level (Between)				
I enjoy learning mathematics	0.487	0.205	2.372	0.018
I wish to study mathematics	0.021	0.071	0.296	0.767
I learn interesting things in mathematics	0.979	0.007	132.530	0.000
It is important to do well in mathematics	0.572	0.208	2.745	0.006

Table E. The values of R^2 for the variables related to the students' attitude of confidence in learning Mathematics

Observed Variables	Estimate	S.E.	z	*p*
Student-Level (Within)				
I usually do well in mathematics	0.500	0.039	12.774	0.000
Mathematics is more difficult for me than my classmates	0.395	0.038	10.378	0.000
I learn things quickly in mathematics	0.451	0.039	11.600	0.000
Mathematics makes me nervous	0.297	0.038	7.892	0.000
My teacher tells me I am good at mathematics	0.327	0.027	12.098	0.000
Mathematics is harder for me than any other subject	0.435	0.038	11.547	0.000
I would like a job that involves using mathematics	0.324	0.030	10.737	0.000
School-Level (Between)				
I usually do well in mathematics	0.967	0.049	19.857	0.000
Mathematics is more difficult for me than my classmates	0.963	0.054	17.676	0.000
I learn things quickly in mathematics	0.870	0.068	12.860	0.000
Mathematics makes me nervous	0.871	0.068	12.811	0.000
My teacher tells me I am good at mathematics	0.772	0.081	9.511	0.000
Mathematics is harder for me than any other subject	0.965	0.063	15.423	0.000
I would like a job that involves using mathematics	0.752	0.108	6.945	0.000

Table G. The values of R^2 for the variables related to lack of school resources

Observed Variables	Estimate	S.E.	z	*p*
Lack of instructional Materials	0.886	0.202	4.395	0.000
Inadequate school Facilities	0.654	0.185	3.540	0.000
Lack of competent teachers	0.289	0.109	2.657	0.008
Mathematics achievement	0.015	0.041	0.378	0.705

Table H. The values of R^2 for the variables related to schools' emphasis on academic success

Observed Variable	Estimate	S.E.	z	*p*
Teachers emphasis on academic success	0.492	0.166	2.968	0.003
Parents emphasis on academic success	0.549	0.175	3.136	0.002
Students emphasis on academic success	0.301	0.185	1.625	0.104
Mathematics achievement	0.115	0.101	1.132	0.258

Table I.The values of R^2 for the variables related to school discipline and safety

Observed Variable	Estimate	S.E.	z	*p*
School absenteeism	0.360	0.149	2.411	0.016
Classroom disturbance	0.682	0.180	3.784	0.000
Exam cheating	0.641	0.182	3.514	0.000
Mathematics achievement	0.005	0.024	0.205	0.838

Appendix VI. Mathematics Achievement Test Items Adapted From TIMSS

A. Mathematics Achievement Test Items Form 01 (English version)

MATHEMATICS ACHIEVEMENT IN TIMSS

ENGLISH - FORM 01

REGION ___________________

SCHOOL NAME__________________________

STUDENT: **SECTION** ____________ **R. N**<u>**o**</u> ____________

GENDER : **MALE** ☐ **FEMALE** ☐

April, 2019

Number of Items: 40 **Time Allowed: 90 minutes**

DIRECTION: Choose the correct answer and circle the letter of your *choice* from the given alternatives or *fill* in the blank spaces with the correct answer for each of the following questions

1. $\frac{4}{100} + \frac{3}{1000} =$__________.

   A. 0.043 C. 0.403

   B. 0.1043 D. 0.43

2. What is the sum of the measures of the interior angles of a 12-sided polygon?

   A. 2160^0 C. 1080^0

   B. 1800^0 D. 1620^0

3. The fractions $\frac{4}{14}$ and $\frac{x}{21}$ are equivalent. What is the value of x?

   A. 6 C. 11

   B. 7 D. 14

4. The simple interest on Birr 4000 invested for 4 years was Birr 480. What is the annual rate of the interest?

   A. 7 % C. 3%

   B. 5% D. 2%

5. Which number is equal to $\frac{3}{5}$?

   A. 0.8 C. 0.53

   B. 0.6 D. 0.35

6. The bases of a trapezium are 15cm and 10cm long and its altitude is 8cm. What is the area of the trapezium?

   A. $60\ cm^2$ C. $120\ cm^2$

B. $100\ cm^2$ D. $150\ cm^2$

7. $42.65 + 5.748 =$ _____________.

8. What is the sum of the three consecutive whole numbers with 2n as the middle number?

A. $6n + 3$ C. $6n - 1$

B. $6n$ D. $6n - 3$

9. If x is a number between 6 and 9, then x + 5 is between

A. 1 and 4 C. 11 and 14

B. 10 and 13 D. 30 and 45

10. Let a fair coin will be tossed once, then what is the probability of getting a tail?

A. $\frac{1}{5}$ C. $\frac{1}{2}$

B. $\frac{1}{4}$ D. $\frac{1}{3}$

11. Given $y = \frac{a+b}{c}$ and a = 8, b = 6, and c = 2. What is the value of y?

A. 7 C. 11

B. 10 D. 14

12. Which of the following operation is **NOT** closed under the set of positive rational numbers?

A. Addition C. Multiplication

B. Subtraction D. Division

13. The sum of the age of a boy and his sister is 24 years. If the boy is 4 years older than his sister, then how old is his sister?

A. 4 C. 14

B. 6 D. 10

14. The area of a square is 144 cm^2. What is the perimeter of the square?

A. 12 cm C. 288 cm

B. 48 cm D. 576 cm

15. What is the solution set of 2(x+1) =3(2x+3)?

A. {7/3} C. {8/11}

B. {-7/4} D. {-8/11}

16. Which one of the following is the measure of dispersion?

A. Range C. Mean

B. Mode D. Median

17. Which shows a correct method for finding $\frac{1}{3} - \frac{1}{4}$?

A. $\frac{1-1}{4-3}$ C. $\frac{3-4}{3\times 4}$

B. $\frac{1}{4-3}$ D. $\frac{4-3}{3\times 4}$

18. Solve the inequality $9x - 6 < 4x + 4$. Answer: _____________.

19. What is the product of (3-x) and $2x^2$?

A. $6x^2 - 2x$ C. $6x^2-2x^3$

B. $6x^3-2x^2$ D. $5x^2-2x^3$

20. Which of the following is equal to $3p^2 + 2p + 2p^2 + p$?

A. 8p C. $5p^2 + 3p$

B. $8p^2$ D. $7p^2 + p$

21. In which quadrant does the point P (3, -5) lies?

A. 1st quadrant C. 3rd quadrant

B. 2nd quadrant D. 4th quadrant

22. There were m boys and n girls in a class. Each of them carried two books. Which of the following expressions represents the total number of books that they carried?

A. 2(m + n)

B. 2 + (m + n)

C. 2m + n

D. m + 2n

23. Given: k =7, t =10. What is the value of p, when $p = \frac{3kt}{5}$? Answer: _______________.

24. A number will be selected at random from {2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12}. What is the probability of getting an odd number?

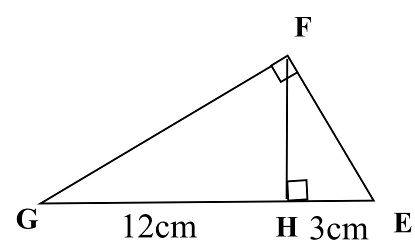
A. $\frac{4}{11}$

B. $\frac{5}{11}$

C. $\frac{6}{11}$

D. $\frac{7}{11}$

25. For the right-angled triangle of ΔEFG shown in the figure below, $\overline{FH}$ is the altitude to its hypotenuse. If $\overline{GH}$ = 12cm, $\overline{HE}$= 3cm, then $\overline{GF}$ is equal to



A. 6cm

B. 15cm

C. 36cm

D. $6\sqrt{5}$ cm

26. There are 10 red, 8 blue, and 4 white balls in a bag. What is the probability of taking out either a blue or a white ball?

A. $\frac{4}{22}$

B. $\frac{8}{22}$

C. $\frac{10}{22}$

D. $\frac{12}{22}$

27. Which one of the following is true?

A. $\frac{1}{2} > \frac{2}{3}$

C. $\frac{5}{6} < \frac{6}{7}$

B. $\frac{3}{4} > \frac{4}{5}$

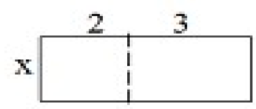
D. $\frac{5}{4} < \frac{2}{3}$

28. Which of the following could represent the expression 2x+3x?

A. The length of this segment

B. The length of this segment:

C. The area of this figure:



D. The area of this figure:

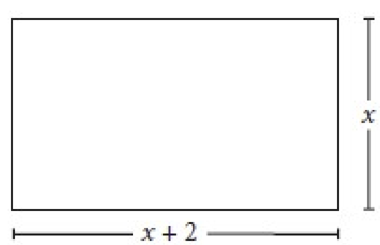
29. What is the area of the rectangle shown below?

A. $x^2 + 2$

B. $x^2 + 2x$

C. $2x + 2$

D. $4x + 4$



30. Out of 80 students in a class 30 are boys. What is the ratio of boys to girls?

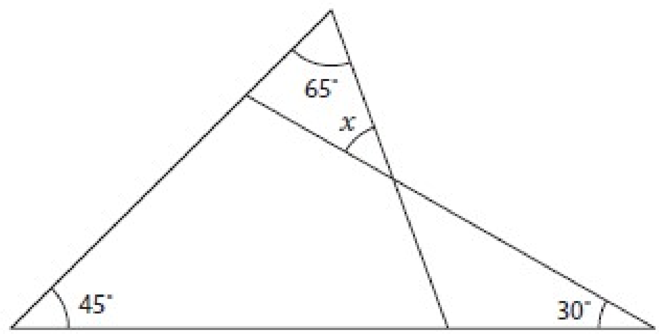
A. 8:5

C. 3:5

B. 5:8

D. 5:3

Answer question 31, based on the figure below.

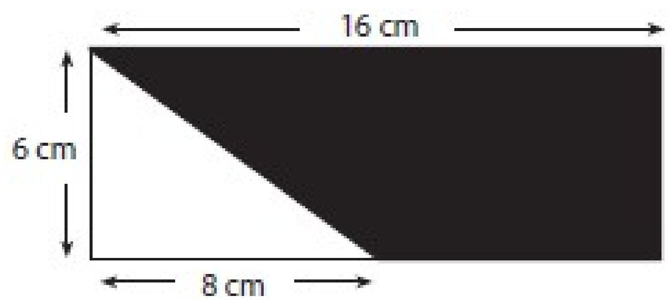


31. What is the value of x?

A. 30° C. 45°

B. 40° D. 65°

32. In the figure below, what is the area of the shaded region in cm^2?



A. 24

B. 44

C. 48

D. 72

33. If $\sqrt{8} \approx 2.83$, then $\sqrt{800} \approx$ ________

A. 40.0 C. 28.3

B. 283 D. 400

34. The sum of the degree measures of all the interior angles of the pentagon is _________.

35. What is the highest common factor (HCF) of $4x^3y$ and $6xy^3$?

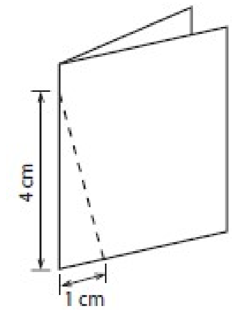
A. 2xy C. $6x^3y$

B. 4xy D. $3x^3y$

36. Aster is packing eggs into boxes. Each box holds 6 eggs. She has 94 eggs. What is the smallest number of boxes she needs to pack all the eggs?

A. 14
B. 15
C. 16
D. 17

Answer question 37 based on the figure below



37. A piece of paper in the shape of a rectangle is folded in half as shown in the figure above. It is then, cut along the dotted line, and the small piece that cut is opened. What is the shape of the cutout figure?

A. An isosceles triangle
B. Two isosceles triangles
C. A right triangle
D. An equilateral triangle

38. Which of the following is TRUE about the image of a plane figure, where k is the constant of proportionality?

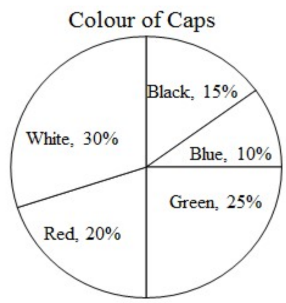
A. If k<1, then the image figure is larger than the original figure

B. If k= 0, then the image figure is the same as the original figure

C. If k > 1, then the image figure is smaller than the original figure

D. If k=1, then the image figure is congruent to the original figure

Refer to the following pie chart to answer question 39 below.



39. The pie chart shows the percentage of caps for sale at a sports shop. If there are 200 caps, what is the total number of caps that are either white or green?

A. 55 C. 110

B. 100 D. 145

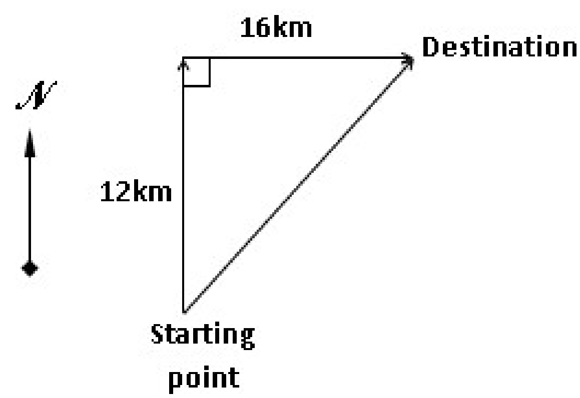
40. A bus traveled 12 km due North and then 16 km due East to reach its destination (see the figure below). How far is the destination of the bus from its starting point?

A. 18km

B. 20km

C. 24km

D. 28km



B. Mathematics Achievement Test Items Form 02 (English version)

MATHEMATICS ACHIEVEMENT IN TIMSS

ENGLISH - FORM 02

REGION ____________________

SCHOOL NAME__________________________

STUDENT: **SECTION** ____________ **R. No** ____________

GENDER : **MALE** ☐ **FEMALE** ☐

April, 2019

Number of Items: 40 **Time Allowed: 90 minutes**

DIRECTION: Choose the correct answer and circle the letter of your ***choice*** from the given alternatives or ***fill*** in the blank spaces with the correct answer for each of the following questions

1. Which fraction is equivalent to 0.125?

A. $\frac{25}{100}$ C. $\frac{25}{10{,}000}$

B. $\frac{25}{1000}$ D. $\frac{25}{100{,}000}$

2. What is the sum of the measures of the interior angles of a 12-sided polygon?

A. 2160^0 C. 1080^0

B. 1800^0 D. 1620^0

3. Which of the following is the BEST estimate of $\frac{7.21\times 3.86}{10.09}$?

A. $\frac{7\times 3}{10}$ C. $\frac{7\times 3}{11}$

B. $\frac{7\times 4}{10}$ D. $\frac{7\times 4}{11}$

4. The simple interest on Birr 4000 invested for 4 years was Birr 480. What is the annual rate of the interest?

A. 7 % C. 3%

B. 5% D. 2%

5. Which of the following shows a prime factorization of 36?

A. 6×6 C. $4 \times 3 \times 3$

B. 4×9 D. $2 \times 2 \times 3 \times 3$

6. The bases of a trapezium are 15cm and 10cm long and its altitude is 8cm. What is the area of the trapezium?

A. $60\ cm^2$ C. $120\ cm^2$

B. $100\ cm^2$ D. $150\ cm^2$

7. Which expression is equivalent to $4(3 + x)$?

A. $12 + x$ C. $12 + 4x$

B. $7 + x$ D. $12x$

8. What is the decimal form of $\frac{35}{6}$? Answer: _________. (Rounded to 2 decimal places)

9. Which of the following is true?

A. $\frac{3}{10}$ of 50 = 50% of 3 C. $50 \div 30 = 30 \div 50$

B. 3% of 50 = 6% of 100 D. $\frac{3}{10} \times 50 = \frac{5}{10} \times 30$

10. Let a fair coin will be tossed once, then what is the probability of getting a tail?

A. $\frac{1}{5}$ C. $\frac{1}{2}$

B. $\frac{1}{4}$ D. $\frac{1}{3}$

11. What does $xy + 1$ mean?

A. Add 1 to y, then multiply by x. C. Add x to y, then add 1.

B. Multiply x and y by 1. D. Multiply x by y, then add 1.

12. Which of the following operation is **NOT** closed under the set of positive rational numbers?

A. Addition C. Multiplication

B. Subtraction D. Division

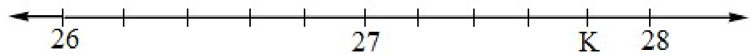
13. The sum of the age of a boy and his sister is 24 years. If the boy is 4 years older than his sister, then how old is his sister?

A. 4 C. 14

B. 6

D. 10

Look at the number line below to answer question 14.



14. Which of the following does K represent on this number line?

A. 27.4 C. 27.9

B. 27.8 D. 28.2

15. What is the solution set of 2(x+1) =3(2x+3)?

A. {7/3} C. {8/11}

B. {-7/4} D. {-8/11}

16. Which one of the following is the measure of dispersion?

A. Range C. Mean

B. Mode D. Median

17. A man cut off 3/5 of a rope. The piece he cut off was 3 meters long. How many meters long was the original rope?

A. 8 m C. 15 m

B. 12 m D. 18 m

18. The simplified form of $\frac{3x}{8} + \frac{x}{4} + \frac{x}{2}$ is ______________________.

19. What is the product of (3-x) and $2x^2$?

A. $6x^2 - 2x$ C. $6x^2-2x^3$

B. $6x^3-2x^2$ D. $5x^2-2x^3$

20. A piece of wood was 40 cm long. It was cut into 3 pieces. The lengths in cm are:

2x – 5, x + 7, x + 6. What is the length of the longest piece?

A. 8 C. 1

B. 11 D. 15

21. In which quadrant does the point P (3, -5) lies?

A. 1^{st} quadrant C. 3rd quadrant

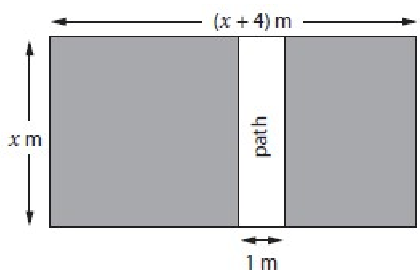
B. 2^{nd} quadrant D. 4^{th} quadrant

22. Which of the following equation is satisfied by BOTH of the ordered pairs (0, −1) and (1, 3)?

A. $x + y = -1$ C. $3x - y = 0$

B. $2x + y = 5$ D. $4x - y = 1$

Below is a diagram of a rectangular garden. The white area is a rectangular path that is 1 meter wide. Answer question 23 with this information.



23. Which expression shows the area of the shaded portion of the garden in m^2?

A. $x^2 + 3x$ C. $x^2 + 4x - 1$

B. $x^2 + 4x$ D. $x^2 + 3x - 1$

24. A number will be selected at random from {2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12}. What is the probability of getting an odd number?

A. $\frac{4}{11}$ C. $\frac{6}{11}$

B. $\frac{5}{11}$ D. $\frac{7}{11}$

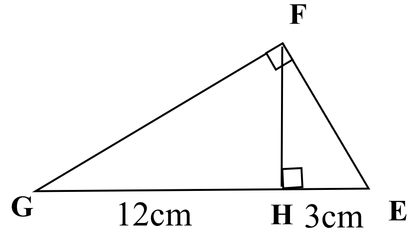
25. For the right-angled triangle of ΔEFG shown in the figure below, $\overline{FH}$ is the altitude to its hypotenuse. If $\overline{GH}$ = 12cm, $\overline{HE}$ = 3cm, then $\overline{GF}$ is equal to

A. 6cm

B. 15cm

C. 36cm

D. $6\sqrt{5}$ cm



26. If a + b = 25. What is the value of 2a + 2b + 4?

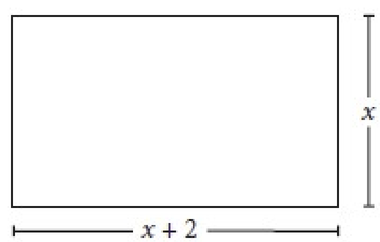
A. 29
B. 31
C. 50
D. 54

27. Which one of the following is true?

A. $\frac{1}{2} > \frac{2}{3}$
B. $\frac{3}{4} > \frac{4}{5}$
C. $\frac{5}{6} < \frac{6}{7}$
D. $\frac{5}{4} < \frac{2}{3}$

28. What is the area of the rectangle shown below?

A. $x^2 + 2$
B. $x^2 + 2x$
C. $2x + 2$
D. $4x + 4$



29 Given $y = 100 - \frac{100}{1 + x}$, what is the value of y when x = 9. Answer: ______________

30. Out of 80 students in a class 30 are boys. What is the ratio of boys to girls?

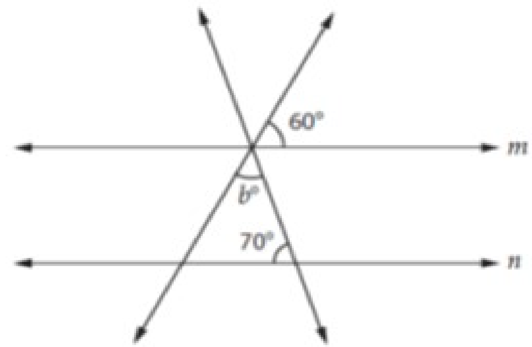
A. 8:5
B. 5:8
C. 3:5
D. 5:3

31. The perimeter of a square is 36 cm. What is the area of this square?

A. 81 cm^2
B. 36 cm^2
C. 24 cm^2
D. 18 cm^2

32. As shown in the figure below, if lines m and n are parallel, what is the value of b?

A. 40^0
B. 50^0
C. 60^0
D. 70^0

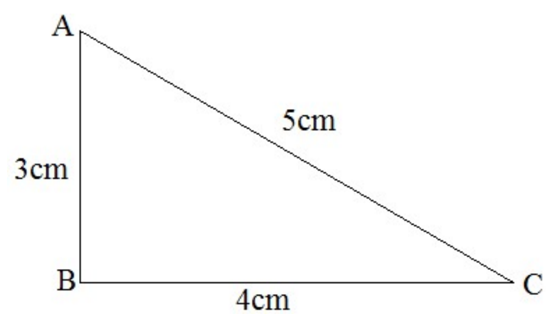


33. If $\sqrt{8} \approx 2.83$, then $\sqrt{800} \approx$ ________

A. 40.0
B. 283
C. 28.3
D. 400

34. Which of the following is the reason for the triangle ABC below is a right-angle triangle?

A. $3^2 + 4^2 = 5^2$
B. $5 < 3 + 4$
C. $3 + 4 = 12 - 5$
D. $3 > 5 - 4$

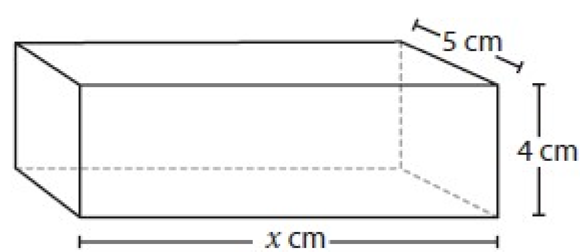


35. What is the highest common factor (HCF) of $4x^3y$ and $6xy^3$?

A. 2xy
B. 4xy
C. $6x^3y$
D. $3x^3y$

36. If the volume of the rectangular box in the figure below is 200 cm^3, what is the value of x?

A. 10
B. 20
C. 100
D. 200



37. Almaz and Ahmed divide 560 Birr between them. If Ahmed gets 3/8 of the money, how many Birr will Almaz get? Answer: ____________.

38. Which of the following is TRUE about the image of a plane figure, where k is the constant of proportionality?

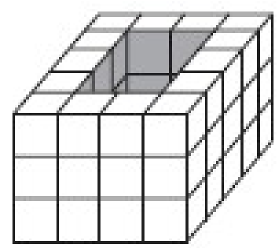
A. If $k<1$, then the image figure is larger than the original figure

B. If $k= 0$, then the image figure is the same as the original figure

C. If $k > 1$, then the image figure is smaller than the original figure

D. If $k=1$, then the image figure is congruent to the original figure

39. The figure below shows a shape made up of cubes that are all the same size. There is a hole all the way through the shape. How many cubes would be needed to fill the hole?



A. 6

B. 12

C. 15

D. 18

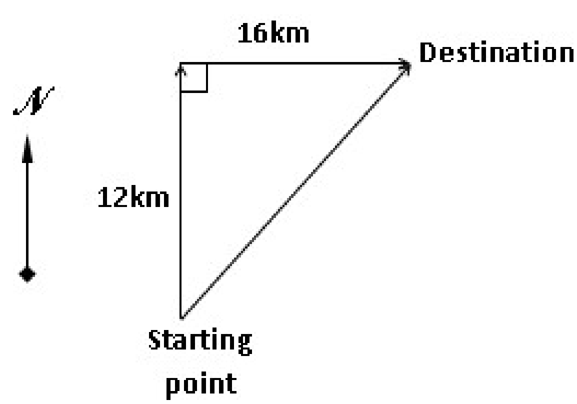
40. A bus traveled 12 km due North and then 16 km due East to reach its destination (see the figure below). How far is the destination of the bus from its starting point?

A. 18km

B. 20km

C. 24km

D. 28km



Appendix VII. Student Contextual Questionnaire (English Version)

This questionnaire addressed to grade eight students who asked to supply information about themselves, their parents, schools and learning Mathematics in order to conduct a research entitled: ***Ethiopian primary school students' Mathematics achievement through the lens of the Trends International Mathematics and Science Study.*** Since you have selected as a sample randomly, your responses are very vital in helping to find out factors related to the mathematics achievement of primary school students in Ethiopia.

Thus, it is important that you answer each question carefully so that the information provided reflects the situation in your school as accurately as possible. It needs few minutes to complete this questionnaire. Therefore, I appreciate the time and effort that this takes and thank you for your cooperation and contribution.

Thank you.

I. Background Information

For each of the following questions, choose the response that you think right and circle the letter of your choice or fill in the blank spaces with the appropriate answers.

1. Gender: A. Male B. Female
2. Age: _ ______________ years
3. How often do you speak the language of instruction at home?
   A. Always B. Sometimes C. Never
4. Highest level of education completed by your father or male guardian
   A. Above Grade 12 D. Able to read and write
   B. Grade 9 - 12 E. Illiterate
   C. Grade 1 - 8
5. Highest level of education completed by your mother or female guardian
   A. Above Grade 12 D. Able to read and write
   B. Grade 9 - 12 E. Illiterate
   C. Grade 1 - 8

6. About how many books are there in your home? (Do not count magazines, newspapers, or your school books)

A. Never D. 11-25 books

B. 1-5 books E. 26-100 books

C. 6-10 books F. 100 books

7. Is there anyone who supports you while you are studying at your home?

A. Yes B. No

8. If your answer to question number 8 is "yes", how many days per week, you get support in your study?

A. 7 days B. 4-6 days C. 1- 3 days

9. In this academic year, how frequently you have attended extra lessons or tutoring not provided by the school in Mathematics?

A. Many times B. few times C. No at all

10. Indicate the availability of the following materials at your home by choosing a letter "A" or "B".

No	Types Materials	Choice	
10.1	Bicycle	A. Yes	B. No
10.2	Telephone/ Mobile	A. Yes	B. No
10.3	Watch	A. Yes	B. No
10.4	Motor bicycle	A. Yes	B. No
10.5	Vehicle	A. Yes	B. No
10.6	Radio	A. Yes	B. No
10.7	Television	A. Yes	B. No
10.8	Refrigerator	A. Yes	B. No
10.9	Table	A. Yes	B. No
10.10	Bed with Mattresses	A. Yes	B. No

10.11. Approximately, how did you rate the economic status of your family?

A. High B. Medium C. Low

11. How many times do you eat your meals per school day?

A. One B. Two C. Three D. More than three

12. How often do the following things happen at home? Choose the letter of your choice as shown below.

No	Statements	Every day	2 or 3 days per week	2 or 3 days per month	Never
12.1	My parents ask me what I am learning in school	A	B	C	D
12.2	I talk about my schoolwork with my parents	A	B	C	D
12.3	My parents make sure that I set aside time for my homework	A	B	C	D
12.4	My parents check if I do my homework	A	B	C	D

II. Your School

13. What do you think about your school? Indicate the degree of your agreement for each of the following statements by choosing the letter of your options.

No	Statements	Strongly Agree	Agree	Disagree	Strongly Disagree
13.1	I like being in school	A	B	C	D
13.2	I feel safe when I am at school	A	B	C	D
13.3	I like to see my classmates at school	A	B	C	D
13.4	Teachers at my school are fair to me	A	B	C	D
13.5	I am proud to go to this school	A	B	C	D
13.6	I learn a lot in school	A	B	C	D
13.7	My teachers understand the problem of the students	A	B	C	D
13.8	I am happy with the ethics of my teachers	A	B	C	D

III. About the Lesson

14. How much do you travel from home to school?

A. Below 15 min. D. 1:01 – 1:30 hrs.

B. 15 - 30 min. E. Above 1:30 hrs.

C. 31- 60 min.

15. For the last two months in this academic year, about how many days you were absent from school?

A. No absence C. 6- 10 days

B. 1- 5 days D. More than10 days

16. How further do you want to study in the future?

A. Complete Grade 8 D. Graduate in Certificate/Diploma

B. Complete Grade 10 E. Graduate in First Degree

C. Complete Grade 12 F. Above First Degree

17. How much do you agree with the following statements about learning mathematics? Indicate your response by choosing the letter of your choice.

No	Statements	Strongly Agree	Agree	Disagree	Strongly Disagree
17.1	I enjoy learning mathematics	A	B	C	D
17.2	I wish I did not have to study mathematics	A	B	C	D
17.3	Mathematics is boring	A	B	C	D
17.4	I learn many interesting things in mathematics	A	B	C	D
17.5	I like mathematics	A	B	C	D
17.6	It is important to do well in mathematics	A	B	C	D

18. How much do you agree with the following statements about mathematics? Indicate your response by choosing the letter of your choice.

No	Statements	Strongly Agree	Agree	Disagree	Strongly Disagree
18.1	I usually do well in mathematics	A	B	C	D
18.2	Mathematics is more difficult for me than for many of my classmate	A	B	C	D
18.3	I learn things quickly in mathematics	A	B	C	D
18.4	Mathematics makes me nervous	A	B	C	D
18.5	My teacher tells me I am good at mathematics	A	B	C	D
18.6	Mathematics is harder for me than any other subject	A	B	C	D
18.7	I think learning mathematics will help me in my daily life	A	B	C	D
18.8	I need mathematics to learn other school subjects	A	B	C	D
18.9	I need to do well in mathematics to get the job I want	A	B	C	D
18.10	I would like a job that involves using mathematics	A	B	C	D

19. How often does your teacher give you homework in Mathematics?

A. Every day D. Less than once a week

B. 3 or 4 times a week E. Never

C. 1 or 2 times a week

20. When your teacher gives you Mathematics homework, about how many minutes do you usually spend on your homework?

A. 1–15 minutes D. 61–90 minutes

B. 16–30 minutes E. More than 90 minutes

C. 31–60 minutes

Appendix VIII. School Principals Contextual Questionnaire

This questionnaire addressed school principals who asked to supply information about their schools. Since your school has been selected as a sample, your responses are very important in helping to describe primary education in Ethiopia. It is important that you answer each question carefully so that the information provided reflects the situation in your school as accurately as possible. It needs few minutes to complete this questionnaire.

Therefore, I appreciate the time and effort that this takes and thank you for your cooperation and contribution.

Thank you.

I. Background Information

1. Sex A. Male B. Female

2. Age ______

3. Qualification:

   A. Certificate C. First Degree

   B. Diploma D. Master's Degree

4. Experience:

   4.1. By the end of this school year, how many years will you have been a principal altogether? ________Years.

   4.2. By the end of this school year, how many years will you have been a principal at this school? _________Years.

II. School Enrollment and Characteristics

5. What is the total number of students in your school by this time?

Male _______ Female________ Total ___________

6. What is the total enrollment of grade 8 students in your school by this time?

Male _______ Female________ Total ___________

7. What is the average number of grade 8 students per classroom in your school? ____students.

8. Approximately what percentage of students in your school have the following backgrounds?

8.1. Come from economically poor families

A. 0 to 10% C. 26 to 50%

B. 11 to 25% D. More than 50%

8.2. Come from economically affluent/well family

A. 0 to 10% C. 26 to 50%

B. 11 to 25% D. More than 50%

9. Does your school provide free meals for students?

9.1. Breakfast

A. Yes, for all students B. Yes, for some students C. No

9.2. Lunch

A. Yes, for all students B. Yes, for some students C. No

III. Instructional Time

10. For the eighth grade students in your school:

10.1. How many days per year is your school open for instruction? __________days

10.2. What is the total instructional time, excluding breaks, per day? ______minutes

10.3. In one calendar week, how many days is the school open for instruction? ______days

11. Does your school provide a place where students can work on their schoolwork before or after school?
A. Yes B. No

12. If your answer is yes for Q11, is someone available to assist them with their schoolwork?
A. Yes B. No

IV. Resources and Technology

13. How many computers does your school have for use by eighth grade students?
_____ Computers

14. Does your school have a school library? A. Yes B. No

15. If your response is yes for Q14, approximately how many books with different titles does your school library has (exclude magazines)?

A. 250 or fewer B. 251–500 C. 501–2,000 D. 2,001–5,000 E. More than 5,000

16. How much is your school's capacity to provide instruction affected by a **shortage** of the following **resources**? Put a tick mark (✓)

School resources	Not at all	A little	some	A lot
a. Instructional materials (e.g., textbooks,)				
b. Supplies (e.g., papers, pencils, materials)				
c. School buildings and grounds				
d. Lighting systems/electricity				
e. Instructional space (e.g., classrooms)				
f. Competent staff				
g. Computer technology for teaching and learning				
h. Resources for students with disabilities				
i. Resources for Mathematics instruction				
j. Teachers with a specialization in Mathematics				
k. Library resources relevant to mathematics instruction				
l. Concrete objects or materials to help students understand quantities or procedures				

V. School Emphasis on Academic Success

17. How would you characterize each of the following within your school? Put a tick mark (✓)

Statements	V. High	High	Medium	Low	V. Low
a. Teachers' understanding of the curricular objectives					
b. Teachers' degree of success in implementing the curriculum					
c. Teachers' expectations for student achievement					
d. Teachers' ability to motivate students					
e. Parental involvement in school activities					
f. Parental commitment to ensure that students are ready to learn					
g. Parental expectations for student achievement					
h. Parental support for student achievement					
i. Parental pressure for the school to maintain high academic standards					
j. Students' desire to do well in school					
k. Students' ability to reach the set academic goals					
l. Students' respect for their classmates					

VI. School Discipline and Safety

18. To what degree is each of the following a problem among grade 8 students in your school? Put a tick mark (✓)

Problems	Not a problem	Minor problem	Moderate problem	Serious problem
a. Arriving late at school				
b. Absenteeism (i.e., unjustified absences)				
c. Classroom disturbance				
d. Cheating				
e. Destruction				
f. Verbal abuse among students				
g. Physical injury among students				
h. Verbal abuse of teachers				
i. Physical injury to teachers or staff				