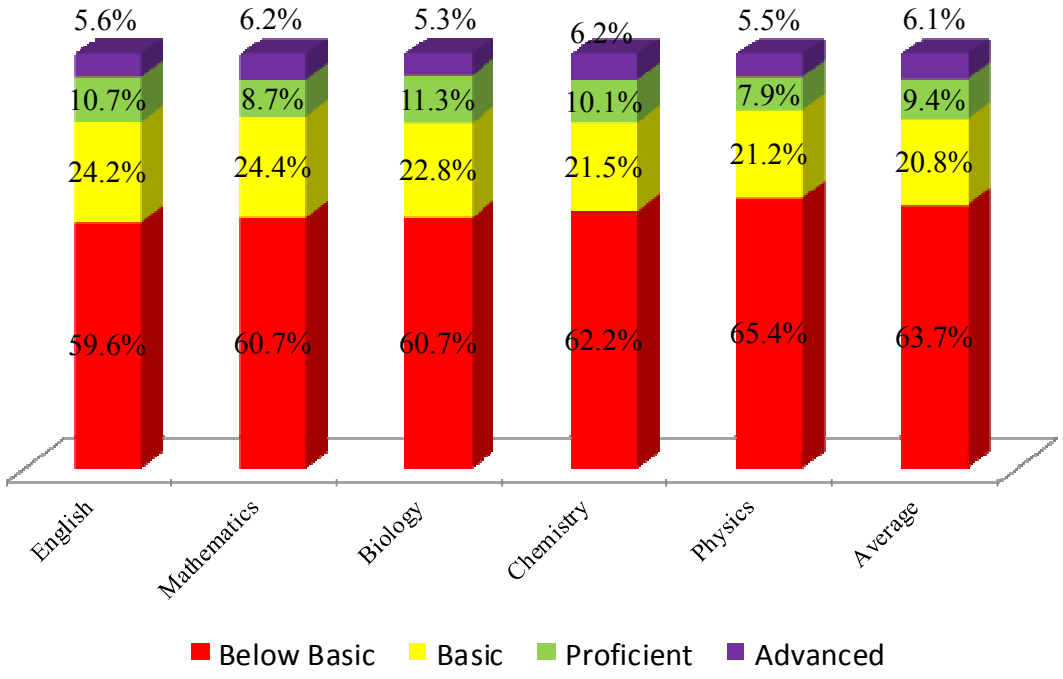
Ethiopian First National Learning Assessment of Grades 10 and 12 Students



National Agency for Examinations

December, 2010
Addis Ababa

This page is left blank intentionally.

Ethiopian First National Learning Assessment of Grades 10 and 12 Students

A study carried out by the National Agency for Examinations

Prepared by:

Zewdu Gebrekidan

December, 2010
Addis Ababa
Ethiopia

This page is left blank intentionally.

Table of Contents

List of Tables

List of Figures

Acknowledgements

The Ministry of Education initiated this study and we are very grateful to H.E. Ato Fuad Ibrahim, State Minister of Education, for his understanding and continued support right from the beginning and throughout the project. We would like to thank the experts from General Education Quality Assurance and Examination Agency (GEQAEA), Curriculum Department of MOE and teachers from different high schools for their role in curriculum analysis, standard setting and test development. The contribution of the experts from GEQEA who were also data collection coordinators, the drivers, the center coordinators and test administrators from the regions are highly commendable.

Special thanks are due to W/o Mulat Azene, secretary at GEQEA, beside with typing the data collecting instruments and capturing the data from the questionnaires she coordinated several activities related to data collection and organization. She was highly stretched and supportive in all major activities and contributed a lot.

The Ministry of Education was undertaking a Business Process Reengineering and the National Learning Assessment was reorganized under the Education and Training Quality Assurance Agency (ETQAA). During the finalization of this report the assessment team members have made comments on the draft report and organized validation workshop.

National Learning Assessment is once again under the National Agency for Examinations and this publication would not have been possible without the support of the Agency's Director, Ato Araya Gebregziabher. We are very grateful to him for his hospitality and strong interest in the National Learning Assessment.

Finally, we would like to thank all students, their teachers and school directors who enthusiastically participated in the assessment.

Front Cover: Grade 10 Proficiency Levels by Subject
Back Cover: Grade 12 Proficiency Levels by Subject

Abbreviations

CTT	Classical Test Theory
EBNLA	Ethiopian Baseline National Learning Assessment
EFA	Education for All
EMIS	Educational Management Information Service
ESDP	Education Sector Development Program
ESNLA	Ethiopia Second National Learning Assessment
ETNLA	Ethiopian Third National Learning Assessment
ETP	Education and Training Policy
IRT	Item Response Theory
GEQAEA	General Education Quality Assurance and Examinations Agency
GEQIP	General Education Quality Assurance and Improvement Program
HLM	Hierarchical Linear Modeling
IIEP	International Institute for Education Planning
MLA	Monitoring Learning Achievement
MLC	Minimum Learning Competency
MOE	Ministry of Education
NOE	National Organization for Examinations
OMR	Optical Mark Reader
SNNPR	Southern Nations, Nationalities and Peoples Region
UNESCO	United Nations Educational, Scientific and Cultural Organization
USAID	United States Agency for International Development

Executive Summary

National learning assessments carried out in different countries are involved in assessing what and how much students are taught, and how much they learn and the conditions of that best facilitates effective learning outcomes. National learning assessments evaluate learning outcomes based on criteria and expectations set forth by national education authorities. National assessments (sometimes called system assessment, learning assessment, and assessment of learning outcomes) may be defined:

> *'as an exercise designed to describe the level of achievements, not of individual students, but of a whole education system, or a clearly defined part of it (e.g. fourth grade pupils or 11-year olds).'* [1]

They are intended to provide national policymakers with systematic information about the status of students' learning outcomes and the extent to which students attain pre-defined standards or proficiencies.

The Ethiopian Baseline National Learning Assessments of Grades 10 and 12 students were initiated by the Ministry of Education and carried out in April 2009 (2001 E.C.) across the nation in all the regions. The State Minister of Ministry of Education, responsible for General of Education, closely followed the whole undertaking.

Ethiopia has carried out three national learning assessments at the first and second cycles of primary education since 2000 (1993 E.C.). This study was the first of its kind to be conducted at secondary education level. The main purposes of the studies were to determine what secondary students know upon completion of General Secondary Education (Grade 10) and the Preparatory Program (Grade 12) in light of the Minimum Learning Competencies set by the Ministry of Education. The achievement levels of secondary school students in selected subjects were described and possible factors that have relations with the achievement scores were identified.

[1] Greany and Kellaghan (2004)

Objectives

The main purposes of the assessments of Grade 10 and 12 students were to measure the academic achievement levels, detect subgroup differences, and identify in-school and out-of-school factors related to achievement scores. Based on the findings, the study tried to answer the following questions:

1. What are the achievement levels of Grades 10 and 12 students in English, mathematics, biology, chemistry and physics?
2. Are there differences in the achievement scores of the tested subjects across subgroups?
3. Are there any relationships between achievement scores and different student and school level variables?
4. What are the major factors that contributed most to variations in achievement scores?

Methods

The Sample

Sample schools and students were drawn following a stratified cluster sampling procedure proportion to the size of the regions. In Grade 10, a total of 140 schools and 5,583 students while in Grade 12, a total of 73 schools and 2,812 students participated in the study sat for the achievement tests. The sample students, their teachers and the school directors responded to questionnaires and perceptioners.

Achievement Measures

Standardized achievement tests in English, mathematics, biology, chemistry and physics were administered at each grade level. These tests were developed based on the MLC (Minimum Learning Competencies) of General Secondary Education (Grades 9 and 10) and Preparatory Program (Grades 11 and 12). A team of curriculum experts, test development experts and experienced teachers guided by assessment specialists developed the tests in the five subjects for both grade levels.

A pilot study was carried out in April 2008, in selected schools. Item and test analyses were carried out following the CTT (Classical Test Theory) and IRT (Item Response Theory) procedures. A validation workshop that brought together the team of experts and subject teachers involved in the test development process was called and the test booklets were finalized. In finalizing the instruments, both the psychometrics statistics and expert judgments were considered simultaneously.

Explanatory measures

Sample students, their teachers and the school directors also responded to questionnaires that addressed a range of in-school and out-of-school factors believed to affect achievement scores. Students were asked about themselves, their family conditions, interest and activities they carry out during their spare time among others. Teachers and school directors were also asked about their students and the school environment.

Data management and organization

Data collection coordinators were selected from GEQAEA (General Education Quality Assurance and Examinations Agency) and participated in a training organized in Addis Ababa and traveled with the booklets to the regions they were assigned. Data collectors were recruited from each region, received training at selected training centers, and traveled to sample schools. Each data collector stayed at least three days at one sample school. The answer sheets were read using OMR (Optical Mark Reader) and the responses from questionnaires were captured by personal computers.

Data analysis

The data analysis techniques used were: item and test analyses, summary descriptive statistics, independent sampled t-tests, analysis of variance, correlations, multiple regression analyses and variance partitioning. Regular and survey data analysis statistical packages such as *SPSS v15, Systat v12* and *Stata v10* were used.

The Findings

Overall achievement scores

The academic achievement of the students as measured by the mean score of the five subjects namely English, mathematics, biology, chemistry and physics was found less than the 50% achievement level set by the Education and Training Policy of Ethiopia. The national mean score (the average of what the students scored in the five subjects) was only 36% for Grade 10 and 47.8% for Grade 12.

Looking at the mean score of each subject, in Grade 10 in all the subjects the mean scores were below the minimum requirement (50%). In Grade 12, in two of the five subjects namely biology and mathematics the mean scores were found higher than 50%.

In Grade 10, in the average of the five subjects only 13.8% scored fifty percent and above. Looking at each subject, the percentages achieving fifty percent and above were only 17.8% in English, 14.7% in mathematics, 24.8% in biology, 17.1% in chemistry and 10.1% in physics.

In Grade 12, in the average of the five subjects only 34.9% scored fifty percent and above. Looking at each subject, the percentages achieving fifty percent and above were only 25.9% in English, 57.7% in mathematics, 60.7% in biology, 44.4% in chemistry and 16.5% in physics.

The standard deviation of the five subjects' mean scores ranges from 13.62% in English to 16.52% in biology. This shows the existence of very wide variation in the mean score of each subject. In terms of distribution the mean score of each subject was found positively skewed indicating that only very few pupils were able to get highest scores.

In Grade 10, only 10% got 54.9% and above in the average score while 50% of the students scored between 28.9% and 35.6%.On the other hand, 10% of the examinees scored at or below chance level in all subjects. In multiple choice tests with four options the chance level is 25%.

In Grade 12, only 10% got 65.2% and above in the average score while 50% of the candidates scored between 38.9% and 55.8%. On the other hand, 10% of the examinee scored 32.4% and less.

Achievement by sex

In all subjects in both Grades 10 and 12 boys performed better than girls and the differences were statistically significant in all cases. In Grade 10, there were 3,151 males and 2,260 females in the national sample. Boys scored an average of 38.7% whereas girls had an average of 32.2%. In Grade 12, there were 2.017 males and 561 females in the national sample. Boys scored an average of 49.2% whereas girls had an average of 42.6%.

Achievement by performance levels

The students' achievement scores were divided into four standards as: 'below basic', 'basic', 'proficient' and 'advanced'. The scaled scores were used instead of raw scores where: 'advanced' is greater than 2 standard deviations from the mean, 'proficient' is between 1 and 2 standard deviations above the mean, 'basic' is within 1 standard deviation above the mean and 'below basic' is below the mean score.

In Grade 10, the proportion achieving each level based on the national achievement of the average score were: 6.1% 'advanced', 9.4% 'proficient', 20.8% 'basic', and the remaining 63.7% below basic level.

In Grade 12, the proportion achieving each level based on the national achievement of the average score were: 3.8% 'advanced', 12.7% 'proficient', 28.3% 'basic', and the remaining 55.3% 'below basic' level.

Region level achievement

The average of the five subjects and mean score of each subject when disaggregated across regions showed that there exist disparities in academic achievement among regions. In Grade 10, no single region achieved the minimum requirement of at least 50%. The mean scores in Gambella (28.5%) and Afar (29.4%) were found much lower when compared with the highest achieving regions. In Grade 12, Dire Dawa

was the only region that achieved a mean score greater than 50%. Somali (40.3%), Gambella (43.5%), and SNNP (43.5%) were the lowest.

Student factors associated with achievement

Multiple regression analysis based on the students' background questionnaires resulted in a model, which was able to explain 17.8% in Grade 10 and 17.3% in Grade 12 of the variations observed in the average scores at student level.

Supports given to pupils' learning have positive effect on their achievement. Among many types, supports given during their study time was a major one. In the subjects tested, mean achievement scores of students who were getting support were found to be higher than those who were not getting.

Schools using plasma system were identified and the mean achievement scores of pupils in the schools using plasma system and those not using the system were compared. No major difference was observed in the achievement scores between the two groups.

The mean scores of students with positive attitude toward the school were slightly better than the other groups. The correlation between positive attitudes toward school with achievement score was found to be statistically significant in both grades.

Time spent on homework was positively related with achievement scores in both grades. The achievement scores of the tested students were also related with the average number of days they were absent from school in that year. Absenteeism was found negatively correlated with the achievement scores in both grades.

Pupils who have the opportunity of getting meal three times in a day scored higher than those getting once or twice a day but the correlations were very week.

Students traveling more distance seem to achieve less than the other groups though the correlations were weak but statistically significant in both grades.

Contrary to our expectations, students from families with high education level did not achieve higher scores compared to families from lower education level.

Family economic status was another factor that was expected to have an influencing potential on pupils' achievement but the correlations were weak and in the negative direction.

Regression models

The regression models built based on students' characteristics and background variables were able to explain 17.8% and 17.3% of the variations in achievement scores observed in Grades 10 and 12 respectively. The models were rather weak fit but statically significant in both cases.

School level effects

The variance partitioning based on fully unconditional hierarchical linear model result, that took student level and school level data simultaneously, showed that in 10^{th} grade, 16.5% and in 12^{th} grade, 8.8% of the observed variations in academic achievement came from differences between schools.

The same model that took school level and region level data simultaneously showed that in both Grades 10 and 12, 5.2 % of the observed variations in academic achievement were due to differences between the regions.

Conclusions

The mean achievement scores in the subjects tested were found very low and most students in both grades were unable to score at least 50%.

Despite the fact that the minimum expected score in each subject is 50%, only 10.1% in Grade 10 and 16.5% in Grade 12 were able to score 50% and above in physics. Physics is a major requirement of further education in the field of science and technology which is the focus of the government at present. However, the achievement scores in this subject were found the least at both grades. It is to be recalled that similar results were found in earlier studies conducted at Grade 8.

In English, which is also a medium of instruction in secondary schools, most students scored far less than the mastery level. As the mean scores of the tested subjects showed strong correlations with each other one confounding factor might be language skill.

The standard deviation of each subject shows the existence of wide variations in the mean scores. Further investigation of the mean scores of each subject also showed the gap between high achievers and low achievers was very wide.

There exist wide variations across subgroups too. Boys are performing better than girls in both grades in each subject. Emerging regions are mostly performing less when compared with the others.

Looking at the standardized scaled scores only very few pupils were categorized in proficient and advanced levels. The positively skewed distributions of the subjects tested in both grades confirm the result of the performance standard.

A number of students' home background variables and characteristics showed statistically significant positive/negative relations but in most cases, the correlations were found statistically significant but weak. The final regression models based on the student level data were able to explain about 17% of the variations observed in the average scores.

Recommendations

1. The observed low achievement scores in all the subjects tested calls for immediate intervention and continued effort to raise the achievement levels. Schools and teachers should be facilitated and supported through professional development trainings in extending the use of all available resources and sharing experiences with one another.
2. A tailor made learning support program should be introduced at national and regional levels. The primary beneficiaries of this scheme should be disadvantaged areas such as the emerging regions. The Ministry of Education and/or the Regional Education Bureaus can track disparities through regular analysis of standards and can take preventive measures which can correct any disparities identified.
3. The existence of wide variations in achievement scores not only between groups but also within group calls for individualized approaches of teaching. Teachers

should be trained and become familiar to techniques that help to diagnose and intervene at individual student level.

4. Teachers need to incorporate greater differentiation of teaching practices into their classroom. Such differentiations should address the needs of both low and high achieving students.

5. All concerned bodies should address provision of additional supports to girls. There should be a concerted effort to raise the achievement levels of girls and narrow the gap. The source of the problem goes beyond the Education Sector hence there is a need to carry out detailed investigations to come up with specific recommendations.

6. Mastery of the medium of instruction is a key to read and understand other subjects. As long as English continues as instructional language due attention should be given to the subject. Beside with remedial actions that should be taken in secondary schools, early intervention that can raise language proficiencies should be introduced at lower grade levels.

7. In all the subjects in general, and in physics in particular, there is a felt need to revisit the curriculum materials and the mode of delivery and intervene immediately. Simplifying the materials and relating the concepts with day to day life are suggested.

8. In order to identify subject specific implications and maximize the benefit of the available data, further exploration using rigorous standard setting and item analysis procedures should be carried out. Production of separate report for each subject will help to maximize the benefits of the assessment.

This page is left blank intentionally.

1 Introduction

The challenge facing Ethiopia today is not only providing equitable access to its growing student population, but also to ensure that education imparted to these students is effective in supplying them with the necessary skills to fully contribute to the development of the society and the economy. Improving the quality of education requires a multi directional approach including the effectiveness of the learning process (by improving teaching techniques, supplying better learning materials and teaching, motivating students to attend school etc). Learning assessments, which are credible and objective, play important roles in this process by providing critical feedback on what and how well students are learning.

The Ministry of Education is committed to improve the quality of education along with its effort to increase enrollment and access. This commitment for improving the quality of education by including quality outcomes such as student achievement scores as performance monitoring indicators is reflected in the Education Sector Development Program III (ESDP III). It is also reaffirmed in the General Education Quality Improvement Program (GEQIP). Learning assessment is increasingly being used around the world to identify strengths and weaknesses of the education system. In a national or regional learning assessment, measures of achievement in key curriculum areas are administered to students of various selected grade levels. Performance of students in selected sub domains can indicate the strengths and weaknesses in students' learning within the curriculum areas and can show how the intended curricula are implemented in schools. Data on important student, teacher and school background factors affecting the learning outcomes are collected along with student achievement data to help identify areas needing improvement. This information could then help policy makers identify major areas of intervention

Some developing countries have tried to assess and measure student achievement and improve their educational systems. Nevertheless, most countries still apply public examinations for certification, selection and promotion. The goal of improving student learning has remained one of the most desired goals of educational processes.

In Ethiopia, quality assurance has been an important part of the reform process. To this end, three national learning assessments have been conducted since 1999 on Grades 4 and 8. This study is the first of its kind to be carried out on Grades 10 and 12 students. The main purposes of conducting this study were to provide information about learning attainments by students and the factors that determine those attainments in the nation so that attention is paid to the improvement of the system as a whole.

This report is organized in five chapters: Chapter 1 starts with introductory statements and provides the rationale. Chapter 2 presents a brief background of learning assessment. Chapter 3 deals with the frame of analysis and the methodology. Chapter 4 presents the findings starting with overall summary descriptive statistics followed by statistical tests to detect differences across subgroups. The relationships of different background variables with the achievement scores are also discussed. In Chapter 5, the findings are summarized, conclusions drawn and recommendations made.

1.1 Objectives of the Study

The Ethiopian Baseline National Learning Assessments of Grade 10 and 12 Students have the following objectives:

1. Analyze the national and regional level student learning achievement results in Grades 10 and 12, English, mathematics, biology, chemistry and physics;
2. Analyze Grades 10 and 12 students' achievement in English, mathematics, biology, chemistry and physics results across subgroups: gender, region and selected home background variables;
3. Identify correlates between achievement scores and different student and school level variables;
4. Explain the major factors that influence Grades 10 and 12 students' academic achievement; and
5. Discuss and summarize the implications of the findings for improvement of school quality and effectiveness.

1.2 Significance of the Study

Student learning assessment involves a systematic process of collecting relevant, valid and timely information about the outcomes of schooling so that decisions are made about the learning and development of students, curriculum, educational programs and educational policy. Student learning assessment provides the necessary feedback and objective evidence required to maximize the outcomes of educational efforts. Such assessments summarize what learners know, understand, and can do in relation to some or all of the learning goals determined in the curricula.

Over the last decade, substantial attempts have been made to expand education, improve access, equity and efficiency in Ethiopia. Now the emphasis has also shifted towards improving quality in all areas and in particular towards student learning achievement. This learning assessment, therefore, provides an indication or feedback of where students' achievement stands in relation to the stated profiles of the curriculum.

A student learning assessment can provide baseline information from which progress can be measured during and at the end of a key stage in education. Since it focuses on actual learning, it enables one to find out the extent to which an educational system is effective as a whole. If it is properly integrated into the system of education, student learning assessment can help actors and stakeholders to focus their collective attention, examine their assumptions, and create a shared academic culture dedicated to assuring and improving the quality of education.

Information on the relationship between student learning outcomes and school inputs provides an immense potential to policy makers to identify, allocate and manage the resources of education to improve quality. This study may provide pieces of information alongside the achievement results so that the most influential determinants of learning are properly known and managed.

Ethiopia expends a considerable amount of its public finance on education. In order for the education sector to justify this expenditure and retain support, both the government and the public require that the money expended produce the required skills.

2 Background

2.1 National Learning Assessments

Concern for educational change and improving its quality has been the focus of educational planners for years. However, the World Declaration on Education for All (EFA) in Jomtien, Thailand (1990) is considered to have uncovered much of the dire necessity of learning assessment. The Education for All declaration gave not only fresh impetus to issues related to assessment but also made clear that there has to be a new form of assessment: system assessment, or national assessment, in order to determine whether children were acquiring the essential knowledge, reasoning ability, skills, and values that schools have promised to deliver. In other words, the basis for learning assessment is a response to both the desirable learning behaviour to take place and ensuring schools' accountability to their stakeholders (the state, the parents, etc) (Kellaghan and Greaney, 2004).

Kellaghan and Greaney (2001) also revealed that one of the most influential statements of concern for learning outcomes is contained in the declaration adopted by the World Conference on Education for All. It emphasizes that the provision of basic education for all was meaningful only if children could acquire useful behavioural skills and values. To this end, Article 4 of the World Declaration on Education for All (1990) stated that focus of basic education should be "on actual learning acquisition and outcome, rather than exclusively upon enrolment, continued participation in organized programmes and completion of certification requirements". Similarly, after a 10-year follow-up to Jomtien declaration, the Dakar Conference (2000) stressed the importance of having "a clear definition and accurate assessment of learning outcomes (knowledge, skills, attitudes, and values)" as governments need to ensure basic education of quality for all, for their citizens (UNESCO, 2000).

The focus on learning has been progressively shifting from input to outcomes in view of learning achievement. Past educational reforms mainly used to emphasize educational structure, curriculum and teacher training, in a view to realize quality. But this trend began to give way to issues related to the improvement of learning achievement, school effectiveness, management and accountability. Consequently,

decentralization, school-based management and learning assessment became the area of focus in the efforts related to educational reforms of the 1990s. In the view of Kellaghan and Greaney (2001), global economic competition has resulted in the critical importance of quality human resources, and the demand for new competencies in the modern information society. All of these demands have therefore, made the educational system, schools, and individual students to be under increasing pressure to perform and work hard. In short, assessing students' learning achievements has instigated due attention and a necessary focus to be made for the former.

The emergence of learning assessment is believed to come up with an objective appraisal system of a given education system before arriving at sound judgment. It is also important to note that one of the modern assessment procedures focus on outcomes. Kellaghan and Greaney argue that unlike past assessments which focus on inputs (e.g. physical facilities, curriculum materials, books, and teacher training) to determine the quality of education, this is no longer the case. Today, the dominant question posed by many stakeholders, including policymakers, has become on the outcomes of education: whether students are acquiring the desired knowledge, skills, behaviour, and attitudes. As a result, policymakers or educational managers need information that would be necessary to reach informed judgment as related to the adequacy of student achievements obtained in the system. They may also need a baseline data on student achievement against which to measure progress or excellence being registered in the educational program. In the mean time, teachers may need similar information on the achievement of their students in order to make some form of comparisons and assess their own professional effectiveness.

Test scores provide one important measure of how well the curriculum is being learned, and help to indicate achievement at the main exit points of the school system. Teachers, schools and national governments have long gathered information on pupils' performance. Teachers and schools use assessment data to monitor pupil progress, identify pupils with difficulties (and suggest appropriate responses), and even to motivate pupil learning. Although these types of assessment are quite varied in form and function, they are used primarily to provide information on individual performance. National governments may also assess educational outcomes (what has

been learned) through examinations. Learning assessments allow the objective measurement of performance at system-level. Although there are many benefits to such type of data, some are of particular relevance. Greaney and Kellaghan (1996) identified eight main uses for such data: *informing policy*, *monitoring standards*, *identifying correlates of achievement*, *introducing realistic standards*, *promoting accountability*, *increasing public awareness*, *directing teachers' efforts and raising pupil achievement*, and *informing political debate*.

2.2 National Learning Assessments in Ethiopia

To date, Ethiopia has carried out three national learning assessments on the primary education. The first learning assessment on students' achievements in Ethiopia was carried out in 2000 at the end of the first cycle (Grade 4) and second cycle (Grade 8) of primary school on a sample basis in four subjects. This assessment serves as a baseline since there was no similar survey that preceded it. The Second National Learning Assessment (ESNLA) was carried out after four years, in April 2004. The Third Ethiopian National Learning Assessment (ETNLA) was then undertaken after three years, in 2007 (GEQAEA, 2008).

In the Ethiopian Baseline National Learning Assessment (EBNLA), the mean scores for Grade 4 students in three subjects were below 50 percent while the mean for all subjects was 47.9%, which is less than the minimum requirement of 50% to pass from one grade to the next per the Ethiopian Education and Training Policy (NOE, 2000).

In the Ethiopian Second National Learning Assessment (ESNLA) of Grade 4, the mean scores for Reading and Mathematics have shown improvements though the changes were not significant. The pupils' performance improved significantly only in Environmental Science. However, the mean for all subjects was 48.5% percent, which is below the minimum requirement. In fact, the pupils' mean score in English in the ESNLA dropped to 38.7% from 40.5% in the EBNLA (NOE, 2006).

In the Ethiopian Third National Learning Assessment (ETNLA) of Grade 4, the mean scores for English and Environmental Science, and the mean for all subjects, were lower than the corresponding figures both in the first and second learning assessments (GEQAEA, 2008). Although the pupils' performance has slightly improved in

mathematics, the decline in the mean scores for English and environmental science has pulled down the mean for all subjects to 39.8%, which is far below the minimum requirement 50% of the Ethiopian Education and Training policy.

2.3 Operational Definitions

Blueprint: A test blueprint, or test specification, details how an exam is to be constructed. It includes important information, such as the total number of items, the number of items in each content area or domain, the number of items that are recall vs. reasoning, and the item formats to be utilized.

Classical Test Theory (CTT): A psychometric analysis and test development paradigm based on correlations, proportions, and other statistics that are relatively simple compared to IRT. It is therefore more appropriate for smaller samples, especially for fewer than 100.

Item Response Theory (IRT): A comprehensive approach to psychometric analysis and test development that utilizes complex mathematical models. This provides several benefits but requires larger sample sizes. A common rule of thumb is 100 candidates for the one-parameter model and 500 for the three-parameter model.

Reliability: A measure of the repeatability or consistency of the measurement process. Often, this is indexed by a single number, most commonly the internal consistency index coefficient alpha or its dichotomous formulation, KR-20. Under most conditions, these range from 0.0 to 1.0, with 1.0 being perfectly reliable measurement. However, just because a test is reliable does not mean that it is valid, i.e., measures what it is supposed to measure.

Scaling: Scaling is a process of converting scores obtained on an exam to an arbitrary scale. This is done so that all the forms and exams used by a testing organization are on a common scale. For example, suppose an organization had two testing programs, one with 50 items and one with 150 items. All scores could be put on the same scale to standardize score reporting.

Validity: Validity is the concept that test scores can be interpreted as intended. For example, a test for certification in a profession should reflect basic knowledge of that profession, and not intelligence or other constructs, and scores can therefore be interpreted as evidencing professional competence. Validity must be formally established and maintained by empirical studies as well as sound psychometric and test development practices.

3 Methodology

3.1 Sampling

The target populations of this study were Grades 10 and 12 students at the end of the 2008/2009 academic year (2001 E.C.). A two-stage cluster sampling procedure was followed. At the first stage, a fixed number of schools proportion to the number of schools in regions. In the second stage cluster of 40 students were taken from the sample schools. School directors and teachers of the sample students were also asked to fill questionnaires.

Table 1 below shows the distribution of the sample schools and students of Grade 10 across regions. Looking at planned and achieved sample sizes the response rate at the first and the second stages were 99.3% and 99% respectively.

Table 1. Distribution of samples G10

Region	Schools		Students	
	Planned	Achieved	Planned	Achieved
Tigray	10	10	400	400
Afar	3	3	120	120
Amhara	33	33	1320	1319
Oromiya	47	46	1880	1828
Somali	4	4	160	157
B. Gumuz	3	3	120	120
SNNP	23	23	920	919
Gambella	3	3	120	120
Harari	3	3	120	120
Addis Ababa	9	9	360	360
Dire Dawa	3	3	120	120
Total	**141**	**140**	**5640**	**5583**

Table 2 below shows the distribution of the sample schools and students of Grade 12 across regions. Looking at the planned and achieved sampled levels the response rate at the first and the second stages were 98.6% and 95% respectively.

Table 2. Distribution of samples G12

Region	Schools		Students	
	Planned	**Achieved**	**Planned**	**Achieved**
Tigray	6	6	240	240
Afar	2	2	80	67
Amhara	11	11	440	432
Oromiya	27	27	1080	989
Somali	2	2	80	71
B. Gumuz	2	2	80	79
SNNP	13	13	520	513
Gambella	2	1	80	61
Harari	2	2	80	80
Addis Ababa	5	5	200	200
Dire Dawa	2	2	80	80
Total	**74**	**73**	**2960**	**2812**

3.2 Instrumentation

To measure pupils' learning achievement and to obtain information on factors that have been found affecting the quality of student learning, two kinds of instruments were used for the assessment. These were achievement tests in English, mathematics, biology, chemistry and physics, and background questionnaires and perceptioners for students, teachers, and directors.

The achievement tests were developed based on the National Curriculum using the Minimum Level of Competency (MLC). Competencies based on the first three levels of the cognitive domain of Bloom's Taxonomy (Knowledge, Understanding and Application) were developed for each subject keeping in view the content areas and objectives of the curriculum. Then a test blue print was designed for each test which included content, learning outcomes and number of items. These instruments were pilot tested in April 2008 in selected schools. Item and Test analyses were carried out following Classical Test Theory (CTT) and Item Response Theory (IRT) procedures and based on the results of the analyses items were further improved and final versions produced. The questionnaires contained questions in order to identify the association of various personal, home, school, and teaching-learning variables with student achievement.

3.3 Data Collection and Organization

Professionals from GEQAEA served as route coordinators. Data collectors were recruited from the regions. They attended a training program on how to administer the tests and conduct interviews. In each school, a data collector stayed for three days to administer the tests and conduct interview.

Data from tests and students' questionnaires were captured using OMR. Data from teachers' and directors' questionnaires were captured using MS Access. For the purpose of data cleaning and checking consistency, MS Excel 2003 and SPSS v15were used.

3.4 Data Analysis

Descriptive summary statistics to summarize central tendencies and dispersion were computed to each subject and to the average score. Correlation and statistical tests of significance were also computed to detect relationships and differences. One-way analysis of variance followed by Post Hoc test was computed to identify homogenous subset groups. Variance component partitioning based on hierarchical linear modeling was computed to see the effects of the schools. Specialized Item Analysis packages, SPSS v15, Stata v10 and Systat v12 were used.

4 Results

4.1 Achievement Outcomes

This part deals with the performance of students on the achievement tests. The subjects were English, mathematics, biology, chemistry and physics. In addition the average score of the five subjects is also reported. Each test was composed of multiple choice items from Grades 9 and 10 contents based on the Minimum Learning Competency (MLC). The raw scores of each subject were converted into percentages. Initially analyses were based on the percentage scores and scaled scores are also reported when appropriate.

4.1.1 Summary Descriptive Statistics

In Grade 10, the summary descriptive statistics shows that the mean score for each subject and the average score of the five subjects were all below the minimum expected score. The minimum passing mark set by the Education and Training Policy is (50%). The median score which is less than the mean score (36.0%) shows that 50% of the students in the average score obtained less than 31.9% (Table 3). The mean score for physics (31.2%) was the least. The distribution in all subjects was positively skewed indicating that only very few pupils achieved the highest scores.

Table 3. Mean scores by subject (%) G10

Subject	N	Mean	SD	Median	Skewness
English	5498	37.4	13.62	33.3	1.06
Math	5525	34.7	14.18	31.7	1.23
Biology	5502	40.3	16.52	35.6	0.97
Chemistry	5476	36.1	14.81	31.3	1.21
Physics	5476	31.2	14.81	31.3	1.45
Average	**5411**	**36.0**	**12.7**	**31.9**	**1.45**

Figure 1 on the next page shows the percentage of puipls with scores of 50% and above in each subject. Accordingly only 10.1% in physics and 24.8% in biology were able to score 50% and above. In the average score 13.8% scored 50% and above.

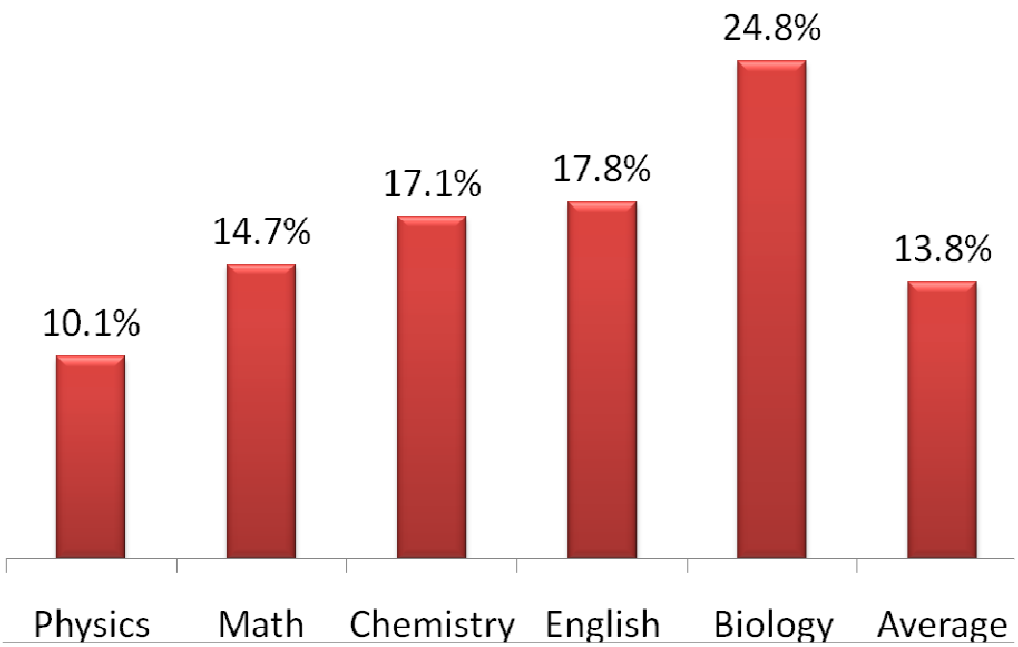


Figure 1. Percent achieving 50% and above by subject G10

In Grade 12, the summary descriptive statistics shows that the mean scores for English, chemistry and physics and the average score of the five subjects were below the minimum expected score. The mean scores of biology and mathematics were found greater than the minimum passing mark set by the Education and Training Policy is (50%). The median score which is less than the mean score shows that 50% of the students in the average score obtained less than 46.1% (Table 4). The mean score for physics (33.3%) was the least. The distribution in all subjects was positively skewed indicating that only very few pupils achieved the highest scores.

Table 4. Mean scores by subject (%) G12

Subject	N	Mean	SD	Median	Skewness
English	2772	43.4	11.39	42.2	.439
Math	2660	54.3	16.43	53.3	.113
Biology	2761	55.5	17.61	55.5	.096
Chemistry	2734	49.1	15.68	49.1	.322
Physics	2727	36.6	14.22	33.3	.917
Average	**2579**	**47.8**	**12.55**	**46.1**	**.520**

Figure 2 below shows the percentage of puipls with scores of 50% and above in each subject. Accordingly only 16.7% in physics and 60.7% in biology were able to score 50% and above. In the average score 34.9% scored 50% and above.

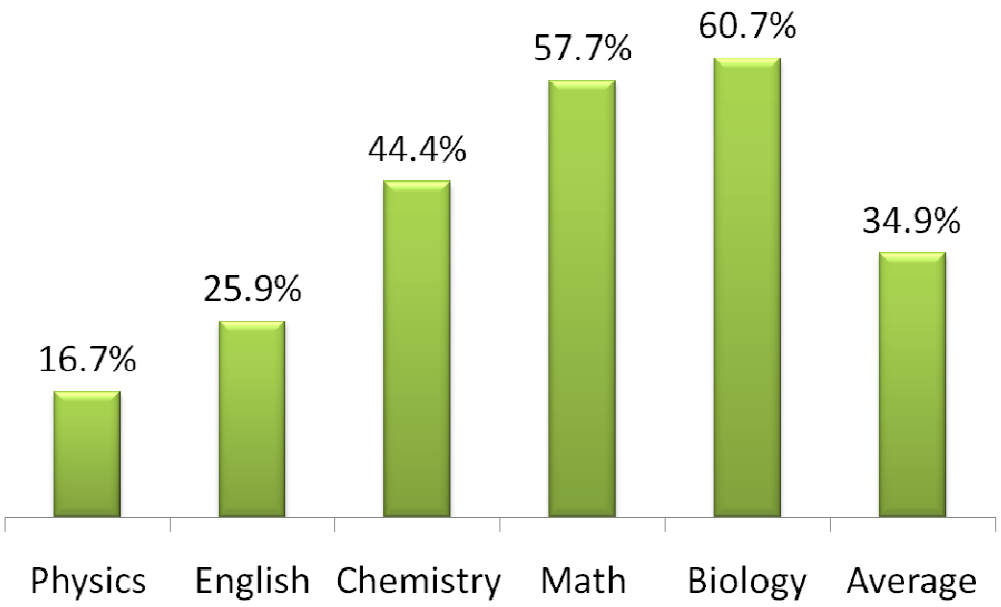


Figure 2. Percent achieving 50% and above by subject G12

4.1.2 Correlations between the mean scores of the subjects

In Grade 10 there exist positive relationship between the five subjects and the correlations were statistically significant in all cases at $p < .01$ (Table 5). This shows that students performing well in one subject did the same in the others. Looking at the relation between the instructional language and other subjects it was the highest with biology (r=.767). Subjects such as biology when compared with the others are more influenced by language ability. The correlation between biology and the average score was found the highest (r=.911).

Table 5. Pearson product moment correlation between the five subjects G10

Subject	Mathematics	Biology	Chemistry	Physics	Average
English	.670**	.767**	.689**	.585**	.854**
Mathematics		.706**	.707**	.630**	.853**
Biology			.789**	.630**	.911**
Chemistry				.705**	.899**
Physics					.819**

** Correlation is significant at the 0.01 level (2-tailed).

In Grade 12 there exist positive relationship between the five subjects and the correlations were statistically significant in all cases at $p < .01$ (Table 6). This shows that students performing well in one subject did the same in the others. Looking at the relation between the instructional language and other subjects it was the highest with biology (r=.607). Subjects such as biology when compared with the others are more influenced by language ability just the same way like Grade 10. The correlation between chemistry and the average score was found the highest (r=.885).

Table 6. Pearson product moment correlation between the five subjects G12

Subject	Mathematics	Biology	Chemistry	Physics	Average
English	.472**	.607**	.530**	.440**	.711**
Mathematics		.630**	.668**	.647**	.838**
Biology			.757**	.579**	.879**
Chemistry				.663**	.885**
Physics					.806**

** Correlation is significant at the 0.01 level (2-tailed).

4.1.3 Relation between the school based and the national assessment scores

The achievement scores of the tested subjects and the first semester score of the academic year obtained from the school rosters were also compared to see relationship. The scores in Grade 10 were correlated positively in all cases and the relationships were statistically significant at $p < .01$ (Table 7). The test development and mode of administration of the school based tests obviously differ from school to school; hence direct comparisons are not possible. Nevertheless the existing relationship shows that those who did well at their school also did the same in the national assessment tests.

Table 7. Pearson product moment correlations between school based and national assessment scores (%) G10

National Assessment (%)	School Based Scores (%)				
	English	Mathematics	Biology	Chemistry	Physics
English	.490**				
Mathematics		.518**			
Biology			.544**		
Chemistry				.539**	
Physics					.421**

** Correlation is significant at the 0.01 level (2-tailed).

In Grade 12 too, the scores were correlated positively in all cases and the relationships were statistically significant at $p < .01$ (Table 8).

Table 8. Pearson product moment correlations between school based and national assessment scores (%) G12

National Assessment (%)	School Based Scores (%)				
	English	Mathematics	Biology	Chemistry	Physics
English	.458(**)				
Mathematics		.324(**)			
Biology			.339(**)		
Chemistry				.478(**)	
Physics					.427(**)

4.1.4 Range of the achievement scores

Table 9 illustrates the range of achievement in the five subjects and the average scores. The table shows scores achieved at key benchmarks: 10^{th}, 25^{th}, 50^{th}, 75^{th} and 90^{th} percentiles. Performance at the 10^{th} percentile may be taken as indicative of the standard among low achievers in a country, while performance at the 90^{th} percentile can be taken as indicative of high achievers. Pupils at the 90^{th} percentile only achieved scores of 54.9% in the average. This means only 10% of the candidates were able to achieve a score of 54.9% and above. On the other hand pupil at 10^{th} percentile scored only 24.4% and this means 10% of the examinee scored at or below chance level in all subjects. Differences between the 10^{th} and 90^{th} percentiles (30.5%) in the average score is an indication of how wide spread the variation between high-achieving and low-achieving students. On the other hand 50% of the candidates scored between 28.9% and 35.6% in the average.

Table 9. Range of achievement scores (%) at five key marker points G10

Percentiles	English	Mathematics	Biology	Chemistry	Physics	Average
10th	22.7	20.0	23.3	21.3	17.8	24.4
25th	27.3	25.0	27.8	25.0	22.2	27.1
50th	33.3	31.7	35.6	31.3	28.9	31.9
75th	43.9	41.7	50.0	42.5	35.6	40.8
90th	57.6	55.0	65.6	58.8	51.1	54.9

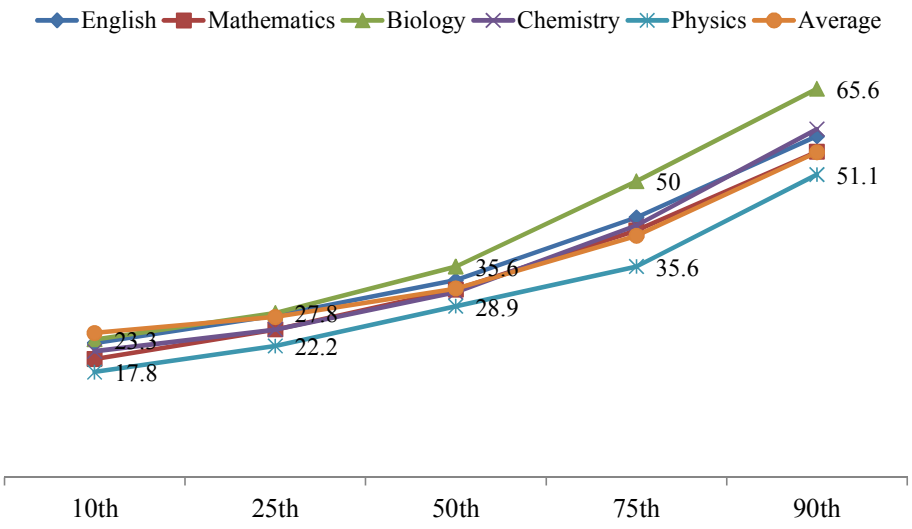


Figure 3. Range of achievement scores at five key marker points G10

In Grade 12 pupils at the 90th percentile only achieved scores of 65.2% in the average. This means only 10% of the candidates were able to achieve a score of 65.2% and above. On the other hand pupil at 10th percentile scored only 32.4% and this means 10% of the examinee scored at or below chance level in all subjects. Differences between the 10th and 90th percentiles (32.8%) in the average score is an indication of how wide spread the variation between high-achieving and low-achieving students. On the other hand 50% of the candidates scored between 38.9% and 55.8% in the average (Table 10).

Table 10. Range of achievement scores (%) at five key marker points G12

Percentiles	English	Mathematics	Biology	Chemistry	Physics	Average
10th	28.9	31.7	32.0	30.0	20.0	32.4
25th	34.9	41.7	42.0	37.5	26.7	38.9
50th	42.2	53.3	56.0	47.5	33.3	46.1
75th	50.6	66.7	68.0	60.0	44.4	55.8
90th	59.0	76.7	80.0	71.3	57.8	65.2

Table 11 presents the range of achievement scores based on the scaled scores; for all subjects the mean score was set at 250 and the standard deviation at 50 arbitrarily. This is especially important to compare the score of one subject with the other plus to make comparisons with similar studies in the future. Pupils who scored at the 10th percentile achieved a score of 144.3 in the average score, which is 2.11 standard deviations less than the mean. The corresponding score for pupils at the 90th percentile was 315.0 that is 1.3 standard deviations greater than the mean. The scaled score at the 50th percentile (median) is slightly less than the scaled mean score indicating the score is positively skewed.

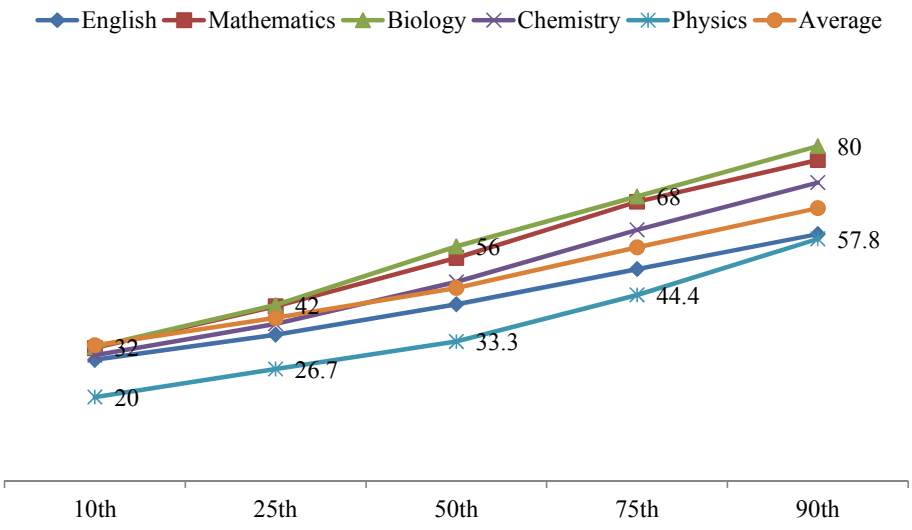


Figure 4. Range of achievement scores at five key marker points G10

Table 11. Range of achievement scores (%) at five key marker points based on scaled score G10

Percentiles	English	Mathematics	Biology	Chemistry	Physics	Average
10th	196.2	198.3	198.5	200.0	201.0	204.1
25th	212.9	215.9	212.0	212.7	217.2	214.7
50th	235.2	239.4	235.5	233.8	241.5	233.6
75th	274.1	274.7	279.3	271.8	265.8	268.9
90th	324.1	321.7	326.4	326.6	322.4	324.5

Grade 12 pupils who scored at the 10th percentile achieved a score of 185.9 in the average score, which is 1.28 standard deviations less than the mean. The corresponding score for pupils at the 90th percentile was 319.2 that is 1.38 standard deviations greater than the mean. The scaled score at the 50th percentile (median) is slightly less than the scaled mean score indicating the score is positively skewed.

Table 12. Range of achievement scores (%) at five key marker points based on scaled score G12

Percentiles	English	Mathematics	Biology	Chemistry	Physics	Average
10th	185.9	181.2	183.3	189.2	191.7	185.9
25th	212.6	211.7	211.7	213.2	215.2	212.6
50th	244.6	247.2	251.4	245.1	238.6	244.6
75th	281.9	287.8	285.5	284.9	277.6	281.9
90th	319.2	318.2	319.6	320.8	324.5	319.2

4.1.5 Performance at Varying Levels of Standards

In this part, the continuum of student's achievement is divided into four levels as '*Below Basic', 'Basic', 'Proficient'* and '*Advanced'* and the proportion achieving at each level is presented below in Table 13 and Figure 3. The classification is based on the scaled scores where '*Advanced'* is greater than 3 standard deviation from the mean *'Proficient'* is between 2 and 3 standard deviations above the mean, '*Basic'* is within 1 standard deviation above the mean and '*Below Basic'* is less than the mean score. Table 13 shows that based on the average scaled score only 6.1% were categorized as *Advanced*, 9.4% *as Proficient,* 20.8% as *Basic* and 63.7% as *Below*

Basic. Looking at the five subjects from 5.3% to 6.2% were in *Advanced* level while from 7.9% to 11.3% were *Proficient* level. On the other hand from 59.6% to 65.4 were in *Below Basic* level.

Table 13. Proficiency levels based on scaled scores G10

Subject	Below Basic	Basic	Proficient	Advanced
English	59.6%	24.2%	10.7%	5.6%
Mathematics	60.7%	24.4%	8.7%	6.2%
Biology	60.7%	22.8%	11.3%	5.3%
Chemistry	62.2%	21.5%	10.1%	6.2%
Physics	65.4%	21.2%	7.9%	5.5%
Average	**63.7%**	**20.8%**	**9.4%**	**6.1%**

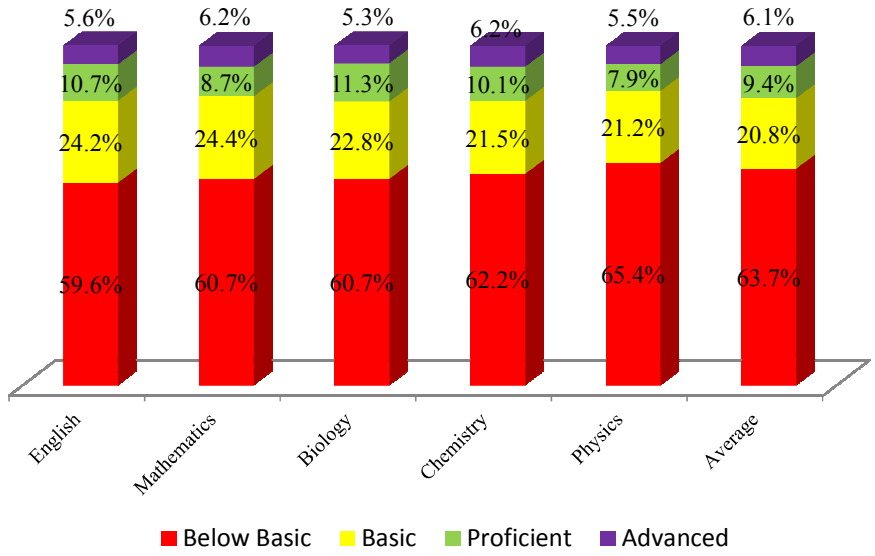


Figure 5. Proficiency levels based on scaled scores by subject G10

In Grade 12, based on the average scaled score, only 3.8% were categorized as *Advanced*, 12.7% *as Proficient,* 28.3% as *Basic* and 58.7% as *Below Basic*. Looking at the five subjects from 1.4% to 4.7% were in *Advanced* level while from 12.2% to 17.0% were *Proficient* level. On the other hand from 50.4% to 58.7% were in *Below Basic* level (Table 14 and Figure 4).

Table 14. Proficiency levels based on scaled scores G12

Subject	Below Basic	Basic	Proficient	Advanced
English	55.8%	28.3%	12.2%	3.7%
Mathematics	50.4%	31.6%	15.9%	2.1%
Biology	49.8%	31.8%	17.0%	1.4%
Chemistry	54.3%	26.9%	16.1%	2.7%
Physics	58.7%	24.3%	12.3%	4.7%
Average	**55.3%**	**28.3%**	**12.7%**	**3.8%**

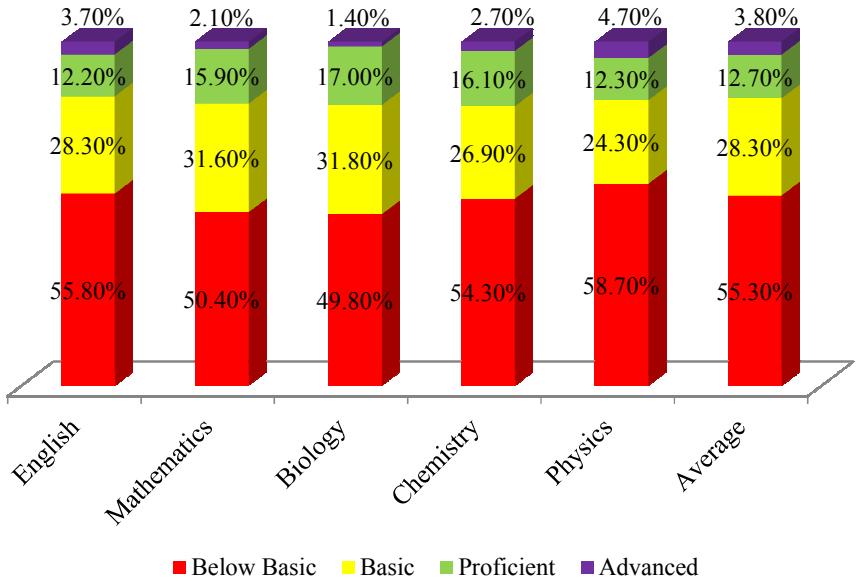


Figure 6. Proficiency levels based on scaled scores by subject G12

4.1.6 Gender and Achievement

Boys achieved mean scores that were higher by 6.53 % in the average score and 4.35% to 8.92% in the five subjects (Table 15) than girls. The differences in all subjects were statistically significant at $p < .001$.

Table 15. Independent sample t-tests between boys and girls G10

Subject	Gender	N	Mean	SD	SE	df	t	MD	Sig.
English	Male	3192	39.9	14.36	.25	5496	16.41	5.97	.000
	Female	2306	33.9	11.70	.24				
Mathematics	Male	3215	37.4	15.44	.27	5523	17.09	6.44	.000
	Female	2310	30.9	11.17	.23				
Biology	Male	3194	44.1	17.54	.31	5500	20.51	8.92	.000
	Female	2308	35.2	13.34	.27				
Chemistry	Male	3182	39.0	16.06	.28	5474	17.87	7.05	.000
	Female	2294	32.0	11.70	.24				
Physics	Male	3181	33.0	15.01	.26	5474	11.71	4.35	.000
	Female	2295	28.7	11.22	.23				
Average	**Male**	**3151**	**38.7**	**13.86**	**.24**	**5409**	**19.33**	**6.53**	**.000**
	Female	**2260**	**32.2**	**9.60**	**.20**				

In addition, looking at the distribution of the scores the mean sores for girls were highly skewed to the right than that of boys indicting fewer girls than boys were at the highest end of the scores (Figure 5).

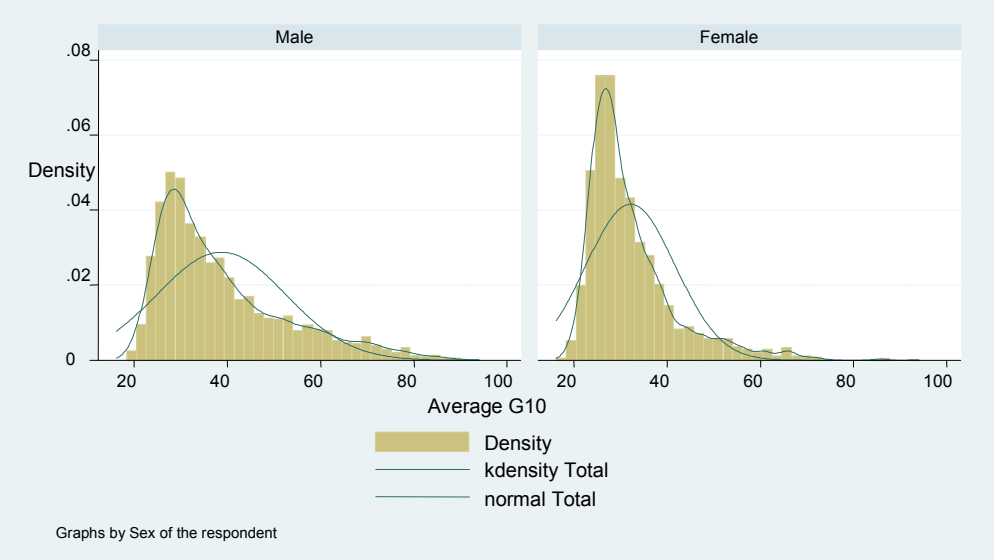


Figure 7. Distribution of the average score by sex G10

In Grade 12 boys achieved mean scores that were higher by 6.6% in the average score and 2.4% to 8.6% in the five subjects (Table 16) than girls. The differences in all subjects were statistically significant at $p < .001$. This shows that the gender gap in academic achievement is persistent as it has been the same with similar studies conducted in Grades 4 and 8.

Table 16. Independent sample t-tests between boys and girls G12

Subject	Gender	N	Mean	SD	SE	df	t	MD	Sig.
English	Male	2174	43.9	11.11	.24	2767	4.56	2.4	.000
	Female	595	41.5	11.80	.48				
Mathematics	Male	2082	56.1	16.37	.36	2657	11.31	8.5	.000
	Female	577	47.6	14.75	.61				
Biology	Male	2169	57.1	17.36	.37	2758	9.61	7.7	.000
	Female	591	49.4	17.14	.71				
Chemistry	Male	2147	50.9	15.49	.33	2730	12.09	8.6	.000
	Female	585	42.3	14.40	.60				
Physics	Male	2142	37.8	14.70	.32	2723	8.90	5.8	.000
	Female	583	32.0	11.06	.46				
Average	**Male**	**2017**	**49.2**	**12.60**	**.28**	**2576**	**11.28**	**6.6**	**.000**
	Female	**561**	**42.6**	**10.82**	**.46**				

Furthermore, looking at the distribution of the scores the mean sores for girls were highly skewed to the right than that of boys indicting fewer girls than boys were at the highest end of the scores (Figure 6).

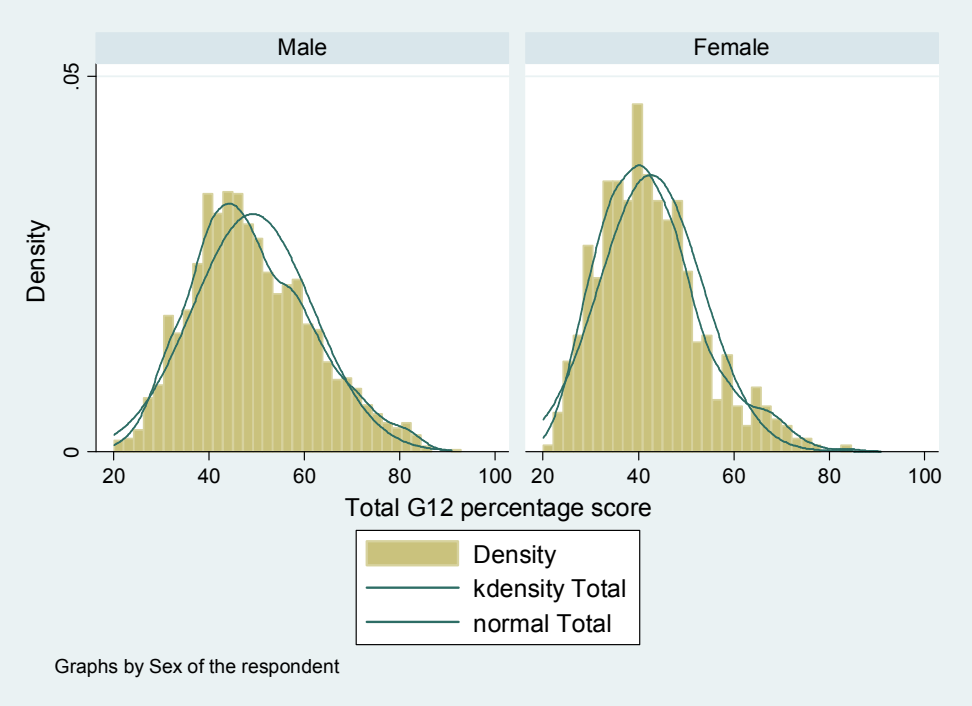


Figure 8. Distribution of the average score by sex G12

4.1.7 Comparisons by Region

The analysis of variance taking the region as independent and the average score as the dependent variable was carried out to identify the existence of statistically significant mean differences. In Grade 10 a statistically significant difference was observed on the average score ($F_{(10,\ 5400)} = 12.9$, $p < .001$). Table 17 shows that SNNP (38.1%) scored highest with a mean difference of 9.6 % from Gambella which achieved the least score (28.5%). The mean scores of Addis Ababa, Amhara, B. Gumuz and Oromia were slightly higher than the national mean. All the other regions achieved mean scores equal or less than the national mean (36%).

Table 17. Average score by region G10

Region	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean	
					Lower Bound	Upper Bound
SNNP	906	38.1	12.81	0.43	37.26	38.93
Addis Ababa	327	36.8	12.48	0.69	35.46	38.18
Amhara	1295	36.6	13.56	0.38	35.87	37.35
B. Gumuz	119	36.4	12.09	1.11	34.20	38.59
Oromiya	1780	36.2	12.76	0.30	35.58	36.76
Dire Dawa	103	34.8	11.81	1.16	32.51	37.12
Tigray	387	33.8	11.24	0.57	32.71	34.96
Harari	115	33.8	11.86	1.11	31.64	36.02
Somali	152	32.8	9.63	0.78	31.25	34.34
Afar	111	29.4	6.68	0.63	28.11	30.63
Gambella	116	28.5	7.80	0.72	27.02	29.89
Total	**5411**	**36.0**	**12.68**	**0.17**	**35.66**	**36.34**

Figure 5 shows a recursive partition analysis procedure which resulted in different groups of the regions based on the average score. Initially there are two groups with group mean below and above the national mean. Further partition shows that additional two groups where SNNP with a mean score of 38.1% distinctly differs from the others.

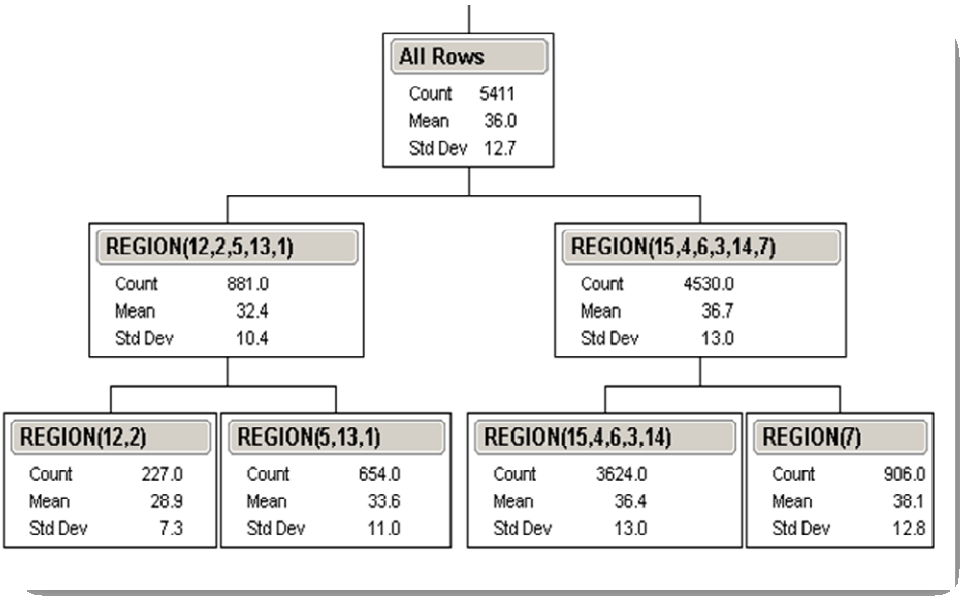


Figure 9. Recursive partitioning analysis by region G10

In Grade 12 the analysis of variance showed a statistically significant difference between regions in the average score ($F_{(10, 2567)} = 7.38$, $p < .001$). Table 18 shows that Dire Dawa (53.3%) scored highest with a mean difference of 13.0 % from Somali which achieved the least score (40.3%). The mean scores of Dire Dawa, Harari and Oromia were higher than the national mean. All the other regions achieved mean scores equal or less than the national mean (47.8%).

Table 18. Average score by region G12

Region	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean	
					Lower Bound	Upper Bound
Dire Dawa	77	53.3	13.64	1.55	50.2	56.4
Harari	78	49.4	12.38	1.40	46.6	52.2
Oromiya	965	49.0	12.17	.39	48.2	49.7
B. Gumuz	479	47.8	11.42	.52	46.8	48.8
Amhara	403	47.6	13.11	.65	46.4	48.9
Afar	63	47.1	11.60	1.46	44.2	50.0
Addis Ababa	190	46.5	13.13	.95	44.6	48.3
Tigray	118	45.3	14.09	1.30	42.8	47.9
SNNP	77	43.5	12.27	1.40	40.7	46.3
Gambella	58	43.5	9.63	1.26	41.0	46.0
Somali	70	40.3	13.43	1.60	37.1	43.5
Total	**2578**	**47.8**	**12.53**	**.25**	**47.3**	**48.2**

The recursive partition analysis procedure based on the average score of the regions resulted in different groups. Figure 6 shows that initially there are two groups with group mean below and above the national mean. Further partition resulted in additional two groups where Dire Dawa with a mean score of 53.3% distinctly differs from the other.

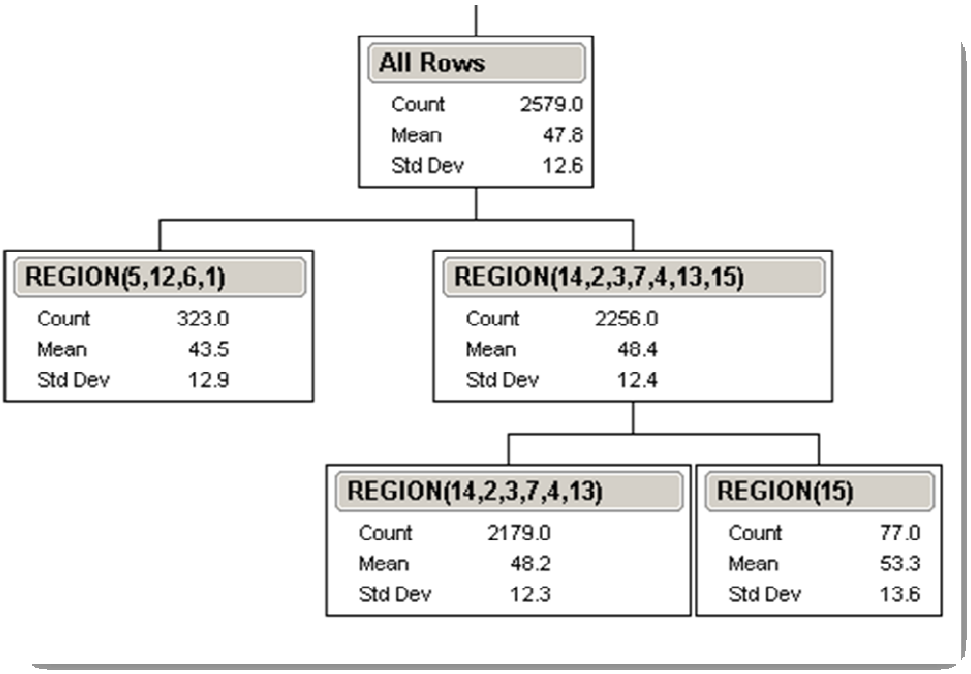


Figure 10. Recursive partitioning analysis by region G12

For the regional comparisons on achievement scores a 0.005 alpha level was used for making decisions about mean differences and for identifying homogeneous groupings based on post-hoc comparisons. The reduced alpha level (i.e., 0.005 as opposed to 0.05) compensates for the inflation of Type I error that can be expected when conducting multiple statistical tests. Scheffe paired comparisons was applied to identify homogenous subset groups for the average score.

In Grade 10 the homogenous subset grouping (Table 19) shows that there are two groups. Gambela and Afar (in Group 1) distinctly differ from those regions (in Group 2) whose mean scores are greater than the national mean.

Table 19. Homogenous subset groupings of average score by region G10

Region	N	Subset for alpha = 0.005	
		1	2
Gambella	116	28.5	
Afar	111	29.4	
Somali	152	32.8	32.8
Harari	115	33.8	33.8
Tigray	387	33.8	33.8
Dire Dawa	103	34.8	34.8
Oromiya	1780		36.2
B. Gumuz	119		36.4
Amhara	1295		36.6
Addis Ababa	327		36.8
SNNP	906		38.1
Sig.		**.008**	**.084**

In Grade 12 the homogenous subset grouping based on the average score (Table 20) shows that there are three groups. Somali (in Group 1) distinctly differ from Oromia, Harari and Dire Dawa and the differences are statistically significant.

Table 20. Homogenous subset groupings of average score by region G12

Region	N	Subset for alpha = .005		
		1	2	3
Somali	70	40.3		
Gambella	58	43.5	43.5	
B. Gumuz	77	43.5	43.5	
Tigray	118	45.3	45.3	45.3
Addis Ababa	190	46.5	46.5	46.5
Afar	63	47.1	47.1	47.1
Amhara	403	47.6	47.6	47.6
SNNP	479	47.8	47.8	47.8
Oromiya	965		49.0	49.0
Harari	78		49.4	49.4
Dire Dawa	77			53.3
Sig.		**.041**	**.297**	**.018**

English

In Grade 10 the analysis of variance showed a statistically significant difference in English score ($F_{(10, 5487)} = 9.16$, $p < .001$). Table 21 shows that Addis Ababa (40.3%) and B. Gumuz (39.7%) scored highest. Gambella (31.5%) and Afar (32.2%) achieved lowest scores and the mean differences were statistically significant with the highest achieving ones.

Table 21. Homogenous subset groupings of English score by region

Region	N	Subset for alpha = 0.005	
		1	2
Gambella	118	31.5	
Afar	115	32.2	
Tigray	391	34.2	34.2
Harari	120	35.7	35.7
Dire Dawa	105	37.0	37.0
Oromiya	1807	37.2	37.2
Amhara	1307	38.0	38.0
SNNP	910	38.3	38.3
Somali	156	38.5	38.5
B. Gumuz	119		39.7
Addis Ababa	350		40.3
Sig.		**0.01**	**0.04**

In Grade 12 the analysis of variance showed a statistically significant difference in English score ($F_{(10, 2791)} = 17.5$, $p < .001$). Table 22 shows that Addis Ababa (49%) and Dire Dawa (48.4%) scored highest. Benshangul Gumuz (37.6%) and Afar (39%) achieved lowest scores and the mean differences were statistically significant with the highest achieving ones.

Table 22. Homogenous subset groupings of English score by region G12

Region	N	Subset for alpha = .005					
		1	2	3	4	5	6
B. Gumuz	79	37.6					
Afar	64	39.0	39.0				
Amhara	421	40.8	40.8	40.8			
Tigray	237	41.3	41.3	41.3	41.3		
SNNP	504	42.2	42.2	42.2	42.2	42.2	
Oromiya	984	44.0	44.0	44.0	44.0	44.0	44.0
Somali	71		46.0	46.0	46.0	46.0	46.0
Gambella	59			46.4	46.4	46.4	46.4
Dire Dawa	78				48.4	48.4	48.4
Addis Ababa	195					49.0	49.0
Harari	80						49.6
Sig.		**.042**	**.012**	**.147**	**.010**	**.020**	**.166**

Mathematics

In Grade 10 the analysis of variance showed a statistically significant difference in mathematics score ($F_{(10,\ 5514)} = 12.2$, $p < .001$). Table 23 shows that SNNP (36.1%) and Amhara (36.1%) scored highest. Gambella (25.4%) and Afar (28.2%) achieved lowest scores and the mean differences were statistically significant with the highest achieving ones.

Table 23. Homogenous subset groupings of mathematics score by region G10

Region	N	Subset for alpha = 0.005		
		1	2	3
Gambella	118	25.4		
Afar	118	28.7	28.7	
Dire Dawa	112	31.5	31.5	31.5
Somali	156	31.6	31.6	31.6
Tigray	397	32.2	32.2	32.2
Harari	120		32.7	32.7
B. Gumuz	120		34.5	34.5
Oromiya	1803		35.0	35.0
Addis Ababa	351		35.5	35.5
Amhara	1314			35.9
SNNP	916			36.1
Sig.		**0.01**	**0.01**	**0.43**

In Grade 12 the analysis of variance showed a statistically significant difference in mathematics score ($F_{(10,\ 2649)} = 16.6$, $p < .001$). Table 24 shows that Dire Dawa

(59.9%) and Oromia (55.1%) scored highest. Gambella (36.6%), Somali (45.1%) achieved lowest scores and the mean differences were statistically significant with the highest achieving ones.

Table 24. Homogenous subset groupings of mathematics score by region G12

Region	N	Subset for alpha = .005			
		1	2	3	4
Gambella	59	36.6			
Somali	70	45.1	45.1		
B. Gumuz	79	45.7	45.7	45.7	
Addis Ababa	196		49.9	49.9	49.9
Tigray	119		51.8	51.8	51.8
Afar	67		54.3	54.3	54.3
Harari	79		54.4	54.4	54.4
SNNP	506		55.1	55.1	55.1
Oromiya	985		55.9	55.9	55.9
Amhara	422			56.6	56.6
Dire Dawa	78				59.9
Sig.		**.063**	**.007**	**.006**	**.024**

Biology

The analysis of variance showed a statistically significant difference in biology score ($F_{(10,\ 5491)} = 14.78$, $p < .001$). Table 25 shows that SNNP (44.1%) and Addis Ababa 42.0%) scored highest. Afar (32.3%) and Afar (33.4%) achieved lowest scores and the mean differences were statistically significant with the highest achieving ones.

Table 25. Homogenous subset groupings of biology score by region G10

Region	N	Subset for alpha = 0.005			
		1	2	3	4
Afar	115	32.3			
Gambella	118	33.4	33.4		
Somali	156	35.6	35.6	35.6	
Tigray	393	35.8	35.8	35.8	35.8
Harari	120	39.1	39.1	39.1	39.1
Oromiya	1809	40.1	40.1	40.1	40.1
B. Gumuz	119	40.3	40.3	40.3	40.3
Dire Dawa	107	40.3	40.3	40.3	40.3
Amhara	1307		40.9	40.9	40.9
Addis Ababa	348			42.0	42.0
SNNP	910				44.1
Sig.		**0.01**	**0.02**	**0.14**	**0.01**

In Grade 12 the analysis of variance showed a statistically significant difference in biology score ($F_{(10, 2750)} = 14.78$, $p < .001$). Table 26 shows that Dire Dawa (63.2%) and Harari (58.7%) and Gambella (57.8%) scored highest. Somali (44.1%) achieved lowest score and the mean differences were statistically significant with the highest achieving ones.

Table 26. Homogenous subset groupings of biology score by region G12

		Subset for alpha = .005	
Region	**N**	**1**	**2**
Somali	71	44.1	
B. Gumuz	79	53.5	53.5
Tigray	237	53.6	53.6
Addis Ababa	195	54.6	54.6
Amhara	410	54.8	54.8
SNNP	504	55.5	55.5
Afar	64		56.0
Oromiya	985		56.4
Gambella	59		57.8
Harari	80		58.7
Dire Dawa	77		63.2
Sig.		**.009**	**.072**

Chemistry

In Grade 10 the analysis of variance showed a statistically significant difference in chemistry score ($F_{(10, 5465)} = 12.4$, $p < .001$). Table 27 shows that SNNP (38.8%) scored highest. Gambella (27.8%) and Afar (29.3%) achieved lowest scores and the mean differences were statistically significant with the highest achieving ones.

Table 27. Homogenous subset groupings of chemistry score by region G10

Region	N	Subset for alpha = 0.005		
		1	2	3
Gambella	118	27.8		
Afar	114	29.3	29.3	
Somali	152	31.6	31.6	31.6
Harari	115	32.7	32.7	32.7
Dire Dawa	107	34.2	34.2	34.2
Addis Ababa	336	34.7	34.7	34.7
Tigray	396	34.8	34.8	34.8
Amhara	1303		36.2	36.2
B. Gumuz	120		36.6	36.6
Oromiya	1803		36.7	36.7
SNNP	912			38.8
Sig.		**0.017**	**0.007**	**0.013**

In Grade 12 the analysis of variance showed a statistically significant difference in chemistry score ($F_{(10, 2723)} = 7.1$, $p < .001$). Table 28 shows that Dire Dawa (40.9%) and Afar (38.7%) scored highest. Gambella (28.3%) and Somali (28.7%) achieved lowest scores and the mean differences were statistically significant with the highest achieving ones.

Table 28. Homogenous subset groupings of chemistry score by region G12

Region	N	Subset for alpha = .005		
		1	2	3
Gambella	60	28.3		
Somali	71	28.7	28.7	
B. Gumuz	77	32.0	32.0	32.0
Harari	79	32.4	32.4	32.4
Addis Ababa	190	34.7	34.7	34.7
SNNP	491	36.1	36.1	36.1
Tigray	237	36.8	36.8	36.8
Amhara	410	37.2	37.2	37.2
Oromiya	971		38.2	38.2
Afar	63			38.7
Dire Dawa	78			40.9
Sig.		**.015**	**.006**	**.015**

Physics

The analysis of variance showed a statistically significant difference in physics score ($F_{(10, 5465)} = 13.0$, $p < .001$). Table 29 shows that SNNP (33.7%) scored highest. Afar

(23.5%) and Gambella (24.3%) achieved lowest scores and the mean differences were statistically significant with the highest achieving ones.

Table 29. Homogenous subset groupings of physics score by region

Region	N	Subset for alpha = 0.005			
		1	2	3	4
Afar	115	23.5			
Gambella	117	24.3	24.3		
Somali	152	26.5	26.5	26.5	
Harari	115	27.9	27.9	27.9	27.9
Dire Dawa	107	29.3	29.3	29.3	29.3
Addis Ababa	336	30.1	30.1	30.1	30.1
B. Gumuz	120		30.6	30.6	30.6
Oromiya	1804			31.5	31.5
Tigray	396			31.7	31.7
Amhara	1302			31.7	31.7
SNNP	912				33.7
Sig.		**0.01**	**0.03**	**0.16**	**0.07**

In Grade 12 the analysis of variance showed a statistically significant difference in physics score ($F_{(10, 2716)} = 7.3$, $p < .001$). Table 30 shows that Dire Dawa (40.9%) and Afar (38.7%) scored highest. Gambella (24.3%) and Somali (28.7%) achieved lowest scores and the mean differences were statistically significant when compared with the highest achieving ones.

Table 30. Homogenous subset groupings of physics score by region

Region	N	Subset for alpha = .005		
		1	2	3
Gambella	60	28.3		
Somali	71	28.7	28.7	
B. Gumuz	77	32.0	32.0	32.0
Harari	79	32.4	32.4	32.4
Addis Ababa	190	34.7	34.7	34.7
SNNP	491	36.1	36.1	36.1
Tigray	237	36.8	36.8	36.8
Amhara	410	37.2	37.2	37.2
Oromiya	971		38.2	38.2
Afar	63			38.7
Dire Dawa	78			40.9
Sig.		**.015**	**.006**	**.015**

4.2 Comparisons of Achievement scores between Boys and Girls across Regions

This part looks the five achievement scores across the regions by taking sex as a disaggregating variable. Looking at the average score in all the regions boys performed better than girls and the differences were statistically significant in most cases (Table 31).

Table 31. Mean scores (%) of the five subjects by gender across regions G10

Region	Sex	English	Mathematics	Biology	Chemistry	Physics	Average
Tigray	Male	36.4	35.2	38.8	36.9	33.1	36.2
	Female	32.2	29.6	33.1	33.0	30.4	31.7
Afar	Male	33.6	30.6	33.3	30.5	23.7	30.6
	Female	30.1	25.8	30.8	27.5	23.2	27.6
Amhara	Male	41.4	39.4	45.8	40.1	34.4	40.3
	Female	34.0	31.7	35.1	31.4	28.5	32.1
Oromiya	Male	39.5	37.7	43.7	39.4	33.3	38.7
	Female	33.7	31.0	34.7	32.7	28.8	32.3
Somali	Male	40.8	33.1	37.4	33.6	26.8	34.4
	Female	32.8	28.2	31.4	26.8	25.6	29.0
B. Gumuz	Male	42.9	37.6	45.1	39.4	33.3	39.7
	Female	33.5	28.5	30.6	31.2	25.4	29.9
SNNP	Male	40.3	38.3	47.1	41.1	34.9	40.2
	Female	34.6	31.8	38.4	34.3	31.3	34.1
Gambella	Male	35.4	26.0	38.3	30.0	24.1	30.7
	Female	27.1	24.6	27.7	25.3	24.5	25.9
Harari	Male	36.7	35.5	41.9	36.8	29.8	36.5
	Female	33.8	27.1	33.7	25.1	24.2	28.7
Addis Ababa	Male	41.7	37.3	44.4	37.6	32.1	38.8
	Female	39.1	34.1	40.2	32.3	28.6	35.2
Dire Dawa	Male	41.8	34.6	46.9	38.2	31.9	39.1
	Female	30.5	27.0	31.4	28.8	25.6	28.9

In Grade 12 looking at the average score in all the regions except Benshangul Gumuz boys performed better than girls and the differences were statistically significant (Table 32). Among those who took the tests in Gambella there were no females.

Table 32. Mean scores (%) of the five subjects by gender across regions G12

Region	Sex	N	English	Math	Biology	Chemistry	Physics	Average
Tigray	Male	192	42.1	55.1	55.4	50.0	38.6	47.7
	Female	47	37.6	39.9	46.4	41.0	29.7	37.0
Afar	Male	51	39.8	56.1	58.1	48.6	38.3	48.2
	Female	16	36.4	48.4	49.2	38.1	40.2	43.1
Amhara	Male	345	41.4	58.2	56.7	50.7	38.5	49.2
	Female	86	38.1	49.6	45.7	38.7	31.0	40.9
Oromiya	Male	817	44.6	57.8	57.9	51.8	39.1	50.3
	Female	172	41.6	47.1	49.1	42.4	33.4	42.8
Somali	Male	54	48.2	48.0	47.4	41.5	29.4	42.5
	Female	17	38.8	36.3	33.5	32.5	26.5	33.5
B. Gumuz	Male	64	38.0	46.3	52.9	48.6	31.3	43.5
	Female	15	36.1	43.4	56.0	48.6	35.0	43.8
SNNP	Male	397	43.1	56.9	57.6	52.2	37.3	49.4
	Female	116	39.1	49.0	48.6	44.4	32.0	42.6
Harari	Male	51	49.8	56.0	58.0	50.9	33.1	50.0
	Female	29	49.3	51.7	59.8	49.5	31.3	48.3
Addis Ababa	Male	111	49.1	52.5	57.3	47.9	38.0	49.1
	Female	85	48.7	46.6	51.1	39.0	30.4	43.0
Dire Dawa	Male	60	49.1	61.5	62.9	54.5	43.6	54.2
	Female	20	46.3	55.2	64.2	55.1	33.0	50.8

4.3 Students Background Variables and Academic Achievement

Series of questions believed to be related to academic achievement were posed to the sample students. These questions were related to back ground information about parents, home environment, socio-economic status, possession of educational materials, and students' characteristics among others.

4.3.1 Correlates with the average score

A number of student level variables showed positive or negative relationships with the average scores. Even though, the relationships were mostly statistically significant they were found very week in almost all cases (Annex 1 – 4).

In Grade 10 the relation between family literacy level and their child's achievement score was low although the correlation was statistically significant [mothers literacy level (r = -.098) and father literacy level (r = -.082)] in the negative direction. Students of families with high education/literacy level are not achieving higher scores compared to pupils of lower literacy level. In Grade 12 too more or less similar result was obtained [mothers literacy level (r = -.129) and father literacy level (r = -.110)].

Family economic status is another factor that was expected to have an influencing potential on pupils' achievement. In this regard, the examinees were asked to label their family's economic status as "Low", "Medium" or "High". In both Grades 10 and 12 the correlations were weak and in the negative direction (r = -.071) and (r = -.051).

Daily meal service for the pupils was assumed to have an impact on their learning and their achievement might be affected by lack or less provision of meal. From the information provided by the examinees themselves, their achievement score was correlated to the number of times they were provided meal in a day. In both Grades 10 and 12, pupils who have the opportunity of getting meal three times in a day scored higher than those getting once or twice a day. However, the correlations were very week and in case of Grade 10 not even statistically significant.

The distance pupils travel to go to their schools was also hypothesized to have an impact on their achievement. Students traveling more distance seem to achieve less than the other groups. The correlations were weak but statistically significant in both grades.

Supports given to pupils' learning have positive effect on their achievement. Among many types, supports given during their study time is a major one. In order to investigate whether this intervention has a contribution to student's achievement, a question "Do you get support during study time?" was posed to them. From the information received, students' achievement scores were organized and compared. In the subjects students tested, mean achievement scores of students who were getting support were found to be higher than those who were not getting.

Pupils tested were categorized on the basis of their attitude towards some school factors and the mean achievements in the composite score were compared. The mean scores of the groups with positive attitude on the school factors were slightly better than the other groups. The correlations between positive attitude toward school with achievement score was found to be statistically significant in both grades.

Some activities those students perform outside school hours were identified and pupils were asked the length of time they spent on. Mean scores of pupils who spent much time watching television outside their school hours were found negatively correlated with the achievement score in Grade 10. In Grade 12, there no relationship was observed. On the other hand, in Grade 12 spending much time on supporting their parents by work seems to have negative effect on their achievement scores but not in Grade 10. Time spent on homework was positively related with achievement score in both grades.

The achievement score of the tested students was also related with the average number of days they were absent from school in that year. Absenteeism was found negatively correlated with the achievement scores in both grades.

In Grade 12, schools using plasma system were identified and the mean achievement scores of pupils in the schools using plasma system and those not using the system were compared. Except in biology, there is no observable mean difference in pupils achievement in the subjects tested between the two groups.

Pupils were also asked to indicate their evaluation of the effectiveness of the plasma system by rating as *High*, *Medium* or *Low*. Their achievement scores in the five subjects were related. It was found that their evaluation of the effectiveness of the system has significant positive correlation with their achievement scores except in physics. Pupils were asked through the questionnaire, whether they understand their plasma teacher or not and their test scores for each subject were related with their response. Those pupils who could understand the plasma teachers performed better.

4.3.2 Correlates with subject specific achievement

Five questions believed to affect academic achievement were included in each test booklet. These were about possession of textbook and reference books, interest in the subject, self concept and self efficacy in each subject.

In Grade 10, possession of textbooks in all subjects showed positive and statistically significant correlation with the respective subject mean score. The relationship was strongest in biology (r = 131) and least in physics (r = .064). As to the use of reference books in all subjects the relationships were positive but weak. Only in the correlations with mathematics (r = .031) and physics (.064) were statistically significant.

In all cases, students who like the respective subject performed better than those who did not and the correlations were statistically significant. The highest correlation was observed in mathematics (r = 176). Students who said they have high competency in the respective subject performed better than those with low competency. The correlations were statistically significant in all cases.

In Grade 12, possession of textbooks in all subjects showed positive and statistically significant correlation with the respective subject mean score. The relationship were strongest in physics (r = .142) and in English (r = .140) and weakest in chemistry (r = .061). As to the use of reference books in physics it was strongest (r = .194) and weakest in biology (r = .082).

In all cases, students who like the respective subject performed better than those who did not the correlations were statistically significant. The highest correlation was observed in physics (r = .214). Students who said they easily understand their subjects performed better than those who did not. Students who said they have high competency in the respective subject performed better than those with low competency level. The correlations were statistically significant in all cases.

Availability of textbooks for the students and their scores in the tested subjects were also related. Pupils having a textbook for his/her own scored better than those sharing or do not have in case of all the five subjects.

Students using reference materials for each of the five subjects were found to be achieving better than those who do not use. Particularly the mean gaps for physics and mathematics are large compared to other subjects. In all cases the mean differences observed between the groups were found to be statistically significant.

Students were asked to indicate their preference for liking/disliking the subjects they were tested. Students liking the subjects are performing better in the tests. The mean scores of the student groups observed were compared and in all cases the mean differences were found to be statistically significant.

Pupils confidently identifying themselves as "*do understand the subjects*" were doing better in the tests and the mean difference observed between those group who say "*understand"* and those who say "*do not"* were found to be statistically significant.

Mean scores of those pupils who level themselves as *high* were achieving higher score than those leveling *medium* or *low* themselves. Their test scores and their ratings as high, medium or low were correlated and coefficients calculate were found to be significant.

4.3.3 Multiple Regression Analysis

The data were analyzed by multiple regressions, using the student level variables believed to be related with academic achievement as regressors.

In Grade 10 the regression model was good fit (R^2 = .5094), and the overall relationship was statistically significant ($F_{22, 3986}$ = 188.14, p < 0.001). This means the regression model was able to explain 50.94% of the variations observed between the students in the average score (Table 33).

Table 33. Multiple regression analysis Grade 10

Source	SS	df	MS
Model	333731.79	22	15169.6268
Residual	321394.888	3986	80.6309302
Total	655126.678	4008	163.45476

Number of obs	4009
F(22, 3986)	188.14
Prob > F	0.0000
R-squared	0.5094
Adj R-squared	0.5067
Root MSE	8.9795

The strongest determinant of the achievement score were the average school based assessment result, gender, self efficacy, attitude towards school, time on homework,

and time in school. When school based assessment score was excluded the model becomes rather weak fit and the variance explained was only 17.8% (Table 34).

Table 34. Multiple regression analysis Grade 10

Source	SS	df	MS
Model	118107.361	20	5905.36804
Residual	545288.054	4052	134.57257
Total	663395.414	4072	162.916359

Number of obs	4073
F(20, 4052)	43.88
Prob > F	0.0000
R-squared	0.1780
Adj R-squared	0.1740
Root MSE	11.601

In Grade 12 the regression model was also good fit (R^2 = 0.4111), and the overall relationship was statistically significant ($F_{(21, 2170} = 72.15$, $p < 0.001$). This means the regression model was able to explain 41.11% of the variations observed between the students in the average score (Table 35).

Table 35. Multiple regression analysis Grade 12

Source	SS	df	MS
Model	140703.282	21	6700.15629
Residual	201519.346	2170	92.8660582
Total	342222.628	2191	156.194719

Number of obs	2192
F(21, 2170)	72.15
Prob > F	0.0000
R-squared	0.4111
Adj R-squared	0.4054
Root MSE	9.6367

The strongest determinant of the achievement score were the average school based assessment result, gender, self efficacy, attitude towards school, time on homework, and time in school. When school based assessment score was excluded the model becomes rather weak fit and the variance explained was only 17.26%.

Table 36. Multiple regression analysis Grade 12

Source	SS	df	MS
Model	64318.7497	20	3215.93749
Residual	308348.176	2373	129.940234
Total	372666.926	2393	155.732104

Number of obs	2394
F(20, 2373)	24.75
Prob > F	0.0000
R-squared	0.1726
Adj R-squared	0.1656
Root MSE	11.399

4.3.4 Variance Partitioning

Using student level data as "*Level 1"* and school level data as "*Level 2"* a fully unconditional model was specified to estimate the variance component for the average score. Taking the average score as dependent variable and the school as random factor the variance component analysis resulted in an intra-class correlation (roh) of 0.165 in Grade 10 and .085 in Grade 12. This means in Grade 10, 16.5% and in Grade 12, 8.5% of the variations in the average score were due to variations that come from the schools. On the other hand when region is taken as random factor the variance component analysis resulted in an intra class correlation of .052 for Grade 10 as well as Grade 12. This means 5.2% of the variation in the average score was due to variations that come from regions in both grades.

5 Conclusions and Recommendations

In Ethiopia, huge progress has been made at all levels of education in the past years. Enrolments are up in both primary and secondary education, with millions more able to go to school. These are no easy achievements; they have required significant political commitment and larger allocations of public resources. What matters most, however, is how these achievements translate into concrete improvements in students' competencies. The truly important measure then is not only enrollment and attendance but how learning is going on. Based on the findings of the studies the following conclusions and recommendations are made.

5.1 Conclusions

The academic achievement of the students as measured by the mean score of the four subjects namely English, mathematics, biology, chemistry and physics was found less than the 50% achievement level set by the Education and Training Policy of Ethiopia. The national mean score (the average of what the students scored in the five subjects) was only 36% for Grade 10 and 47.8% for Grade 12.

Looking at the mean score of each subject, in none of the subjects tested did the scores of students' amount to 50% for Grade 10. In Grade 12, the mean scores for biology and mathematics were found higher than 50%.

In all subjects in both Grades 10 and 12 boys performed better than girls and the differences were statistically significant in all cases. In Grade 10, there were 3,151 males and 2260 females in the national sample. Boys scored an average of 38.7% whereas girls had an average of 32.2%. In Grade 12, there were 2017 males and 561 females in the national sample. Boys scored an average of 49.2% whereas girls had an average of 42.6.

The students' achievement scores were divided into four standards as: *Below Basic*, *Basic, Proficient* and *Advanced*. The scaled scores were used instead of raw scores where: *Advanced* is greater than 2 standard deviations from the mean, "*Proficient*" is between 1 and 2 standard deviations above the mean, "*Basic***"** is within 1 standard deviation above the mean and "*Below Basic***"** is below the mean score.

In Grade 10 the proportion achieving each level based on the national achievement of the average score are: 6.1% *Advanced*, 9.4% *Proficient*, 20.8% *Basic*, and the remaining 63.7% *Below Basic* level.

In Grade 12, the proportion achieving each level based on the national achievement of the average score are: 4.7% advanced, 12.3% proficient, 28.3% basic, and the remaining 55.3% below basic level.

The average of the five subjects and mean score of each subject when disaggregated across regions showed that there exist disparities in academic achievement among regions. In Grade 10, no single region achieved the minimum requirement of at least 50% and the mean scores in Gambella (28.5%) and Afar (29.4%) were found much lower when compared with the highest achieving regions. In Grade 12, Dire Dawa was the only region that achieved a mean score greater than 50%. Somali (40.3%), Gambella (43.5%), and SNNP (43.5%) were the least.

Multiple regression analysis based on the students' background questionnaires resulted in a model which was able to explain 17.8% in Grade 10 and 17.3% in Grade 12 of the variations observed in the average scores at student level..

The variance partitioning based on fully unconditional hierarchical linear model result, that took student level and school level data simultaneously, showed that in Grade 10 16.5% and in Grade 12 8.8% of the observed difference in academic achievement came from differences in schools.

The same model that took school level and region level data simultaneously showed that in both Grades 10 and 12, 5.2 % of the observed difference in academic achievement was due to differences between the regions.

5.2 Recommendations

Measures that will help to raise the achievement levels of Grade 10 and 12 students should be taken at all levels. Based on the findings of the studies the following recommendations are made.

The observed low achievement scores within and between groups call for individualized approaches of teaching. Teachers should be trained and become familiar to techniques that help to diagnose and intervene at individual student level.

The gender gap in primary as well as secondary schools is persistent and there is a need to make further investigations that would help to narrow the gap between boys and girls. All concerned bodies should address provision of additional supports for girls.

The existence of wide variations with in groups call for teachers to incorporate greater differentiation of teaching practices into their classroom. Such differentiations should address the needs of both low and high achieving students.

Schools and teachers should be facilitated and supported through professional development trainings in extending the use of all available resources and sharing experiences with one another.

A tailor made learning support program should be introduced at national and regional levels. The primary beneficiaries of this scheme should be disadvantaged areas such as emerging regions.

Mastery of the medium of instruction is a key to read and understand other subjects. As long as English continues as instructional language due attention should be given to the subject at all levels.

References

NOE (2000). *Ethiopian Baseline National Learning Assessment of Grades 4 & 8 Students.* Addis Ababa: NOE/MOE

NOE (2004). *Ethiopian Second National Learning Assessment of Grades 4 & 8 Students.* Addis Ababa: NOE/MOE

Kellaghan, T. (2001a). The Globalization of Assessment in the 20th Century. In *Assessment in Education*, 8, pp. 87-102

Kellaghan, T. (2001b). *Using Assessment to Improve the Quality of Education.* Paris: UNESCO, IIEP

Kellaghan, T. and Greaney, V. (2004). Assessing student learning in Africa, Washington D.C: The World Bank

Kellaghan, T. and Greaney, V. (2001). The globalization of assessment in the 20th Century. In *Assessments in Education*, 8, pp. 87-102

UNESCO (2000). *The Dakar Framework for Action. Education for All: Meeting Our Collective Commitments.* Text adopted by World Education Forum. Dakar Senegal, 26-28, Paris: UNESCO

UNICEF (2000). *Education Update, "Curriculum & Learning"*. UNICEF Vol. 3, Issue I

World Declaration on Education for All (1990). *World Declaration on Education for All*. Adopted by the World Conference on Education for All: Meeting Basic Learning Needs, New York: UNDP/ UNESCO/ UNICEF/ World Bank

Appendix i. Correlates with the average score Grade 10

	N	Correlation	Sig. (2-tailed)
female	5379	-.238**	.000
age	5204	.075**	.000
mother (or stepmother or female guardian) literate	5353	.067**	.000
highest level of education completed by mother	3395	-.098**	.000
father literate	5293	.019	.163
highest level of education completed by father	3664	-.082**	.000
economic status of the family	5336	-.071**	.000
meals a day	5321	.003	.815
distance from home to school	5342	-.054**	.000
help at home in studies	5312	-.174**	.000
I like school	5368	.237**	.000
my teachers care about me	5361	.130**	.000
I make friends easily at school	5328	.097**	.000
reading for all the subjects outside schools	5363	.198**	.000
reading for enjoyment and/or general interest	5352	-.064**	.000
doing sports or other activities	5351	-.079**	.000
taking extra school lessons or going to tutors	5355	-.015	.285
watching television or movies	5363	-.082**	.000
Supporting my parents by work	5360	.058**	.000
time on homework	5340	.163**	.000
days absent from school	5349	-.112**	.000
school use the plasma system	5033	.069**	.000
effective is the plasma system	4635	.079**	.000
understand the language plasma teacher	4581	.037*	.013

Appendix ii. Subject specific correlates with academic achievement Grade 10

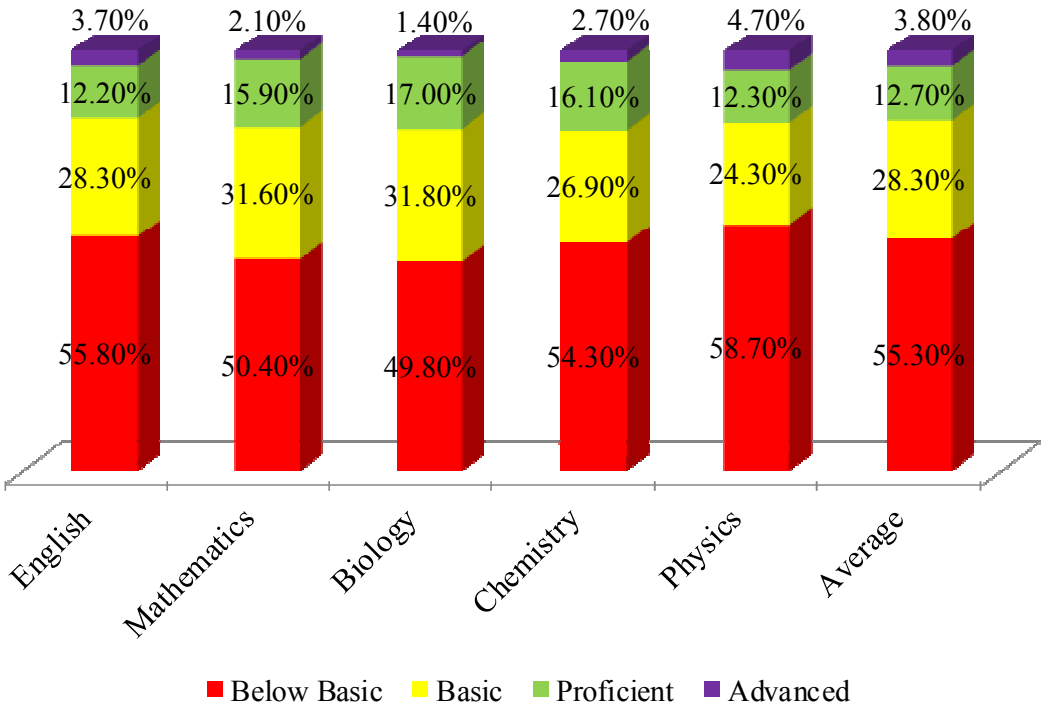
	N	**Correlation**
possession of English textbook	5261	.109**
possession of mathematics text book	5357	.116**
possession of biology text book	5412	.131**
possession of chemistry text book	5363	.089**
possession of physics textbook	5317	.064**
use English reference books	5283	.021
use mathematics reference books	5369	.042**
use biology reference books		
use chemistry reference books	5358	.019
use physics reference books	5338	.042**
like English	5283	.144**
like mathematics	5301	.176**
like biology	5399	.129**
like chemistry	5359	.153**
like physics	5356	.098**
understand English	5275	.073**
understand mathematics	5276	.031*
understand chemistry	5320	.081**
understand biology	5389	.104**
understand physics	5342	.022
competency in English	5160	.170**
competency in mathematics	5193	.154**
competency in biology	5281	.211**
competency in chemistry	5179	.194**
competency in physics	5155	.112**

Appendix iii. Correlates with the average score Grade 12

	N	Correlation	Sig. (2-tailed)
female	2552	-.221**	.000
age	2574	-.046	.020
mother (or stepmother or female guardian) literate	2547	.051**	.009
highest level of education completed by mother (or stepmother or female guardian)	1904	-.101**	.000
father (or stepfather or male guardian) literate	2533	.006	.781
highest level of education completed by your mother (or stepmother or female guardian)	2067	-.133**	.000
economic status of your family relative to the others	2543	-.051*	.010
how many meals a day do you eat	2536	.075**	.000
distance from home to school	2544	-.078**	.000
help in your studies at home	2548	-.166**	.000
like school	2550	.173**	.000
my teachers care about me	2547	.019	.343
I make friends easily	2546	.031	.122
outside-of-class reading for all the subjects	2544	.219**	.000
reading for enjoyment and/or general interest	2544	-.114**	.000
doing sports or other school and community activities	2546	.001	.952
taking extra school lessons or going to tutors	2542	-.042*	.033
watching television or movies	2548	.017	.379
supporting my parents by work	2542	-.068**	.001
time on homework in all of your school subjects	2542	.127**	.000
days absent from school over this school year	2544	.026	.198
school use the plasma system	2504	.023	.247
effectiveness of the plasma system	2367	-.079**	.000
understand the language of your plasma teacher	2372	-.104**	.000

Appendix iv. Subject specific correlates with academic achievement Grade 12

	N	Correlation	Sig. (2-tailed)
possession of English textbook	2569	.140(**)	.000
possession of mathematics text book	2548	.091(**)	.000
possession of biology textbook	2643	.080(**)	.000
possession of chemistry textbook	2568	.061(**)	.002
possession of physics textbook	2583	.142(**)	.000
use of English reference books	2524	.164(**)	.000
use of mathematics reference books	2528	.150(**)	.000
use of biology reference books	2631	.082(**)	.000
use of chemistry reference books	2553	.103(**)	.000
use physics reference books	2554	.194(**)	.000
like English	2534	.178(**)	.000
like biology	2637	.144(**)	.000
like mathematics	2535	.234(**)	.000
like chemistry	2553	.147(**)	.000
like physics	2564	.241(**)	.000
understand English	2515	.188(**)	.000
understand mathematics	2510	.322(**)	.000
understand biology	2612	.201(**)	.000
understand physics	2536	.237(**)	.000
understand chemistry	2536	.241(**)	.000
competency in English	2511	.270(**)	.000
competency in mathematics	2466	.369(**)	.000
competency in biology	2530	.311(**)	.000
competency in chemistry	2490	.280(**)	.000
competency in physics	2481	.304(**)	.000


3.70%
12.20%
28.30%
55.80%
2.10%
15.90%
31.60%
50.40%
1.40%
17.00%
31.80%
49.80%
2.70%
16.10%
26.90%
54.30%
4.70%
12.30%
24.30%
58.70%
3.80%
12.70%
28.30%
55.30%
English
Mathematics
Biology
Chemistry
Physics
Average
Below Basic
Basic
Proficient
Advanced