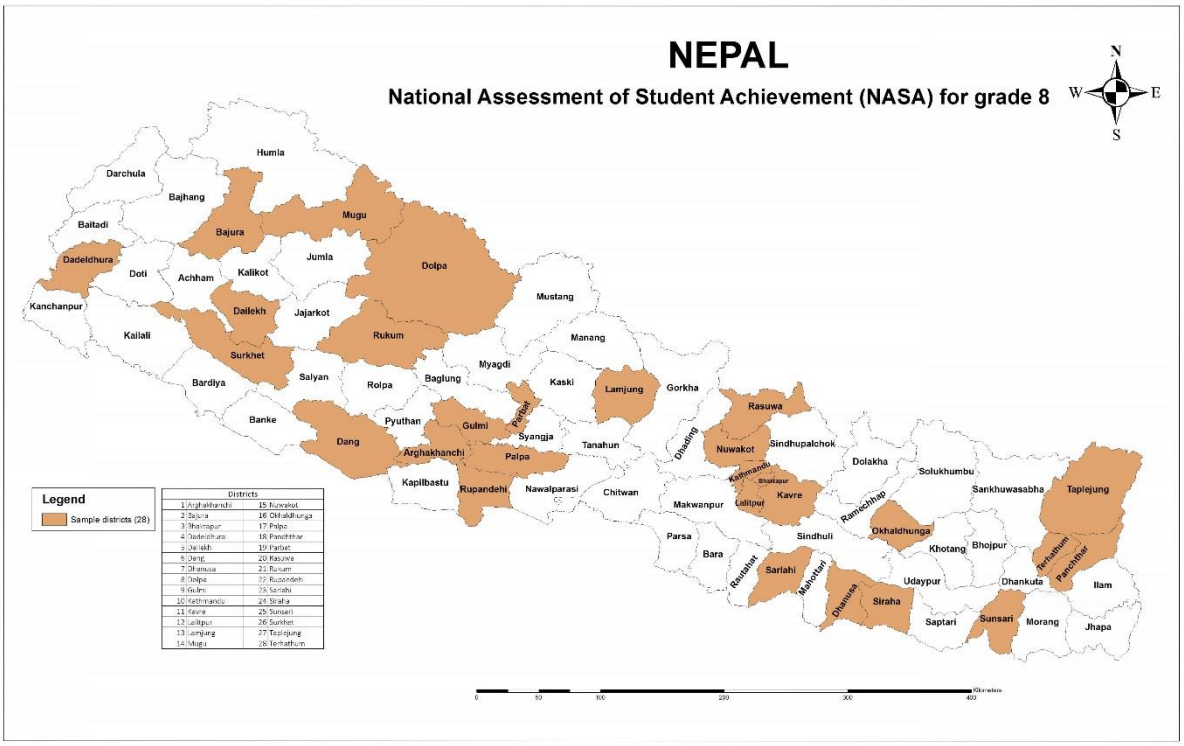
Final Draft

Report on National Assessment of Student Achievement (NASA) 2013

(Grade 8: Mathematics, Nepali and Science)



Government of Nepal

Ministry of Education

Education Review Office

Sanothimi, Bhaktapur

2015

Report on National Assessment of Student Achievement, 2013

(Grade 8: Mathematics, Nepali and Science)

Editing (Content):

Dr Lekha Nath Poudel
Dr Tulasi Thapaliya
Dr Peshal Khanal

Language correction:

Dr Binod Luitel

Data analysis and report writing:

Gopal Prasad Bhattarai
Sushil Babu Khanal
Shyam Acharya
Devi Ram Acharya
Sunita Shakya

Government of Nepal
Ministry of Education
Education Review Office
Sanothimi, Bhaktapur
Nepal

Acknowledgement

This report on National Assessment of Student Achievement is a result of collaborative efforts of various contributors. First, our appreciation goes to the students who took part in this assessment, teachers and head teachers who responded questionnaire. Second, our sincere appreciation also goes to subject teachers and subject specialists including subject committee members for their effort in developing and finalizing the tools for this assessment. Third, we acknowledge teachers, head teachers and district education officers, ERO focal points and other staff at respective districts for their support on test administration.

Our sincere thanks go to Dr Tirth Raj Khaniya and the consulting firm ESDC for timely administrating the test, marking the answer paper and providing all datasets in OMR. We thank Mr Shyam Prasad Acharya, Mr Devi Ram Acharya and Ms Sunita Shakya from Education Review Office for their efforts on timely analysis of data. We thank Mr Gopal Prasad Bhattarai and Sushil Babu Khanal for coordinating the whole process of data analysis and preparing the draft report.

We express our gratitude to Dr Tulasi Prasad Thapaliya and Dr Peshal Khanal for their contribution on content editing of this report. Similarly, we also appreciate Dr Binod Luitel for correcting English language of the report. We express our gratitude to Dr Jari Metsämuuronen who worked as a consultant for previous NASA in Nepal and supported for data analysis of this study through responding emails. Due thanks also goes to Dr Bhojraj Kafle for his supportive role for initiating the process. We acknowledge preceding Joint secretaries of ERO Dr Kamaleshwar Kumar Sinha, Mr Khaga Raj Baral and Dr Hari Prasad Lamsal for their leading role during planning and data generating process of this study.

The Steering Committee headed by the secretary of the Ministry of Education deserves sincere thanks for facilitating National Assessment of Student Achievement process by providing policy guidelines. Our sincere appreciation also goes to the Technical Committee headed by Dr Iswhor Prasad Upadhyya for their technical support.

Finally, we are grateful to everyone who contributed for accomplishing the NASA 2013 and preparing the report in this form.

Dr Lekha Nath Poudel
Joint Secretary
MOE, Education Review Office

Executive Summary

Context

Government of Nepal (GON) established Education Review Office (ERO) in 2010 as an integral part of reform in education quality under the School Sector Reform Program (SSRP) 2009-2015. The main aim of establishing ERO is to inform regularly to the educational stakeholders including the government, teachers, parents, schools, students and civil society about the effectiveness, efficiency, equity in education and the quality of education.

Since its establishment, ERO has continuously been engaged in conducting large-scale assessment of student achievement for grades eight, five and three in order to examine the quality, equity and efficiency of the education system in Nepal. In the first year (2011), ERO conducted National Assessment of Student Achievement (NASA) in 28 sample districts for grade eight students in three subjects: Mathematics, Nepali language and Social Studies. In the subsequent year, the assessment was conducted in 28 sample districts in Mathematics and Nepali subjects for grade three, and Mathematics, English and Nepali for grade five. The present study, NASA 2013, is the second cycle of a large-scale grade eight students' assessment in Mathematics, Nepali language and Science subjects, conducted in 28 sample districts.

Objectives

The overall objective of this assessment was to find out whether the students of grade eight had achieved the goals set by the national curricula in Mathematics, Nepali language and Science. In order to achieve the overall objectives of this assessment, the study not only assessed the overall achievement levels in the selected subjects, but also analysed the variations in student achievements among different strata of population with considerations for and associations of various school related diversity factors on achievement. Furthermore, it examined the extent to which the home background and other pupil related factors have influenced the learning achievement.

Population, Sample and the Method

The population for this assessment was all eighth grader students of Nepal from the 8000 schools that were running grade eight in 2013 across 75 districts of Nepal. The sample size for this assessment was 44067 students, including 48% boys and 52% girls from the 1199 randomly selected schools of 28 sample districts, representing each strata of population. Besides, all the head teachers of the sample schools and 1199 subject teachers responded to related questionnaires. The schools were selected from both the 'community' and 'institutional' categories as well as from rural and urban areas across the sample districts representing all development regions (Eastern, Central, Western, Mid-Western, and Far-Western) including the Kathmandu Valley and all ecological zones (Mountain, Hill and Tarai). So, the results of the assessment can credibly be generalized to the whole students and school population of the entire country.

For this assessment, three parallel versions of test items in each subject were administered. Assuring the reliability and validity of the test, items were constructed following the specification grids and national curriculum of the respective subjects.

From methodological standpoint, the study followed standard measures available in the present context. Further, it has attempted to contextualize the procedures to the Nepali education system. To establish comparability, the results were linked with the set of results from the first assessment cycle of 2011. Further, for comparing the study results with the international assessment results, linking items were included from the item banks of international tests like Trends in International Mathematics and Science Study (TIMSS) and Program for International Student Assessment (PISA).

The tests were administered at a time in one shot in all the sample schools throughout the country in the scheduled day. Each selected school was assigned to conduct test in one of the three subjects. Thus, the students were required to participate in one of the three subjects assigned to the school. The answer sheets were marked and achievement scores were tabulated using Optical Mark Reading (OMR) machine, and the assessment results were presented in percentage of mean score.

To develop standardized items, contents were analysed against the expected performance of the national curriculum, and relative weightages in various content areas and various levels of cognitive domain were prepared. Individual items developed through the workshops of trained subject teachers and subject experts were further revised and selected, and at least six different versions of items were pre-tested. The results of pre-test were analysed by calculating difficulty level of each item. Based on the pre-test results the difficulty levels of items were set around 50% to 60%. In addition to pre-tested items, some linking items from NASA 2011 and international tests like TIMSS and PISA were selected for the test.

Summary of Key Findings

The following are the summarized key results drawn from the analysis of the datasets of this assessment:

1. **Wide differences in achievement among development regions, ecological zones, districts, schools and students**

When district is taken as a unit of analysis, the disparity in the level of achievement is big. For example the lowest performing districts (below 20%) in Mathematics are Panchthar, Rasuwa, Surkhet, Mugu, Terhathum and Dolpa, while in the highest performing district Bhaktapur the achievement score is 59%. In all subjects, the average achievement score of students in the Kathmandu Valley are far ahead compared to the rest parts of Nepal. For example, the achievement level of the students in Science in Eastern Development Region (32%), Far-Western Development Region (32%), and Mid-Western Development region (34%) are far behind compared to that of the Kathmandu Valley (56%).

The most important unit of analysis for the purpose of educational management appears to be the school itself. Schools were categorized in terms of average achievement scores: high performing (above average scores), average performing and under-performing schools. In the under-performing schools, the average achievement of students is less than 15 in a scale of 100. On the other hand, in the high performing schools, the average achievement of students is above 90%. It depicts the huge level of disparity in the achievement of students among the schools. Moreover, there is a greater concentration of students in low performing categories, particularly in the community schools.

Some students in the study could not answer correctly even a single test item and did not obtained any score, for example in Mathematics 0.90 % students obtain 0% score. At the same time, there are some students, who scored more than 95%. For example, in Science 0.30 % students scored more than 95%.

Further analysis of the highest and lowest performing community schools reveals that those students tend to perform higher who afford more time on homework, have positive attitude towards the subjects, receive the required support from siblings or private tuition from teachers, do not need to work for earning while studying, receive textbooks timely, and reach the grades at their correct age. On the other extreme, lowest achievers are those who belong to illiterate parents, especially the mothers who are involved in agriculture work, need to work for earning, receive neither private tuition nor support from family members, and the children not receiving textbook, and so on.

2. Remarkable gap in achievement between institutional and community schools

Students' average achievement score in institutional schools (also known as private schools) is higher than that in the community schools. For example, in Mathematics the average achievement score is 57% in the private schools whereas it is 26% in the community schools. However, it is not clear whether it is due to the effectiveness of instructional processes in the institutional schools or manifestation of the disparity of socio-economic status of students in these two different school system.

It is noteworthy that there are some community schools where the average results are at the same level as in the private schools even though the SES of students is remarkably lower than that of private schools. In these schools, either the educational processes are more effective than in the private schools or the students are of the same ability as those in the private schools and are not adversely affected by the processes within the school or their socioeconomic status. The sporadic success stories in the community schools do not allow for attributing their success to the effect of systemic reforms, as there are many cases to the other side.

3. **Students' performance found better in lower levels of cognitive skills, but poor in higher cognitive skills**

Students are good in lower level of cognitive ability, like knowledge and comprehension, but they are poor in higher level of cognitive ability, like analysis, evaluation and applying the gained knowledge and skills in a new situation. Moreover, students are comparatively poorer in the ability to solve problems, analyse issues, deduce logic, generalize the concept, justify an argument or viewpoint, and in the ability to transfer learning from one context to another. A remarkably higher number of students were able to solve only 15% or less of the practical problems (15% of the students in Mathematics, 11% in Nepali, and 5% in Science).

In Mathematics, students are able to do basic calculations, but are weak in reasoning, problem solving, proving theory or formula, and in constructing shapes and figures. In many cases, the students did not even attempt to complete the open-ended questions of higher cognitive level.

In Nepali, students performed well in the tasks requiring the recognition of correct answer, recalling simple facts from the texts, fundamental thinking, and the basic interpretation of paragraphs. However, they are much weaker in producing fluent texts or essays, and in preparing synthesis and abstracts from a text. In Nepali, the students responded open-ended tasks, but could not meet the expected standards to gain maximum scores. In case of applying the gained knowledge in new situation, they are poorer than in the case of higher level thinking ability.

In Science, the result reveals the fact that the students are good in recognizing the correct answer and in very fundamental knowledge such as choosing the facts and numbers, and writing the definitions. They are much weaker in reasoning, problem solving, proving the principle, and constructing the figures. In many cases, the students did not even start to do the open-ended questions and hence, obtained low score.

4. **Caste/ethnicity and home language based variations in student achievement despite having no definite trend in achievement based on home language**

The achievement level of the Dalit students is lower in all the subjects comparing the students from other communities. However, achievement level for Dalits in the Kathmandu Valley is higher than the total national average. Particularly, achievement of Dalit girls is low in Mathematics

(23%), and Science (33%), compared to the Dalit boys in Mathematics (25%) and Science (37%). Similarly, the achievement of Madhesi students is also lower than the total national average. For example, average achievement of Madhesi students is 29% in Mathematics, 37% in Nepali and 36% in Science.

On the other hand, there is no definite trend in the achievement of various language groups, for instance in Nepali subject, the students from Non-Nepali home language groups Newari and Sherpa have outperformed the rest of the language groups, including the students from Nepali language group. However, the achievement of the students from other language groups (except Newari and Sherpa) are even lower than the achievement of students from the Nepali language group.

5. **Remarkable rural urban differences in student achievement across all subjects**

Urban students have outperformed the rural students in all subjects. For instance, students from urban schools scored 57% in Nepali whereas it is only 45% in the case of students from rural schools. The achievement gap between the students from urban and rural schools is wider in Mathematics as the gap is 24%. Similar level of difference is in Science in which students from the urban schools scored 51% while the students from the rural schools scored only 37%.

6. **Achievement scores going down over the years instead of improvement**

When the results of 2013 were compared with those of 2011, the achievement has not been improved over these years. For instance, in 2011 it was 49% for Nepali, whereas it is only 48% in 2013 - going further down by around 1% in 2013. In Mathematics, it is more worsening as the achievement score has gone down further by almost 8% from 2011 to 2013. The reasons behind such downfall in the learning achievement demand further enquiry to get the root of the fact.

Comparing the achievement scores of NASA 2013 of Mathematics and Science with the achievement scores of an international assessment TIMSS; it is found that the achievements in both subjects are lower than the international means.

7. The unavailability of textbooks associated with the lower level of achievement

Despite the efforts to make the textbooks available in the hands of students timely, the datasets reveal a distressing fact that a significant number of students (5.1% in Mathematics, 2.7% in Nepali and 2.8% in Science) are found studying the subjects without receiving the textbooks of respective subjects even up to the end of the academic year. The strong association regarding the timeliness in textbook availability and student achievement was observed. For instance, the students reporting to have received textbook at the beginning of the academic session have scored 36% in Mathematics, 49% in Nepali and 42% in Science whereas those who did not receive textbooks even at the end of academic year have scored 24%, 35% and 34% respectively in the same subjects.

8. School bullying impeding the learning potential of students

A significant number of students experienced bullying at school, though bullying at school goes unnoticed for many of the parents and school teachers. It worsens the learning environment and negatively affects the learning potential of students, which is taken as negative activities at school. The obtained datasets reveal that bullying is found rampant in Nepalese schools, which has negatively been affecting the students' learning potential. Those students who have not experienced bullying are found to have scored 35% in Mathematics, 50% in Nepali and 42% in Science, whereas students experiencing extreme types of bullying have scored only 26% in Mathematics, 29% in each of Nepali language and Science.

9. Reluctancy to assign and check homework preventing students' progress

Properly assigned and checked homework (feedback provided) become supportive to boost up the learning achievement as it provides opportunities for self-learning and engages students in problem solving. However, the datasets of student assessment show that nearly 37% in Mathematics, 35% in Nepali and 32% students are found never assigned or checked their home works. The students who were assigned homework and provided appropriate feedback were found to have score 36% in Mathematics, 50% in Nepali and 42% in Science, whereas the students who were not assigned home works were able to score just 27%, 34% and 30% in these subjects respectively. It further shows that regular assigning of home works along with feedback has positive impact on achievement score.

Abbreviations

ANCOVA	Analysis of Covariance Variance
ANOVA	Analysis of Variance
BPEP	Basic and Primary Education Programme
CBS	Central Bureau of Statistics
CERID	Research Centre for Educational Innovation and Development
CERSOD	Centre for Educational Research and Social Development
CML	Conditional Maximum Likelihood
DEO	District Education Office
DOE	Department of Education
DTA	Decision Tree Analysis
EDSC	Education and Development Service Centre
EFA	Exploratory factor analysis
ERO	Education Review Office
ES	Effect Size
ETC	Education Training Centre
FNBE	Finnish National Board of Education
GLM	General Linear Modelling
IRT	Item Response Theory
Max	Maximum
Min	Minimum

MML	Marginal Maximum Likelihood
N	Number
NASA	National Assessment of Student Achievement
OECD	Organization for Economic Co-operation and Development
OMR	Optical Mark Reader
OPLM	One Parametric Logistic Model
PEDP	Primary Education Development Project
PIRLS	Progress in International Reading Literacy Study
PISA	Programme for International Student Achievement
SD	Standard Deviation
SE	Standard Error
SES	Socio-economic status
SPSS	Statistical Package/Programme for Social Science
TIMSS	Trends in International Mathematics and Science Study

Table of Contents

Chapter 1: Introduction

This introductory chapter briefly describes the context of the assessment, presents the history of the practice of National Assessment of Student Achievement in Nepal, deals with the student assessment as a process, and points out the characteristics of NASA 2013. Similarly, it includes the objectives of this assessment as well as a brief indication of methods and process applied for this assessment. Finally, it includes the structure of this report.

1.1 The Context

Nepal is a country with about 27 million people having numerous diversities not only with geographical but also with socio-cultural and linguistic diversities. Within the area of 147,181 square kilometres, Nepal is divided into three ecological zones, namely, Mountain, Hill and Tarai in which each ecological zone is situated from east to the west. Ecological zones differ not only in terms of the landscape but also in climatic conditions and settlement patterns. For example, there is cold climatic condition and thinly populated settlements in mountain zone, whereas in Tarai, the climate is warm and population is dense in comparison to other two zones. Administratively, Nepal is divided into 75 districts with five development regions in which each region includes all the ecological zones (i.e, Mountain, Hill and Tarai).

In Nepal, the socio-cultural and linguistic diversities are visible as there are more than 123 ethnic/caste groups and 125 linguistic groups (CBS, 2012). Besides, about a quarter of population is still below the poverty line. Nevertheless, there is a large population below the poverty line, a small section of population is with high economic conditions, and the rest is with medium economic conditions.

Nepali education system is structured into basic, secondary and higher level. The basic education is of eight years, including the first five years of primary and the next three years of lower secondary education, and four years of secondary education, including secondary and higher secondary each of two years. In Nepal, there are 34,743 schools with 4,401,780 students at primary level (grades 1 to 5); 34,923 schools with 6,230,131students at basic level (grades 1 to 8); 8,726 schools with 896,919 students at secondary (grades 9 and 10) level, and 3,596 schools with 415,343 students at higher secondary level (grades 11 and 12) (MOE, 2071 BS). About 20 percent of schools are institutional, usually called private schools, and the rest are public schools. The literacy rates and school enrolments in Nepal have been improved noticeably in the recent years. However, a large mass of population is still illiterate and about four percent children are still

out of school (see DEO, 2014). The high dropout and repetition rates in early years of schooling has resulted into low completion rates at basic level education.

Several interventions, such as rapid expansion of schools; incentives to girls, Dalits and other children from disadvantaged communities; arrangement of alternative schooling programmes; fee waive and free distribution of textbooks the students studying at public schools, and so on have been put to increase access to education, particularly focusing on basic education. Similarly, in order to improve quality of education, teachers were trained and technical support provision was included through supervision system, some programmes have been conducted to improve management of school, and to make classroom as well as school child friendly. Some programmes have also been conducted in order to manage diversity and promote inclusion of diverse student population in education system. After such interventions, some noticeable progress have been made in providing access to education, particularly in basic education. However, quality of education has not been improved as expected and the equity and efficiency of education system has always been questioned. In such situation, a regular mechanism for assessing student achievement as well as performance of the school education system could provide feedbacks for the improvement of quality, equity and efficiency in education system.

Education Review Office (ERO) in Nepal was established in 2010 as an integral part of the reform in education quality under the School Sector Reform Program (SSRP), 2009-2015. The main aim of establishing ERO is to inform regularly to educational stakeholders, including the government, teachers, parents, schools, students and civil society about the effectiveness, efficiency, equity in education and the quality of education so that equity in and quality of education will be improved regularly. In order to provide feedback for policy formulation and programme implementation in education system, ERO is entrusted to assess student achievement regularly, carry out the performance audit of educational institutions and schools, and make assessment and audit report public. In such context, ERO for the first time conducted National Assessment of Student Achievement (NASA) for grade eight students in 2011, and in 2012, it conducted NASA for grades three and five students. This assessment, NASA 2013, is the second cycle of grade eight assessment, which was conducted in March 2013.

1.2 National Assessment of Student Achievement in Nepal

The aim of the national level student assessment is to produce objective, accurate and comparative information about the achievement of students. Normally, national level assessment of student achievement is carried out to analyse the efficiency, effectiveness and equity in education system. The national level

assessment primarily does not prefer to assess the students; rather it focuses on the analysis of education system based on the result of assessment of student achievement. Hence, the assessment can be used as a tool to evaluate the state of the current educational system against the curricula goals.

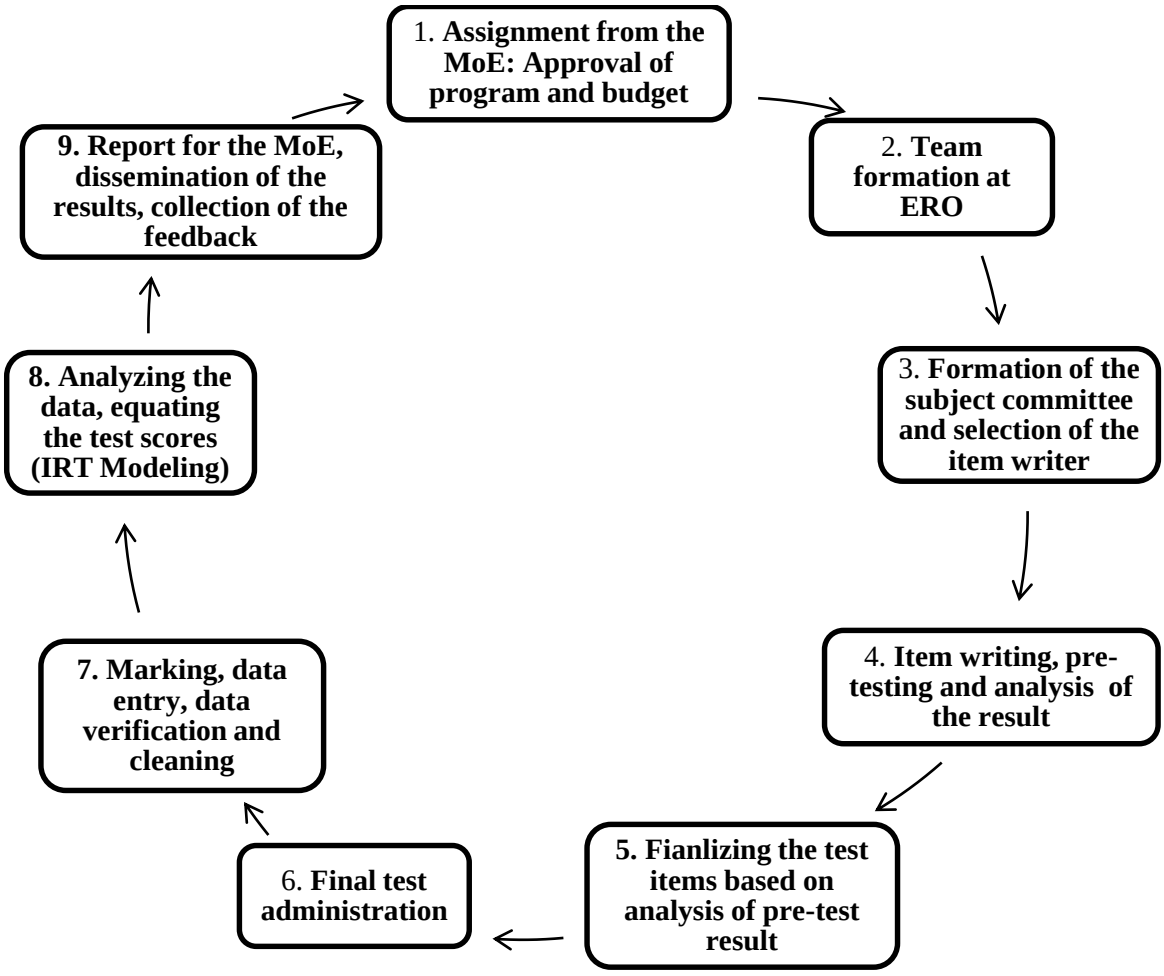
Assessment should be just and fair, valid and reliable, transparent, motivating, and able to reveal the best performance of all (Race, Brown & Smith, 2005). The finding of an assessment having the above five principles informs all concerned agencies and persons so that everyone could contribute to improve quality, efficiency and equity in the education system.

The assessment of grade eight (NASA 2011), and grades three and five (NASA 2012) were respectively the first and second large scale assessments in Nepal, though several small scale assessments of student achievement have been carried out since 1995 (see BPEP, 1995; 1997; EDSC, 1997; BPEP 1998; PEDP, 1998; EDSC, 1999; CERID, 1999; EDSC, 2001; EDSC, 2003; CERSOD, 2001; EDSC, 2008; Fulbright, 2008). This assessment (hereafter NASA 2013) is the third large-scale assessment targeted to grade eight students in three subjects, viz. Mathematics, Nepali language and Science. The samples for this assessment were taken using proportional stratified sampling method with random selection of schools from 28 sample districts covering regional as well as ecological variations. Initial plan was to take total of 48000 students covering 16000 students in each subject as the sample size. When the tests were administered, the actual number of students attending the tests were slightly reduced and differed in three subjects.

1.3 Student Assessment as a Process

National Assessment of Student Achievement (NASA) in Nepal takes about two years to complete each cycle, which includes several nested cycles. The figure 1.3.1 shows general cycle as the process starting from the approval of budget and program to releasing the report.

Figure 1.3.1 Administrative cycle of NASA 2013



Government of Nepal, Ministry of Education has approved a plan for NASA until 2016. According to the plan, starting from 2011 for grade eight and 2012 in grades three and five, the National Assessment of Student Achievements shall be carried out in every second year. In line with this road map, NASA 2011 for grade eight and NASA 2012 for grades three and five have already been completed as the first cycle of large-scale assessment. As the second cycle of student assessment, NASA 2013, has been is carried out for grade eight in three subjects, viz. Mathematics, Nepali language and Science.

In order to conduct NASA 2013, Education Review Office (ERO) formed a team to coordinate NASA activities. The NASA unit under ERO works as the team to coordinate NASA activities. The subject committees for each subject have been activated during the whole process of NASA 2013. In the beginning of the process, several experienced classroom teachers, teachers from universities and Resource Persons from the DEOs in the Kathmandu Valley contributed in item writing. The task was to create sufficient items representing different sub-topics having appropriate difficulty levels and hierarchies for six versions of the

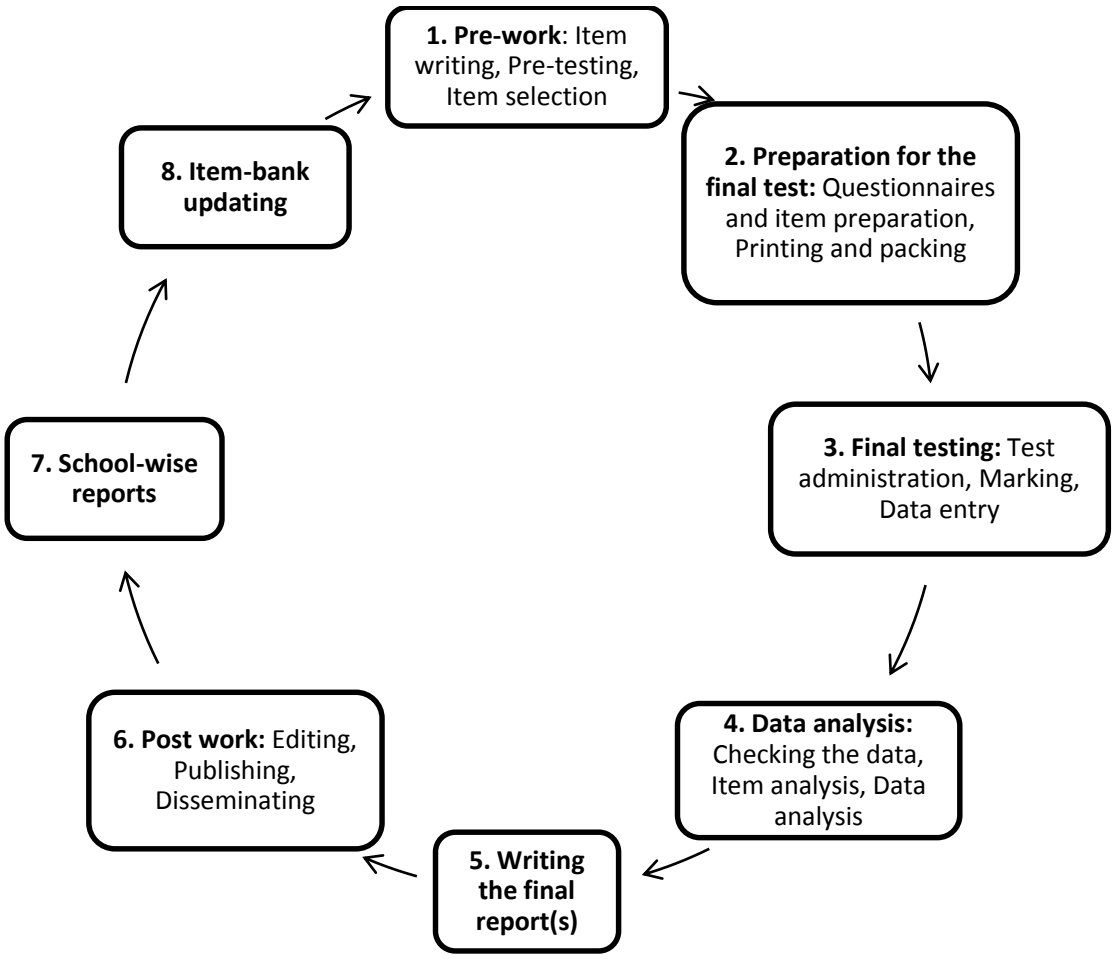
pre-test items, and ultimately for three versions of the final test items in each subject and grade. The items were pre-tested in two layers; in the first layer, pre-test was carried out to find out to what extent the language used in the tests effect the scores and in the second layer, items were tested in 130 schools in 12 districts to find the stable item parameters for the final test. Items were finalized incorporating feedbacks from pre-tests.

The final versions of test were prepared using the following six principles: (1) Curriculum based, (2) Content coverage (i.e., content validity), (3) Proper structure of various levels of the cognitive domain (i.e., ecological validity), (4) High power of discrimination (i.e., reliability), (5) Proper difficulty level, and (6) Comparability of the results with the international results, particularly, with the TIMSS and PIRLS. All the tests have high discriminatory power as the reliability of the total score is higher than α = 0.93 for all versions of tests in all subjects.

The final tests were administered in 1199 schools by a consulting firm with the support of the respective District Education Offices (DEOs) in 28 districts. The sample districts represented different ecological zones and developmental regions; rural and urban areas; and community and institutional schools. Marking of the students' test papers and data entry were done centrally by an outsourced consulting company. Data management and analysis including the equating of the test scores by using the IRT modelling was done by ERO. The preliminary result of this report was shared with various stakeholders in the last week of March 2015.

From methodological point of view, there were 8 phases in this assessment process. They are: 1) Pre-phase, 2) Preparation phase, 3) Final testing phase, 4) Data analysis phase, 5) Post-work phase, 6) Report writing phase, 7) School reporting phase, and 8) Item banking phase (see figure 1.3.2).

Figure 1.3.2. Phases of the student assessment process



The pre-phase includes the task of item writing, reviewing, pre-testing, and item analysis and selection of items for the test. At the preparation phase the background questionnaires were prepared, items and questionnaires were printed and packed, the relevant stakeholders like DEO and schools were informed and orientation was conducted for test administration. At the phase of the final testing, the test was administered at the same time in the different parts of the country in controlled circumstances, and the papers were checked based on the marking scheme so that the results are comparable all over the country. The data entry was done using OMR sheet so that accuracy and uniformity could be maintained. The data analysis phase started with the screening of the data set to find and correct the possible errors in the data. At this phase, the final item analyses through IRT modelling were done where a few poor items were omitted from the test. Finally, data analysis was done and the necessary information was provided to the report writers. The report-writing phase took several weeks to prepare a draft report. The report was

finalized, which took about two months, after sharing the preliminary findings to the stakeholders and receiving comments from relevant persons and agencies. At the phase of reporting to schools, the schools participated in the assessment could also provide feedback about their performance. At the final phase, before the next assessment round, the items and their parameters will be developed, revised and banked for later use.

1.4 Characteristics of NASA 2013

Most of the features, such as large scale and wider coverage, use of Item Response Theory (IRT), comparison with previous studies, international flavour, item banking, item analysis, and shared approach are similar characteristics as previous NASA of 2011 and 2012. There are some elements different from the previous NASA processes. One of the differences is that NASA 2013 has been administered and led technically without any international support from the beginning to the analysis and draft report preparation phase. As per requirement, the ex-consultant virtually supported personally to the NASA team in test equating, data-analysis and reporting phase. Hence, the process itself shows that the ERO is capable of carrying out such tasks.

The major characteristics of NASA 2013 can be summarized as follows:

Led by the national system

Unlike other assessments of student achievement that were conducted for grades three, five and eight by external agencies, NASA for grade eight in 2013 has been carried out and fully led by ERO. ERO personnel led the entire process-from the very beginning of test items writing and their pre-testing to the item analysis and finalization by mobilizing the subject experts working within MOE system such as from CDC, NCED, OCE, school teachers and universities. The ERO itself decided the sample making representative of all development regions, ecological belts, rural and urban locations, and types of schools.

Contrary to the first cycle of NASA 2011 for grades eight, three and five for which an expatriate led the process of data analysis and final report writing, NASA 2013 has been led entirely by the ERO personnel from preliminary data analysis and data processing to the preparation of final report writing. However, during the NASA process, some national expertise and agencies were utilized as required. Built upon the knowledge and experiences of the past two achievement studies, ERO has been found capable to lead the entire process of large-scale student assessment and to produce its result in an accepted form by maintaining standards.

Use of shared approach

Similar to the earlier NASA studies, present study on student achievement has been accomplished adopting a shared approach among national consultancy firms, school teachers and the ERO. Although the ERO personnel took a lead in the entire process, it solicited support from a national consultancy service in getting test item printed, transporting them to the examination centres, administering the tests into schools, checking and scoring answer sheets including tabulating and processing of marks in OMR sheets. Similarly, personnel from the respective DEOs were also heavily involved while administering the test and monitoring it in the schools. During the data analysis and report preparation, some support was received from national experts. Report finalization, including content editing and language editing were done with the support from external national experts. In this sense, this study is a product of collective efforts between in-house and out-source expertise at national level.

Capacity development

The exercise of NASA of this type has also provided an opportunity to build capacity of various agencies in test item construction and selection, marking and scoring answer sheets as well as tabulating the data. During the process, many teachers have been trained on writing standardized test items, selection of test items against the curriculum and international standards and developing the background questionnaires. Besides, ERO personnel have been well familiarized with the process equating test items with internationally accepted standards, correlating the achievement scores with various perspectives, and analysing the assessment results.

Large-scale assessment

Similar to the previous assessments carried out in 2011 and 2012, present study is also one of the large-scale national assessments conducted at national level with the samples of 44067 students from 1199 schools of 28 districts across the country. The sample population covers all strata: rural and urban location; community and institutional schools; Mountain, Hill and Tarai across all development regions proportionately to represent the whole population of grade eight students of the year.

This study not only assesses the students' knowledge and mastery over the contents of the subject, but also measures the cognitive ability of pupils in relation to their age and level along with making comparison with international standards. The study is not only limited to assessing the content specific knowledge and skills, it generates various information on students' family background, status of teachers,

schools and teaching learning practice in the classroom as well, establishing correlation with achievement in academic aspects.

Use of Item Response Theory

Like the earlier assessments, present study also follows the Item Response Theory (IRT) modelling in analysing test items and test construction in order to make the assessment results comparable to earlier assessments in terms of the various difficulty levels. IRT is the statistical way of equating several data sets that is essential to validly monitor performance over time and to describe growth between grades and between groups of interest. In this assessment, the data sets have been equated and compared with the results of 2011 and with some international assessment like TIMS.

Comparison with previous assessment

As stated earlier, the present NASA for grade 8 is not the first one in Nepal. So, it follows almost all procedures and methodology used in previous assessments in the entire process from item writing to data interpretation and presentation. So, its results are made comparable to the earlier ones in all respects.

Earlier NASA was conducted in Mathematics, Nepali and Social Study whereas the present NASA is conducted in Mathematics, Nepali and Science. So, except for Science, the results of Mathematics and Nepali of both assessments are exactly comparable in all respects. Furthermore, some of the items of previous assessment for each of the subjects have also been linked with present ones.

Preparation of item bank and analysis of every item

One of the unique features of NASA is establishing and updating the item banks for each subject. As new items are added every year, item bank is rich with items of varied types and difficulty level. Each item is finalized after its pre-testing in required sample students. Furthermore, the items are also drawn from the international tests like TIMSS, PISA and PIRLS linking to their respective difficulty level. The items included in the banks cover all levels of cognitive levels of varied tasks as expected by the respective subjects of national curriculum.

Validity and reliability assured

Present NASA, like the previous ones, used three parallel versions of items of the same difficulty level for each subject, equating them by IRT modelling. The reliability of the test is high and their validity is assured as they are constructed against the learning outcomes specified in the respective subject curriculum by strictly following the specification grid prepared by the CDC. The items drawn from the international test have also been linked to the test after contextualizing them into the Nepalese setting and

standard. The methodological process and procedures of the test is the same as some of the international tests like TIMSS, PIRLS and PISA. The results were linked to the set of results of 2011 assessment study and to the item banks of aforesaid international assessments.

International comparison

As this study uses some items for Mathematics and Science as linking used in TIMSS, its standard and result are found comparable to the international tests. So, its results for Mathematics and Science have been compared against TIMSS. As the tests to the extent possible attempt to give some international flavour by developing items following the standards of those tests and adopting the linking items, it is an indication that Nepal is inclined towards aligning its process and procedures in line with the international standard.

1.5 Objectives of the Study

The main objective of NASA 2013 was to find out whether the students of grade eight achieve the goals set by the national curricula in Mathematics, Nepali language and Science. Specific objectives of NASA 2013 are as follows:

- To determine the current national level of achievement of grade eight students in Mathematics, Nepali language and Science subjects;
- To determine variations in student achievement among different ecological zones, developmental regions, districts, school location (rural/urban), school type (community/ institutional schools), ethnicity and language groups, gender, and socio-economic conditions;
- To examine the extent to which the home background and other pupil related factors influence the learning achievement;
- To compare student learning achievement in the current study with that of the previous studies;
- To create reliable baseline data for the future studies.

1.6 Method and Process Used in Assessment

This assessment is carried out using mostly quantitative method based on a large-scale survey of student assessment using a standardized tool, that is, a set of questions in each subject based on curriculum of the subject approved by the government of Nepal. Set of questions were drafted by a group of qualified subject teachers and subject specialists, pre-tested in some schools and revised based on pre-test results. During the process of item writing and selection major performance, expected by the curriculum were

analysed to ensure content validity of the set of questions. While revising the question set based on pre-test, each item was analysed by determining difficulty level. As a result of pre-testing of the items, the difficulty levels of the tests were set around 50 to 60%.

Item Response Theory (IRT) modelling was used from the beginning of item construction and preparing the marking scheme for the analysis of the data, which also helped compare NASA 2013 results with NASA 2011 as well as international assessments like TIMSS. Besides, a set of background questionnaires was also used among the students in order to identify variables that influence the achievement of the students. Questionnaires were asked for teachers and head teachers about classroom and school management.

In this assessment, the school population was about 8000 and the number of schools taken as the sample was 1199, the number of sample districts was 28 and the number of students taken as the sample was 44067 in three subjects, almost equally distributed in each subject. Sampling strategy used is therefore the proportional stratified method with random selection from different strata. The following are the strata considered while selecting the samples:

i. Ecological zones (Mountain, Hill, Tarai, and Kathmandu Valley);
ii. Developmental regions (Eastern, Central, Western, Mid-Western, Far-Western, and Kathmandu Valley);
iii. Districts (75 altogether);
iv. School type (Community and Institutional); and
v. School location (Rural and Urban).

Each test paper was marked and item-wise scores were tabulated and analysed using statistical methods, descriptive as well as inferential statistics as appropriate. As the descriptive statistics, univariate analysis including distribution of scores in various categories by calculating percentages and frequencies, calculation of mean score and dispersion of data by calculating standard deviation were carried out for the analysis. Similarly, Pearson's product moment correlation coefficients were calculated to correlate various results, and mean achievements were compared using t-test as well as inferential statistics like p-value and effect sizes.

1.7 Structure of the Report

This report is structured into six chapters. Chapter one has introduced NASA 2013. This chapter mainly consists of a brief background and characteristics of the NASA 2013. The second chapter deals with the methodologies applied in NASA 2013. This chapter gives details on the methods that were applied in NASA 2013 including sample selection and determination of size; item writing and selection procedures; reliability and validity; test administration, marking scheme and data entry; tool development, analysis of the results and statistical tools used. The analysis of achievement results of each subject are detailed out from the third to fifth chapters. The third chapter covers the results of Mathematics, which is followed by Nepali language and Science subjects respectively. Each of these chapters is categorized into three sub-sections namely: basic results, diversity factors and achievement, and selected explanatory factors and achievement. The final chapter presents major findings, implications and conclusion of the study.

Chapter 2: Methodology

The National Assessment of Student Achievement (NASA) for grade eight administered in the year 2013 (hereafter, NASA 2013) is one of the largest national assessments of students administered in Nepal. Utilizing the experiences from the first cycle of NASA, that is, NASA 2011 and NASA 2012, the Educational Review Office (ERO) carried out the core functions of the assessment including sampling, developing the tools and data analysis. This section describes the methodology applied in NASA 2013. As the methodology it describes sampling methods and process, item writing and test construction process, preparation of background questionnaires, equating of the test scores over three versions used in the final testing, the statistical methods used in getting the result or findings. This section describes the common process of the core functions for all tested subjects: Mathematics, Nepali language and Science. It starts with the overall method and process used in this assessment, and describes item writing and test construction process in Mathematics, Nepali language and Science subjects. Similarly, it includes method and process of constructing background questionnaires, describes variables used in the analysis, method and process of equating the test items over three versions used in the final testing, discusses principles considered while conducting the assessment, and describes statistical methods used in analysing data.

2.1 Population and Sample

The population of NASA 2013 covers all schools of Nepal that are conducting grade eight class in 2013 and the students studying at grade eight in the same year. Therefore, the school population for this study is about 8000. In the initial plan, 1200 sample schools and 48000 students were assumed to be the sample size. When tests were actually administered, the number of schools taken as samples remained 1199 and the number of students taken as samples remained 44067 in three subjects, almost equally distributed in each subject (see table 2.1). In this section sampling method including definition of strata and sample sizes of each sample district and subject, and ecological zone are identified and justified.

Population Strata

The basic unit for sampling in NASA 2013 is school. While selecting sample schools, it was considered that the schools should be represented the country as widely as possible, and the students should be

selected in such a way that they represent the school as much as possible. In this case, sampling strategy used was therefore, the proportional stratified sampling with random selection from different strata. The following are the strata considered while selecting the sample:

i. Ecological zones (Mountain, Hill, Tarai, and Kathmandu Valley);
ii. Developmental regions (Eastern, Central, Western, Mid-Western, Far-Western, and Kathmandu Valley);
iii. Districts (75 altogether);
iv. School type (Community and Institutional); and
v. School location (Rural and Urban).

Within the ecological strata, the Kathmandu Valley was taken as a single geographical stratum as it is the most densely populated area in the country with more opportunities than other areas. From not only the population point of view, also the mixed ethnicities, weather conditions, economic activities, urbanization, as well as the dense human capacity make the Valley a unique fourth geographical area in the analysis. Hence, there were 16 basic strata in the sampling. Information such as school size, ethnic group and language used at home were also considered when selecting the samples but not as the strata. Selecting 28 districts among 75 districts with a representative number of schools from each district would automatically have considered a good coverage of both the development regions and the ecological zones. In the first phase (i.e., NASA 2011) 25 districts were selected as the sample districts. In the second phase (i.e., NASA 2012) another 25 districts together with three districts from the Kathmandu Valley were selected as the sample districts. These sample districts were selected randomly by representing each of the geographical and regional strata (see figure 2.1.1). In NASA 2013 remaining 25 districts together with three districts from the Kathmandu Valley were selected as samples. The selected sample districts are located in the given map of Nepal, and sample districts, numbers of schools and numbers of the students are presented in table 2.1.1.

Figure 2.1.1 Sample districts at NASA 2013

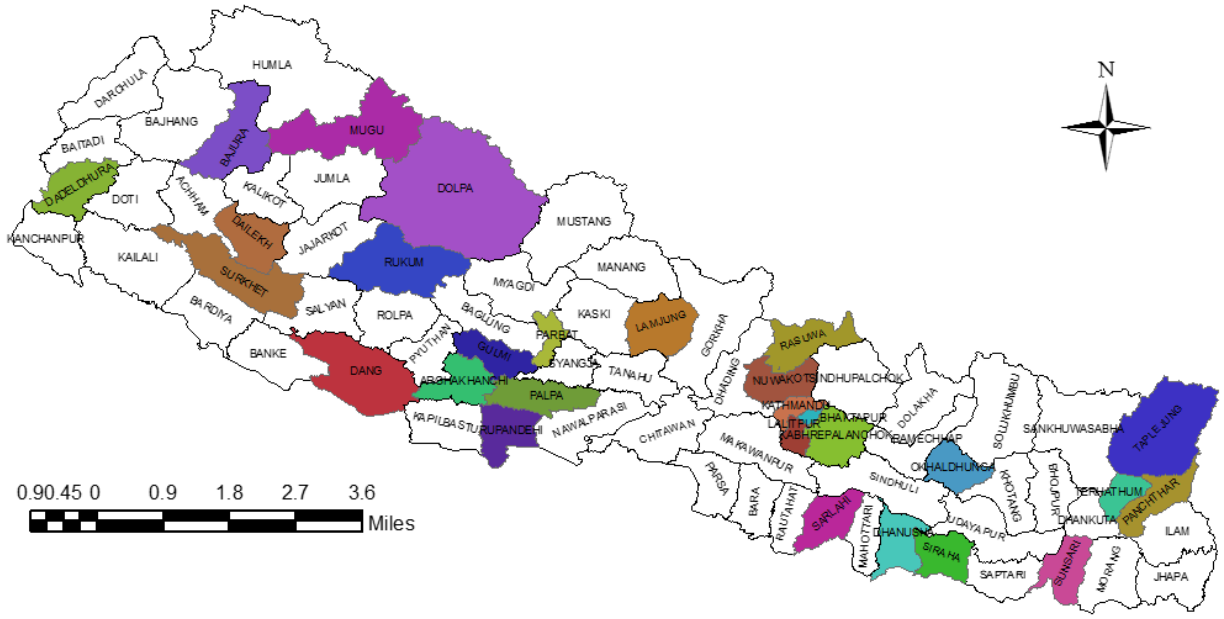


Table 2.1.1 Basic information on samples in NASA 2013 for Grade 8

Districts	Development Region	Ecological Region	Mathematics		Nepali		Science		Total	
			Schools	Students	Schools	Students	Schools	Students	Schools	Students
Arghakhanchi	Western	Hill	10	373	9	304	9	351	28	1028
Bajura	Far-Western	Mountain	8	288	8	339	8	259	24	886
Bhaktapur	Central	Valley	19	882	18	438	19	654	56	1974
Dadeldhura	Far-Western	Hill	9	325	10	394	9	254	28	973
Dailekh	Mid-Western	Hill	13	453	13	497	13	564	39	1514
Dang	Mid-Western	Tarai	17	686	17	715	17	649	51	2050
Dhanusa	Central	Tarai	9	469	8	438	9	395	26	1302
Dolpa	Mid-Western	Mountain	2	82	2	86	2	108	6	276
Gulmi	Western	Hill	15	480	15	508	15	510	45	1498
Kathmandu	Central	Valley	66	2332	66	2230	65	2352	197	6914
Kavre	Central	Hill	23	712	23	811	22	702	68	2225
Lalitpur	Central	Valley	23	755	24	914	23	681	70	2350
Lamjung	Western	Hill	10	303	10	344	10	318	30	965
Mugu	Mid-Western	Mountain	3	104	3	120	4	184	10	408
Nuwakot	Central	Hill	13	382	14	536	13	400	40	1318
Okhaldhunga	Eastern	Hill	10	320	9	310	9	267	28	897
Palpa	Western	Hill	16	602	16	577	16	487	48	1666
Panchthar	Eastern	Hill	11	363	13	595	12	371	36	1329
Parbat	Western	Hill	9	327	9	297	9	263	27	887
Rasuwa	Central	Mountain	3	97	4	101	3	104	10	302
Rukum	Mid-Western	Hill	8	321	9	376	9	426	26	1123
Rupandehi	Western	Tarai	24	1068	25	1210	24	1057	73	3335
Sarlahi	Central	Tarai	11	322	12	356	12	295	35	973
Siraha	Eastern	Tarai	8	556	11	516	9	499	28	1571
Sunsari	Eastern	Tarai	25	730	25	1142	25	863	75	2735
Surkhet	Mid-Western	Hill	16	778	16	744	16	719	48	2241
Taplejung	Eastern	Mountain	9	216	9	294	9	266	27	776
Terhathum	Eastern	Hill	7	174	6	144	7	233	20	551
Total			397	14500	404	15336	398	14231	1199	44067

Sample Size

At the second phase of sampling, the number of sample schools in each of the 28 districts was fixed proportionally based on the number of schools in each stratum. In the sample on national level student assessments, the conventional maximum sample size is less than 5% of the population (see Cochran 1977; Bartlett, Kortlik & Higgins, 2001). Based on the latest official list of schools, the number of schools with grade eight exceeded 8000. The number of sample schools in three subjects covered 1199 schools in which Nepali language test was conducted in 404 schools, Mathematics in 397 schools and Science in 398 schools. Calculating proportionally based on the number of schools within the sample district, the number of schools in each stratum was fixed, but the schools were not selected at this phase.

For convenience, all the students in the selected schools were included in the final test so that the school did not have to arrange an alternative programme. Additionally, because the number of students was not cut to a fixed number (as is a convention in the international comparisons), there is no need for using complex sampling procedures. Initially 400 schools per subject were chosen based on a rough estimation of the number of the students in the schools as discussed above. After fixing the number of schools, it was estimated that 48,000 students would attend with 40 students per class on average.

Selection of Schools

At the third phase of sampling, each of the District Educational Offices (DEOs) of the selected 28 districts provided a list of schools running grade eight to ERO. This list included information such as the number of the students, school type (community and institutional), and school location (rural and urban), which were then used as the basis for the random selection of the schools in the district.

2.2 Item Writing and Pre-testing

Item Writing

For item writing, two week-long workshops of subject teachers from schools, subject experts and university teachers were organized in each of the three subjects. A separate workshop was conducted in order to prepare background questionnaires for the students. The item writers drafted sufficient amount of items for pre-test. Five versions of test items were pre-tested, and ultimately three versions of test items were selected as the final test.

The item writers, within two weeks, drafted altogether 600 items for Mathematics, 550 items for Nepali and 450 items for Science. The number of items was too big for testing, and thus best items were selected from this item bank. After selecting the test items, all the items were translated into English and they were screened, edited and rewritten when needed. Pre-test papers were printed in the same secured manner as the examination papers.

Pre-testing

During June 2013, items for three subjects with 12 versions of test papers including four versions in each set were pre-tested in 12 districts and 90 schools by ERO with the support from the respective District Education Offices (DEOs). The districts which conducted pre-test were Ilam, Sankhuwashabha, Morang, Saptari, Dhading, Chitawan, Lalitpur, Kaski, Gorkha, Rolpa, Bajhang and Kailali. About 300 items were tested in each subject for final use. To conduct pre-test, DEOs organized an orientation session, monitored the processes in the schools, coordinated for the collection of papers and sent them to ERO for marking and data entry. The District Education Officers themselves were oriented to the process in a two-day seminar where an elementary understanding about objective testing was shared in which ERO monitored the pre-test process. To avoid the leaking of the items, all the papers were counted before and after the process so that no papers were left in schools or DEO offices.

Calibration of International Achievement Test Items

On top of the main pre-test held in June 2013, four other smaller scale pre-tests were administered during August and September to acquire the item information about the released items from the Trends in International Mathematics and Science Study (TIMSS) in Mathematics and Science, and of the Programme for International Student Achievement (PISA) reading items in Nepali. The item parameters based on the Item Response Theory (IRT) of the TIMSS and PISA items were released, but it was important to collect other relevant information, such as classical item parameters and the time on task of the items for the final item selection. Some items were taken from PISA and used in NASA 2013, however, the parameter of those items used in Nepali were not released (OECD, 2009. p.16). Thus, in Nepali subject calibration of NASA result with PISA was not done but in Mathematics and Science subjects calibration was carried out with TIMSS.

2.3 Principles of Item Selection for Final Test

From the pre-tested set of items, three sets of items were selected as the final test items. Six basic principles were considered while selecting items for the final tests. These six principles were as follows: (1) Curriculum Based (2) Content coverage (content validity), (3) Proper structure of cognitive levels of the cognitive domain (ecological validity), (4) High test discrimination (reliability), (5) Proper difficulty level, and (6) the comparability of the results with 2008 results and with the international results (TIMSS and PISA).

Curriculum Based

In this national assessment, the main idea is to test how well the objectives expressed in the national curricula are fulfilled. Those documents were used as the main criteria for the validity of the tests. Specification grids (or Table of Specifications) were prepared for each subject based on the curricula. In these grids, the time spent on the tasks in the curricula was operationalized as percentages for each topic and sub-topic. This information was used as a basis in item writing and item selection so that the marks on the tests will be proportional to those percentages in the grid.

Content Coverage

From the view point of content validity, items were selected to cover wide range of topics as for as possible. All main topics are covered by the specification grid. However, the tests were not that long to make it possible to cover all the sub-topics. A sub-test length of 3 to 4 items may be taken technically as a minimum length to discriminate the test takers from each other sufficiently. Thus, an attempt was made to include as many subtopics as possible in the tests. The selection was, however, proportional. When there were sub-topics of wider coverage under a particular topic in the curriculum than the other topics, more items were selected from those areas in test constructions.

The content coverage was widened using three different versions in the final testing in each subject of both grades. The three versions of tests were linked to each other by several linking items and to the international standards with the use of linking items from the TIMSS and PISA question banks.

Proper Structure of Cognitive Levels

The Bloom's taxonomy of the cognitive domain was used as the basis of cognitive levels. Bloom's original classification (Bloom et al. 1956; Metfesser, Michael & Kirsner 1969) of Knowledge, Comprehension, Application, Analysis, Synthesis, Evaluation was shortened into four categories: Knowledge, Comprehension, Application, and Higher Skills. There is always a risk that objective tests are measuring merely the rote memory or the recalling type of knowledge rather than higher level of thinking. At the phase of item writing, the decision was made to gear the final tests towards comprehension and application type, rather than knowledge or comprehension. This also matches with the international practice in TIMSS and PIRLS tests.

High Test Discrimination (Reliability)

Two technical areas related to high-test reliability were addressed in the test. These were: item discrimination and item difficulty. Two main item parameters, that is, item difficulty and item discrimination, classically estimated by using the proportion of correct answers (p) and the item-total correlation (ρ_{gX}), are interrelated so that the item discrimination is the highest when the difficulty level is around 0.50. When knowing that the variance of the dichotomous item is strictly related with the item difficulty, that is, $s^2 = p\,(1-p)$, the classical formula of Alpha reliability can also be expressed with these two indicators as follows:

$$rel = \frac{k}{k-1}\left[1 - \frac{\sum_{i=1}^{k} s_i^2}{\left(\sum_{i=1}^{k} r_{it} s_i\right)^2}\right]$$

where k = number of items s_i^2 = variance of the scores on item i and s_i = standard deviation of the scores on item i $r_{it} = \rho_{gX}$ = item test correlation

It is noteworthy that there are only two sources of information needed to estimate the reliability of the test: item discrimination (r_{it}) and item variance (s_i^2). It is also noteworthy that the Alpha reliability is maximized when the sum of the elements r_{it} and s is the highest. The variance is the highest when the proportion of the correct answer is $p = 0.50$, it makes sense why it is wise to select items with 1) as high item discrimination as possible and 2) mediocre difficulty level in the test. Generally, the values for reliability lower than $\alpha = 0.70$ were not taken as accurate enough to accept when comparing the scores of different groups with each

other. On the other hand, the values higher than $\alpha = 0.60$ can be accepted for a new instrument (see Nunnally, 1978; DeVellis, 1991; Hair *et al*., 1998). However, the boundaries are not strict (see Knapp & Brown, 1995).

Proper Difficulty Level

In the pre-test phase, it was noticed that many items were too difficult to measure the average achievement level in students' population. Generally, the pass rate for these items was less than 50%. Especially in the Mathematics test, higher number of students did not even start to do the open-ended questions, which lowered the proportion of correct answers. This means that the pre-test versions were able to discriminate the best students from each other, but they were not able to discriminate the lowest performing pupils from each other. From the national assessment viewpoint, the latter would be important. Lord (1952) calculated what the average facility level should be in order to gain the maximal discrimination for the test. According to his calculations, this maximally discriminating test is achieved when the percentage of the correct answers is,

- 50 % in the Completion and Short-answer type items,
- 70 % in the Multiple Choice type items with five options,
- 74 % in the Multiple Choice type items with four options,
- 77 % in the Multiple Choice type items with three options,
- 85 % in the Wright/Wrong and True/False type items.

There are some other boundaries while selecting the items, for example, the most balanced test for the national student assessment is assured when the items are selected from the whole range of ability (see Mehrens & Lehman, 1991). Thus, there should be easy items, mediocre items and demanding items on the test. These kinds of items can discriminate the best students and the poorest students as well as the mediocre students. One possible solution is to select the items so that,

- 10 % of the items should be very easy,
- 20 % of the items should be quite easy,
- 40 % of the items should be of medium difficulty,
- 20 % of the items should be quite demanding,
- 10% of the items should be very demanding.

Combining the principles, by selecting the easiest items from the pre-tests, the aim was to raise the average difficulty level in the tests near $p = 0.60$ that is, 0.50 + 0.10 (0.10 could be added for guessing in the multiple-choice questions) or even higher when possible. This is the way to construct a test, which could discriminate not only among the mediocre pupils but also among the highest and lowest performing pupils. Item selection was made based on the classical item difficulty parameter p though the IRT parameter β would have been more accurate.

Comparability with the NASA 2011 Results and International Results

The fifth principle in item selection was that the results should be comparable with 2008 results and with international test results. New concept used in NASA 2011 onwards is the use of IRT modelling where marks are given in whole number. The logic for this is that if there is not enough ability to give the correct answer, it is more probable that they will fail the item. Using IRT, any two tests can be made comparable with the help of linking items regardless the length and parallelism of tests. One possibility is to compare the Nepalese results with the international test results. TIMSS and PISA communities have released some items and their parameters and it makes sense to use those international items as the baseline in the comparison (TIMSS, 2007; 2009a; 2009b; Adams & Wu, 2000; PISA, 2006c; 2009). The idea is that when one knows the difficulty parameter of the international items, those values can be fixed in the datasets in Nepal, and thus the local items are calibrated onto the same scale as the international items are. Pre-tested TIMSS Mathematics and Science items were taken from the year 2007 and PISA. In Nepali, reading comprehension texts and items were taken from the years 2000 and 2006. As the entire item parameters were not published in PISA reports, the parameter of the chosen items were not release and therefore comparison of Nepali language results with PISA were not possible.

2.4 Final Tests

The general process followed to develop final tests in the three subjects - Mathematics, Nepali language and Science- are the same. Along with general process, some variations occurred appeared during the process of finalization of the tests in these subjects as the differences may appear due to the nature of subject. The process involved in producing final tests in each subject is discussed separately in the following paragraphs.

Mathematics Test

Mathematics assessment of grade 8 in NASA 2013 is the second cycle of assessments since 2011, and therefore, it is similar to NASA 2011 Mathematics test in nature, several items from NASA 2011 test were used as linking items between the years 2011 and 2013. With these linking items it is possible to assess what kinds of changes have happened between the two measurement points.

Several published items from the international bank were borrowed in order to compare the results in Nepal with an international standard. By using the Item Response Theory (IRT) modelling, the 2011 dataset, TIMSS and the NASA 2013 dataset were linked together to give unique information on the achievement level in Mathematics.

In the process of testing the Mathematics achievement of eighth graders, 14,500 students, 397 schools, 397 head teachers and 397 teachers from 28 districts participated in the survey. Three test versions, M_1, M_2, and M_3 were administered simultaneously in all sample districts in the same classroom.

In NASA 2013, IRT modelling was used from the beginning of item construction and preparing the marking scheme for the analysis of data. In practice, there were certain restrictions on the marking of the papers. These restrictions are: Firstly, no decimal number scores are allowed in the assessment during the marking of answer sheets. If the students are not qualified to secure full score, 0.5 score is not provided in any case i.e., students' responses are marked in whole numbers. Secondly, the marking scheme has to be more rigorously prepared because of the need to make exactly the same judgments in the years to come with the linking items. Thirdly, IRT modelling requires that all the possible marks have to be observed in the dataset; this made some difficulties when analysing some of the productive items because of the markers' tendency of not giving full marks to the students. Finally, IRT modelling requires a linking procedure between the different versions in the test. Hence, the common items for each test version, that is, the linking items were carefully selected from the NASA 2011 test booklet, TIMSS dataset, as well as the pre-test items banked for item selection. Classical item and test analysis methods were used in the pre-test phase to find the percentage of correct answers, item discrimination power and the test reliability whereas IRT was used for item calibration, finding the latent ability, comparing and equating the three versions M_1, M_2 and M_3, comparing and equating the dataset from the year 2011 and TIMSS database. SPSS software was used for classical analysis, and One Parametric Logistic Model (OPLM) software was used for IRT modelling.

In Mathematics versions were not parallel and among the three versions M_2 was the longest one and M_3 was more difficult version than the other two. All the versions were linked with each other by the use of the identical linking items. The longer version M_2 was scored out of a maximum of 74 marks whereas a maximum score of M_3 was 73 and that of M_1 was 72. There were 8 test items from NASA 2011, and 11 test items were the linking items all over the three versions. The parameters of the test items were fixed during the item calibration so that all the test items of the years 2013 and 2011 were calibrated in the national scale. Hence, at first phase the comparison was done at the national scale. Further, the parameters of the 12 TIMSS items were fixed in the item calibration and all the test items of the years 2013 and 2011 were calibrated in the international TIMSS scale to make comparison with the international level.

After the calibration of the items, all the scores in the versions M_1 to M_3 were transformed into the same scale, that is, the scores were equated. This means that all the scores in each test version are made comparable; the different difficulty levels of the tests have been modelled by using IRT modelling. The original output is the latent ability (Theta, θ), which is a standardized normal score ranging usually from – 4 to + 4. These values in each test versions were later transformed to equated scores and further the equated scores were converted into percentage. In this report, mark/score or average/mean score refers to the percentage of the scores ranging from 0 to 100. Table 2.4.1 shows the average marks calculated for three versions, M_1, M_2, and M_3, separately.

Table 2.4.1 Comparison of the characteristics of Mathematics test versions

Version	N	Max. Marks	Original score	Equated score*	Mean	SD	CV
M_1	4,915	72	26	35		23.1	66.7
M_2	4,847	74	26	35		24.4	69.3
M_3	4,738	73	25	34		22.8	67.9
Total	14,500			35		23.5	68.0

* Percentage of the equated maximum score rounded in whole number.

The test items were classified into five categories: *Algebra, Arithmetic, Geometry, Sets,* and *Statistics.* The number of items and their weightage were based on the recommendations of the curriculum. Overall internal consistencies, given by the reliability of the test, of the whole tests on each version were very high ($\alpha > 0.90$); however, some of the categories, for example, Sets and Statistics, contain a few items and hence the reliability could be relatively low. The overall summary of the content wise test analysis is given below.

The reliability of the score in the total sample cannot be given in a classical way because it can be estimated only in specific version of test.

Table 2.4.2 Validity and reliability of the scores

Topics	Marks			Percentages			Percentage in	Reliability		
	M1	M2	M3	M1	M2	M3	Curriculum	V1	V2	V3
Total	72	74	73	100	100	100	100	0.92	0.93	0.93
Algebra	16	14	22	22.2	18.9	30.1	30	0.73	0.74	0.77
Arithmetic	21	16	15	29.2	21.6	20.5	20	0.74	0.75	0.69
Geometry	24	24	24	33.3	32.4	32.9	30	0.81	0.81	0.83
Sets	8	12	5	11.1	16.2	6.8	10	0.40	0.72	0.57
Statistics	6	8	8	8.3	10.8	11.0	10	0.46	0.58	0.56

Nepali Language Test

Nepali language is a compulsory subject in grade eight. Due to the additional 30 to 45 minutes time required for reading comprehension in Nepali language the tests took half an hour more to complete the test in comparison to other two subjects, that is, it takes almost 2:30 hours. Average difficulty level of item was 55% in different versions. The total marks of each version was 79, which was the same in all versions, but in the final analysis one item from version 1 and 1 item from version 3 were omitted. Among versions 1, 2 and 3, there were 14 linking items. Similarly, six items from PISA and 17 items from 2011 item set were used in the test, which was sufficient to equate the 2013 results with the results of 2011.

Nepali language gives 65% of the total time for writing, only 10% for reading and the remaining 25% for grammar and vocabulary. Due to unavailability of good items from pre-test, challenging task was to maintain the curricular weightage. There were not enough varying texts in the pre-tests and thus some international texts (released PISA texts) and related items were selected to cover different text types such as, short opinion type of texts, short letters, medium length informative texts, long informative texts, and long narrative texts. These items were pre-tested with a small scale. Due to the nature and score of test items there were variations in the weightage of various areas of contents among versions. Variations also exist between the versions and curricular weightage.

Table 2.4.3 Validity and reliability of Nepali language tests

Topic	VER1	VER2	VER3	VER1%	VER2%	VER3%	Curriculum%	Reliability[2]
Total	78[1]	79[1]	78[1]				100	0.93
Reading	22	21	24	28.2	26.6	30.8	10	0.80
Writing	32	19	29	41.1	24.0	37.2	65	0.82
Grammar	20	24	13	25.6	30.4	16.7	15	0.83
Vocabulary	4	15	12	5.1	19.0	15.3	10	0.68

1) marks 2) on weighted mean

Reliabilities of the versions were round α = 0.80, that is, the test scores can discriminate the individual pupils with high accuracy. Though there is no intention to use the tests as the examinations; the more discriminating the tests, the more accurate will be the outputs.

As in Mathematics and Science, weightage in Nepali language test was high in comprehension and application type of items (see table 2.4.4).

Table 2.4.4 Characteristics of Nepali test with various level of cognitive domain

Various level of cognitive domain	VER1	VER2	VER3	VER1%	VER2%	VER3%
Knowledge	8[1]	21[1]	17[1]	10.1	26.6	21.5
Comprehension	21	12	19	26.6	15.2	24.1
Application	37	29	24	46.8	36.7	30.4
Higher ability	12	17	18	15.2	21.5	22.8

1) marks

Science Test

NASA 2013 test is the first cycle of the grade eight Science assessment. Several published items from the international item bank were borrowed in order to link and compare the results in Nepal with an international TIMSS standard. By using the Item Response Theory (IRT) modelling, TIMSS datasets and the NASA 2013 dataset were linked together to give unique information on the achievement level in Science.

In the process of testing the Science achievement of eighth graders, 14,231 students, 398 schools, 398 head teachers and 398 teachers from 28 districts participated in the survey. Three test versions, S_1, S_2, and S_3 were administered simultaneously in all sample districts in the same classroom, taking these versions alternatively by students.

In Science the average difficulty level was around 50% in aggregate of all versions. In version 1, nine items from 2008 national assessment were used as linking items in this test. Between versions 1 and 2 there were linking items of 8% marks and in version 3, nine items were used as linking items from the TIMSS. Selected TIMSS linking items came from physics in general.

Versions 1 and 2 are parallel when comparing them in item characteristics (see table 2.4.5). Different versions include a wider variety of sub-topics under each topic/content area. Reliabilities of the sub-topics in Science tests is around 0.93 (i.e., $\alpha \approx 0.93$). This means the total test score can discriminate the individual pupils with high accuracy.

Table 2.4.5 Characteristics of Science tests with various content areas

Topic	VER1	VER2	VER3	VER1%	VER2%	VER3%	Reliability[2]
Total	67[1]	72	72	100	100	100	0.93
Biology	18	21	17	27	31	25	0.79
Chemistry	18	11	15	27	16	22	0.73
Geology & Astrology	17	15	12	25	22	18	0.72
Physics	19	26	29	28	39	43	0.82

1) marks 2) on weighted mean

As in the Mathematics and Nepali language tests, high weightage is given in comprehension and application type of items (see, table 2.4.6). Higher skills are measured mainly by the open-ended, productive type of questions.

Table 2.4.6 characteristics of Science test with various cognitive levels Science tests

Cognitive level	VER1	VER2	VER3	VER1%	VER2%	VER3%
Knowledge	13[1]	15[1]	17[1]	19	22	25
Comprehension	20	18	20	30	27	30
Application	30	34	26	45	51	39
Higher skills	4	6	10	6	9	15

1) marks

2.5 Marking Schemes

Marking scheme of each test paper was prepared through the workshops of subject teachers and experts. A marking scheme was prepared by adding specific examples on what varieties of correct answers were allowed for the marks and which kinds of answers were not to be credited by marks.

Later, in the analysis phase, it was noted that in the Science test some items needed to be discarded, and therefore two items, each carrying 1 score, were discarded from version 1. Hence, the maximum score of version 1 was reduced from 69 to 67. Similarly, in Nepali language test two items were discarded, one item from version 1 and another from version 3. In Mathematics test two items were also discarded, one item from version 1 and another from version 3.

2. 6 Final Test Administration, Marking and Data Entry

Test administration, marking and data entry were outsourced to a consulting firm. The following paragraphs elaborate the process of test administration, marking and data entry.

Test Administration

To set the final test administration, NASA focal persons of all districts (28) were oriented for two days. The contents of orientation were final test administration procedure, maintaining peace and fairness in the test, verification of sampled schools and numbers, and contacting the schools. Similarly, one day orientation for examination centre heads was conducted in district headquarter in presence of the consulting firm EDSC and ERO monitors at least one day prior to the final test administration. The final test was administered on the same day throughout the country which was on March 13, 2014. The administration was outsourced to

the consulting firm EDSC in collaboration with the respective District Education Offices in which the officers from the ERO monitored the process. Test arrangement was based on one subject each school/student, mostly different versions to the students sitting together.

Marking and Data Entry

The marking of the papers as well as the data entry was outsourced. Altogether 44,067 students' papers, around 1,200 teachers' questionnaires and 1,200 head teachers' questionnaires were collected into an "examination centre" in Kathmandu. Within three months of April, May, and June 2014, the outsourced consulting firm completed marking the papers. Scoring was done based on approved marking scheme in order to maintain marking uniformity among the markers. Students' and teachers' responses were filled in Optical Mark Reader (OMR) sheets and scanned for data entry. Consulting firm EDSC submitted the dataset in Excel as well as SPSS format.

2.7 Background Questionnaires

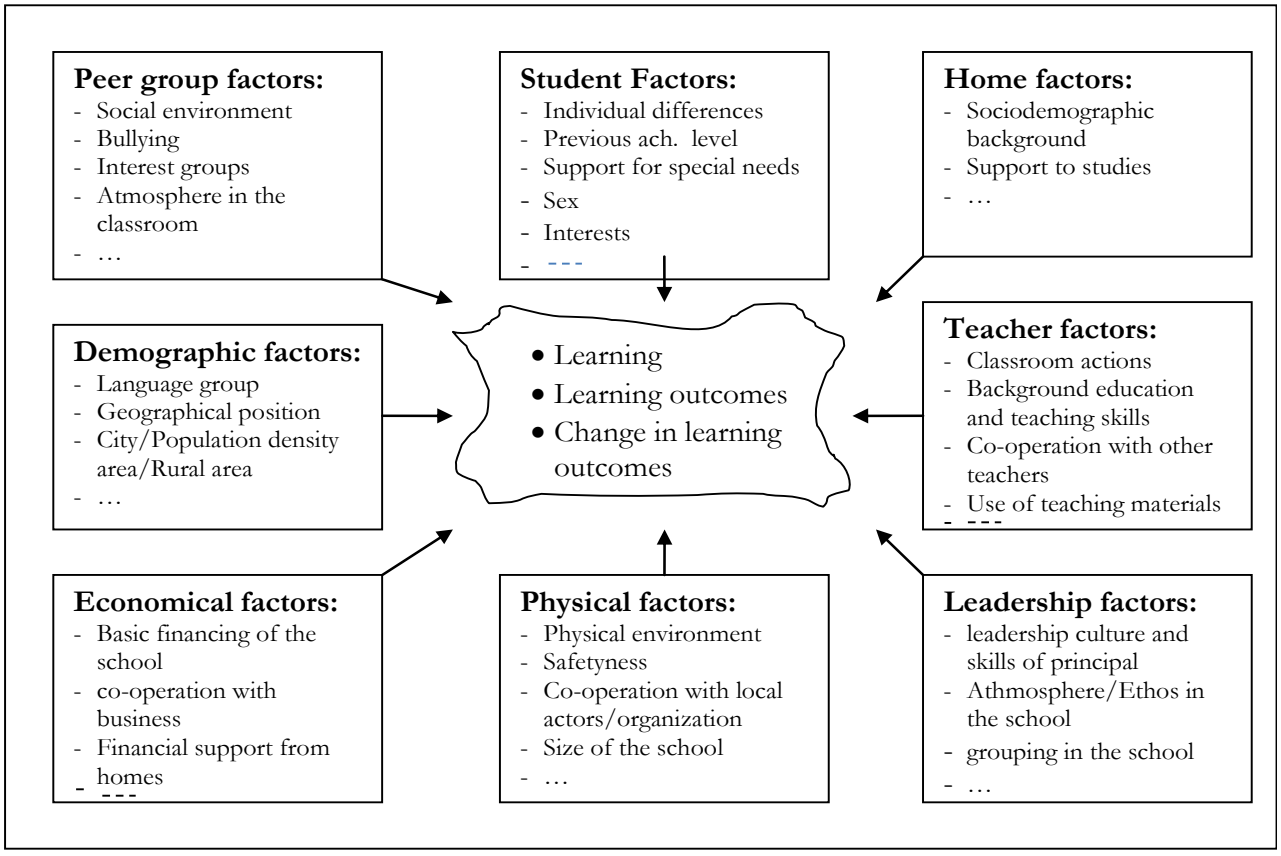
In addition to assessment questions, a set of background questionnaires including the relevant contextual information about the students, students' family as well as the school and teaching were prepared for the students. There are several indicators which may be valuable to measure during the student assessment process. The question relevant to ask was 'how to make the selection of such influential variables?' In order to select such variables, a conceptual framework was developed based on the previous studies and relevant literature. Background questionnaires were developed to represent the selected variables.

Conceptual Model for the Background Questionnaires

A sketchy modelling of this complex phenomenon of learning is in use in the Finnish National Board of Education (FNBE, Metsämuuronen, 2009). The same model, with contextual modification, was used as the basis for compiling the background questionnaires (see fig. 2.7.1). The idea in the model is that the main factors explaining the learning outcomes of the individual students are the *student factors*: motivation, attitude, working habits, and so on. Other influential variables are the *family factors*: SES, support to the studies, literacy in the family and so on. Third set of related factors near the pupils are the *peer group factors*: social support for studies, bullying, atmosphere in the classroom, and so on. All these matters were asked straight to the students and their answers were solicited accordingly . *Teacher factors* evidently play some role in student achievement; however, usually much less of them are expected. Such factors as

classroom activities, teaching skills, use of teaching materials, for example, may affect student achievement. Questions about the teacher related factors were asked to the teacher and their responses were solicited.. School factors can be divided into two groups: *managerial factors* and *physical factors*. These factors, such as atmosphere in the school, the condition of the school premises, safeness and so on, were asked from the head teacher or principal. *Economic factors* and *demographic factors* were found in national statistics. In NASA 2013, the demographic factors are part of the sampling scheme. Figure 2.7.1 illustrates these factors.

Figure 2.7.1 Conceptual framework for the background information (adapted and contextualized from Metsämuuronen, 2009; NASA 2011)



Selection of Questions

The variables for the background questionnaire were selected in several phases. At the first phase, background items were screened and updated with both TIMSS and PISA questionnaires (PISA 2003a; 2003b; 2006a; 2006b; TIMSS, 2003; 2006) as well as NASA 2011. In the second phase, a discussion was

organized to finalize the background variables. The structure of most of the questions was based on NASA 2011 questionnaire (see NASA 2011 report for detail).

2.8 Specific Variables Used in Analysis

Three sets of variables were handled carefully in order to fully understand the results while analyzing the results: the concept of equated scores, Fennema-Sherman attitude scale, and the indicator for socioeconomic status (SES). Three different versions were used in student achievement, the scores are not automatically comparable, and hence, the scores have to be equated before the analysis. The logic and procedure of these equated scores are further described in the following paragraphs of this section. The Fennema-Sherman test is a widely used test (e.g. in PISA and TIMSS questionnaires) to measure attitudes towards school subjects, which was handled as mentioned below in this section. The socioeconomic variables were used to explain the achievement results and explain the differences between the students' achievement. The motivation and construction of the SES variables were handled as described below in this section.

Equated Scores and IRT Modelling

The final tests were constructed so that a certain amount of identical items, representing different content areas, linked the tests with each other. Thus, it was possible to equate the test scores with IRT modelling (Rasch, 1960; Lord & Novick, 1968; Lord, 1980; Hambleton, 1982; 1993; Béguin, 2000) and finally to acquire the comparable latent ability of each student over the different versions. IRT modelling is the tool for equating test scores in the well-known international comparisons of PISA and TIMSS studies. The testing procedure used in NASA 2011 and 2013 typically involve this kind of complex process facilitated by IRT modelling. IRT modelling is the only credible way to assess on how the achievement level is changed from 2011 to 2013, and to compare the results in Nepal with the national as well as international standards such as PISA and TIMSS results.

The need for equating comes from four facts. First, to widen the number of items so that the range of topics and sub-topics will be covered in testing. For this purpose, it is preferable to use several versions of tests in testing. In NASA 2013, three versions of tests were used. Second, in order to compare the results of 2011 and 2013, IRT modelling is the most accurate method in order to perform the comparison when equal (or parallel) tests are not used. Third, to compare the national and international level results IRT modelling is necessary because the only knowledge available of the items was the IRT difficulty parameter β.

Additionally, using IRT modelling it is possible to use tests of unequal lengths and make tests free from leaking of the items as well as unfair behaviour among the students. When the tests were of unequal length and deliberately of somewhat different difficulty level, IRT modelling is actually the only sensible method of making the final scores comparable.

Equating the test scores with IRT modelling was administered with the following principles and practices. The scores are transformed into the same scale on the basis of characteristics of IRT models that the latent ability level of a learner (θ) and difficulty level of an item (β) are identical when certain pre-conditions are met (see Wright, 1968). The latent ability level for each pupil can be determined in the same metric for every test as far as there are the linking items connecting the versions. The estimation was run with OPLM program (Verhelst, Glas & Verstralen, 1995). A brief technical description of the equating process is as follows (see Béguin, 2000, 17–36):

i. Define the structure of the test so that the linking items are connecting the tests with each other. The values of difficulty parameter of the linking items are exactly the same in each version because the difficulty levels of all other items are calibrated into the same scale as the linking items are.
ii. Use *Conditional Maximum Likelihood* (CML) procedure to estimate the difficulty level (β parameter) for each item.
iii. Use *Marginal Maximum Likelihood* (MML) procedure to estimate the distribution of each student's latent ability (θ parameter) in each version.
iv. Estimate the θ parameter of the scores of each version using means and deviations of distributions of β and θ. This results in a unique latent value, however measured in a common scale, for each observed value of the scores in all versions.

The success of equating depends on three things. First, the linking items should represent a sufficient range of ability level; too easy and too difficult items should, however, be avoided. Second, the linking items should represent a short test inside the test; the items should cover the different content areas as widely as possible. Third, the stable parameters in the equating process are dependent on the sample; the better the sample represents the target population the better the calibration corresponds with the population parameter. Though the item parameters are to some extent vague, the results are much more accurate than if only the classical matrices (the proportion of correct answers) were used in comparison.

Normally, in the equating of the test scores, an average student with average ability would get Theta value zero ($\theta = 0$). The better the student is, the higher is his/her θ above the zero line and parallel, the weaker the student the lower is θ below the zero line. However, when borrowing the items from the international item bank, their difficulty level is calibrated to that international level where an "average international student" would get the zero for Theta. All the new items written in Nepal will be calibrated into that international scale and hence the mean of the "average Nepalese student" does not get the value of zero but either above or below. This makes it possible to assess the achievement level of Nepalese students compared with the international standard.

At the final phase, Mathematics and Science scores were calibrated into the TIMSS scale. The original scores are transformed into the equated scores and the equated scores are changed to the percentage of maximum score. Hence, the score 100 means that the student was able to solve all the tasks and obtained the maximum marks.

Fennema-Sherman Attitude Scale

A shortened version of Fennema–Sherman Attitude Scales (FSAS, Fennema & Sherman, 1976) are used in several international comparisons, like in Trends in International Mathematics and Science Study 2007 (TIMSS, Mullis, Martin, & Foy, 2008) and its predecessors 1995, 1999, and 2003 as well as in Programme for International Student Assessment (PISA). Original scales include nine dimensions but in these international comparisons only three dimensions with four items on each (see table 2.8.1) and two negative items on each of the first two dimensions are used. The names of the factors can be "Liking Maths", "Self-Efficacy in Maths", and "Experiencing utility in Maths" (see Metsämuuronen 2012a, 2012b; compare naming in, e.g., Kadijevich, 2006; 2008). Metsämuuronen (2012a; 2012b) noted that this kind of "expected factor structure" can be found in all Western countries including European countries (except in Bulgaria and Romania), Australia, Canada, Israel, the United States and Russia. However, after performing exploratory factor analysis (EFA) with Principal Axis Factoring, 3 factors, and Promax rotation with Kaiser Normalization separately in all countries, it is notable that in several countries, this structure cannot be found.

Table 2.8.1 Expected Factor structure of FSAS in North America in TIMSS 2007 *(Pattern Matrix[a])*

	Factor[b]		
	1	2	3
*Math is more difficult for me than many of my classmates	-.804[b]		
*Math is not one of my strengths	-.778		
Usually I do well in Math	.768		
I learn things quickly in Math	.725		
*I hate Math		-.878	
I enjoy learning Mathematics		.865	
*Math is boring		-.740	
I would like to take more Math in school		.615	
I need Mathematics to get into the <university> of my choice			.706
I need to do well in Math to get the job I need			.698
I need Math to learn other school subjects			.555
I think learning Mathematics will help me in my daily life			.552

a. Extraction Method: Principal Axis Factoring. Rotation Method: Promax with Kaiser Normalization. Rotation converged in 5 iterations.
b. Values > .30 are seen

Students from Nepal responded on some items of "liking science", "self-efficacy in science" and "experiencing utility in science". In this test it is not found structured as exactly as it was expected. However, it is closer to the structure in three components, which is given in the table 2.8.2.

Table 2.8.2 Unstructured factor structure in NASA 2013 (Nepal)

Pattern Matrix[a]

	Component[b]		
	1	2	3
q17a FS: I think learning Science will help me in my daily life	0.55		
q17b FS: I need Science to learn other school subjects	0.53		
q17c FS: I need to do well in Science to get into the <university> of my choice	0.63		
q17d FS: I would like a job that involved us in Science	0.55		
q17e FS: I need to do well in Science to get the job I need	0.62		
q18a FS: I usually do well in Science			0.55
q18b FS: I would like to take more Science in school	0.58		
q18c FS*: Science is more difficult for me than many of my classmates	0.61		
q18d FS: I enjoy learning Science	0.54		0.51
q18e FS*:Science is more difficult for me than many of my classmates		0.75	
q18f FS*: Science is not one of my strengths		0.81	
q18g*: Sometimes, when I do not initially understand a new topic in Science, I know that I will never really understand it		0.75	

a. Extraction Method: Principal Component Analysis
b. Three components extracted, based on Science database

In NASA 2013, although response came into three components, it is interesting to note that '*Science is more difficult for me than many of my classmates*' comes in the same component as it is with positive attitude. Instances of similar type could also be found in case of other subjects.

SES Variables

According to Bradley and Corwyn (2002) socio-economic status (SES) is one of the most studied constructs in social science. The construct seems to have interested the researchers because of the belief that high SES families provide for their children an array of services, goods, parental actions, and social connections that potentially rebound to the benefit of children and a concern that many low SES children lack access to the same resources and experiences, thus putting them at risk for developmental problems (see Brooks-Gunn & Duncan, 1997). Specifically, SES matters because it has been related to health and life outcomes for as long as social groups have existed (Oaks, 2011), and it has been shown to have a strong connection with cognitive and academic attainment (see for example, Bradley & Corwyn, 2002; APA, 2011).

In the literature, social status is commonly conceptualized in terms of socio-economic standing, taking into account the various combinations of income, education, and occupation (APA 2007, 5). The challenge in measuring SES is that there has not been a complete consensus on precisely what it represents (Liberatos, Link & Kelsey, 1988; McLoyd 1997), the economic position or social status, and hence there is not a single measure for SES (Bradley & Corwyn, 2002; APA, 2007: 5).

Although general consensus is that income, education and occupation together represent SES better than any of these alone (White 1982), there is no consensus on: (a) how best to composite the set of indicators; (b) whether it works the best to examine relations between SES and child outcomes using a composite or single indicator at a time; or (c) how best to measure each component (Krieger et al. 1997).

Hence, as there is no consensus of the measures for SES, NASA team developed indicators relevant to Nepalese context. The following seven indicators of SES were selected into the final SES indicator:

- Father's education
- Mother's education
- Father's occupation
- Mother's occupation
- Home possessions (e.g. "*Do you have* a dictionary at your home?")
- Home accessories (e.g. "*How many* mobile phones do you have in your home?")
- Attending the private school (yes or no)

Because the variables were of different scales (from nominal to ordinal scales) and because of incomparable scores (from 0 – 1 to 0 – 12), all the variables were re-scaled first to fit with each other. At the first phase, the variables were analysed with respect to educational outcomes. Decision Tree Analysis (DTA), the data mining tool in SPSS software, and ANOVA, the basic tool for analysing the differences between the groups mean were used to find the best classification of each variable with regard to the statistical differences in learning outcomes. At the second phase, five variables comprising the home accessories were dichotomized on the basis of DTA and ANOVA and summed up. At the third phase, all seven variables for SES were dichotomized on the basis of DTA and ANOVA. Hence, all the variables, regardless of their original scale, were scaled as 0 or 1 where 1 indicates the higher SES. This makes all the individual indicators equal weighted. At the final phase, seven indicators were summed up as the final SES indicators. The indicators and their cut-offs are presented in table 2.8.3.

Table 2.8.3 Indicators of SES in NASA 2013

Variable	cut-off[1]	Effect on total score[2] (for example in grade eight Science)
Father's education	less than SLC = 0, other = 1	+14% points, η^2 = 0.08
Mother's education	less than SLC = 0, other = 1	+14% points, η^2 = 0.052
Father's occupation	Agriculture = 0, other =1	+10% points, η^2 = 0.053
Mother's occupation	Agriculture = 0, other =1	+10% points, η^2 = 0.049
Home possessions	5 or less out of 12 possessions = 0, 6 or more = 1	+9% points, η^2 = 0.040
Home accessories		
Mobile phone	1 to 3 = 1, none = 0	+ 8% points, η^2 = 0.01
Television	1 to 3 = 1, none = 0	+12% points, η^2 = 0.056
Computer	1 to 3 = 1, none = 0	+12% points, η^2 = 0.075
all together	0 out of 3 = 0, other = 1	+14% points, η^2 = 0.019
Attending to private school	no = 0, yes = 1	+23% points, η^2 = 0.236
Total SES		+35% points, η^2 = 0.209

1) based on NASA 2011 2) one-way ANOVA/ univariate GLM

2.9 Statistical Methods Applied for the Analysis

Analytical Tools Used for the Statistical Analysis

Basic statistical methods used to analyse assessment results includes tools of statistical description such as means, standard deviations, percentages and frequencies; correlations such as Pearson's product moment correlation coefficient, and comparison of two means such as t-test as well as statistical inference like p-values and effect sizes. These methods are described in all standard books of statistical description and inference (e.g. Metsämuuronen, 2013). Analyses are done using SPSS software. The Analysis of Variance (ANOVA) and Covariance (ANCOVA) are used in the General Linear Modelling (GLM) when several means are compared. All the p-values are corrected by using Multi-level modelling (Goldstein, 1986) or Hierarchical Linear modelling (Bryk & Raudenbush, 1987) by using SPSS Linear Mixed models module. In some cases (for example, when explaining the schools' average student achievement level with SES variables), traditional Linear Regression Analysis and Logistic Regression Analysis are also used.

A set of methods called Decision Tree Analysis (DTA) are used in some cases when willing to find the best predictors of achievement out of hundreds of possible meaningful variables. DTA is one of the methods used in data mining; it is very effective when it comes to finding statistically the best groupings of the independent variables.

Some Statistical Concepts Used in the Text

Within the text, three important concepts should be understood. Statistical significance, that is, p-value which refers to the possibilities to generalize the result to the population. Practically speaking, behind the p-value (from "probability") is the fact that measuring human mental processes – as learning outcomes or attitudes are – there is always a measurement error. This means that the result of each individual student as well as each mean score carries error. Especially, when the population is examined by using a sample, all the means carry both measurement error and sampling error. In the sample, there can be a small difference between the boys and girls, for example. The p-value tells us how probable the same result could be in the population as a whole. If the probability is $p < 0.05$, this means that the difference would be found at risk of 5% only in five samples out of 100 the results differ from those obtained. If the p-value is $p = 0.002$, the risk for a faulty decision (or difference) is only 0.2%.

When the sample size is huge, like the sample of 44067 students, the p-value very easily gives a signal that the difference between the groups is real in population. P-value does not, however, tell whether the difference is small or big. For this purpose there is another statistic, Effect Size (ES). ES indicates how far the lowest and highest groups are from each other. The especially used indicators of ES are Cohen's d used for two means and Cohen's f for several means (Cohen 1988). Cohen has given boundaries for small, medium and large ESs. During the text, these boundaries are used as a "measurement stick" to indicate whether the difference is small, medium or large. The rough boundaries of the small, medium, and high ESs are collected in Table 2.9.1.

Table 2.9.1 Rough boundaries of Effect Sizes

size	Cohen's *d*	Cohen's *f*
small	< 0.2	< 0.1
medium	round 0.45	round 0.2
high	> 0.8	> 0.4

Technically, ES also gives a preliminary indication as to how well the grouping factors, such as the gender, explain the results. Hence, in the text one may read that "the difference between boys and girls is statistically significant ($p < 0.001$) but the effect size is small." This means that first, the difference between the boys and girls is real, but second, the difference is very small in reality and third, gender as a grouping variable does not effectively explain the variation in the data.

The third related concept is the explanatory power of the variable. Especially when using the Analysis of Variance (ANOVA) as an analytic tool, the output allows for the possibility to show how well the factor explains the variation in the data. The usual indicator for this is Eta squared (η^2) which is actually a correlation between a grouping variable and continuous variable. When the Eta squared equals $\eta^2 = 0.30$, this means that the grouping factor (such as the geographical region) explains 30% of the variation in the dataset. Cohen's *f* strictly uses this information:

$$f = \sqrt{\frac{\eta^2}{1-\eta^2}}$$

Hence, if $\eta^2 = 0.30$, then the effect size *f* equals $f = \sqrt{(0.3/0.7)} = \sqrt{0.43} = 0.65$, showing high effect size (see Cohen, 1988, 284). While comparing two groups, an effect size is measured by *Cohen's d* which uses the information $d = \frac{t \cdot (n_1 + n_2)}{\sqrt{df \times n_1 \times n_2}}$, where n_1 and n_2 are number of students in first and second group, df is the degree of freedom.

2.10 Data Analysis and Report Preparation

The report preparation started when the data were received from the consultancy firm. Once the data were received, the first task was the data cleaning process in each subject. During this process, the background variables and items were recoded and many indicators were also created for the analysis purpose. Data were also verified even from the OMR sheets, mainly when errors were found or data were missing.

The second step for the analysis was the preparation of database for the OPLM design and analysis. During this phase, the item characteristic curves of each item was analysed and the items having negative correlation were removed from the analysis. At the end of this phase, the version, cognitive level, content area, item types based students' ability scores (theta) were generated.

In the third phase, the three versions of students' score and latent ability (theta) generated from OPLM were used to equate all the three versions of each subject to maintain the difficulty level (created parallel version). Further equating was done on the basis of cognitive level, content areas and item types. All those equating processes were done manually in Excel application.

In the fourth stage, the equated data were merged into the original SPSS file to make a complete set of database. In this stage, all the versions and background variables (student, teacher, head teacher, and sample) were also merged and database was prepared for the analysis. In the process of data analysis various statistical results as mentioned in the previous section were calculated and the results were compared.

The data analysis and draft report preparation were carried out simultaneously. Several rounds of workshops were organized in order to analyse data and prepare draft report. A team of content editors worked on draft to finalise the contents of the report. To prepare the final draft of the report a language editor was assigned to edit language of the draft report that was edited by the content editors. To complete the whole process of data analysis, draft report preparation and finalization of report took about nine months.

2. 11 Professional Standard and Ethical Consideration

This study has been accomplished following and maintaining the professional and ethical norms in the process. This study followed the standard norms and procedures of National Assessment during random sampling, preparation of items and test administration. Test items were based on National Curriculum rather than following a particular textbook or reference book. All the data and information presented in this report were based on student achievement test and responses on background questionnaires as well as teachers' responses and head teachers' responses from sample schools. In order to make assessment technically sound and analysis and interpretation unbiased, several experts were involved during the process of this assessment. A consulting firm was hired to carry out test administration, mark answer sheets, enter the data using OMR and clean the data. Furthermore, during the report writing, a team of experts, including external experts, was assigned the task of editing the contents and finalize the report. Such external involvement helped to make assessment process and result more credible.

In addition to the above mentioned professional considerations, basic ethical standards were taken into consideration during the process of this assessment. This study respected the confidentiality and anonymity of respondents. Participation of the students and teachers were made voluntary, and no forced and paid recruitment of students and teacher was done. This study does not expose the information related to any particular student, teacher and school unless self-request is made. ERO can deliver the report of each school to the concerned schools only. Partial database can be provided to researchers on request provided that no any individual information will be disclosed. Background information and international linking items are borrowed from TIMSS and PISA accepting their condition of non-profit use.

Chapter 3: Analysis of Student Achievement in Mathematics

Mathematics is one of the compulsory subjects in the school curriculum and accorded five out of 40 credit hours a week. The present assessment is entirely based on the learning outcomes (curricular competencies) as set in approved curriculum for grade eight. As per the approved curriculum, the overall objective of grade eight Mathematics is two-fold. One of the objectives is to use mathematical knowledge and skills in solving daily problem; and the second objective is to provide students with the basic prerequisites for secondary education

This chapter presents the overall distribution of achievement. Further, the analysis is presented for achievements across various content areas of the curriculum, various levels of cognitive domain, and types of items. For establishing comparability, the analysis includes the temporal perspective of comparing with the first cycle of study i.e., NASA 2011.

In the second section, there is a discussion to make sense of the level of achievement through comparison with international studies- particularly with TIMSS, an international assessment. The third section of this chapter provides results associated with various diversity factors, including district, ecological zone, development region, school type, school location, language at home, and caste/ethnicity.

Section four of this chapter deals with the selected explanatory factors about student achievement in grade eight Mathematics. The factors that explain achievement in Mathematics are parents' education and occupation, home possessions and accessories, socio-economic status (SES) of parents, work beyond school hour, age of the student, support provided for the study, availability of textbook, homework given and feedback provided, and negative and positive activities at school. The final section of this chapter briefly summarizes the findings of the achievement in Mathematics.

3.1 Basic Achievement Results in Mathematics

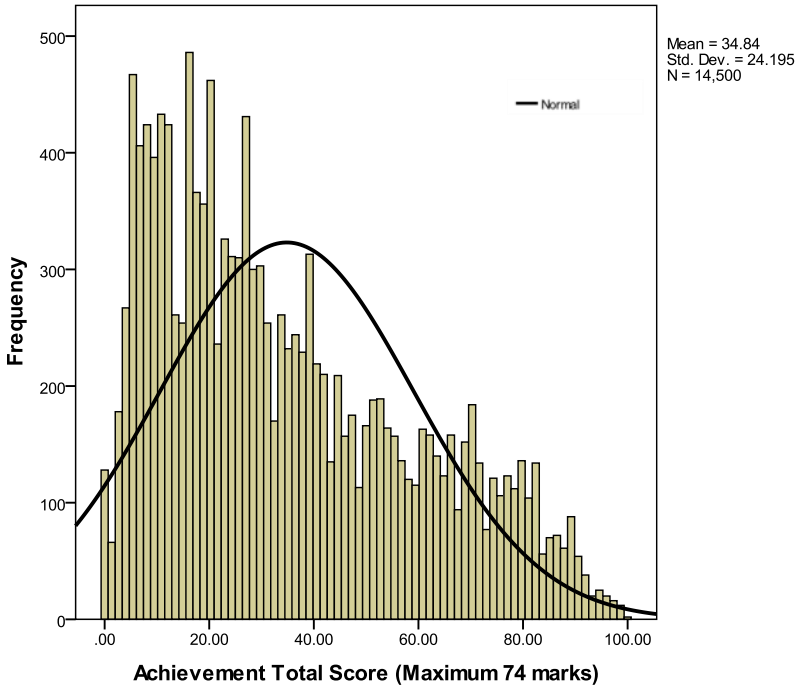
Overall Distribution of Achievement Scores

The population is usually distributed normally in a large sample in the study of student achievement. In the assessment of Mathematics, the sample was big enough to form a normal distribution, as there were 14,500

students in the sample. However, figure 3.1.1 shows that the achievement score in Mathematics is not normally distributed. Based on the distribution of achievement scores, the student population can be grouped into three types of population - namely, low performing, medium performing, and high performing, although the majority of the students fall in the low-performing group.

The low-performing students achieved 20–25%, the medium-performing students achieved 40–50% and the high-performing students achieved 70–80% of the score.

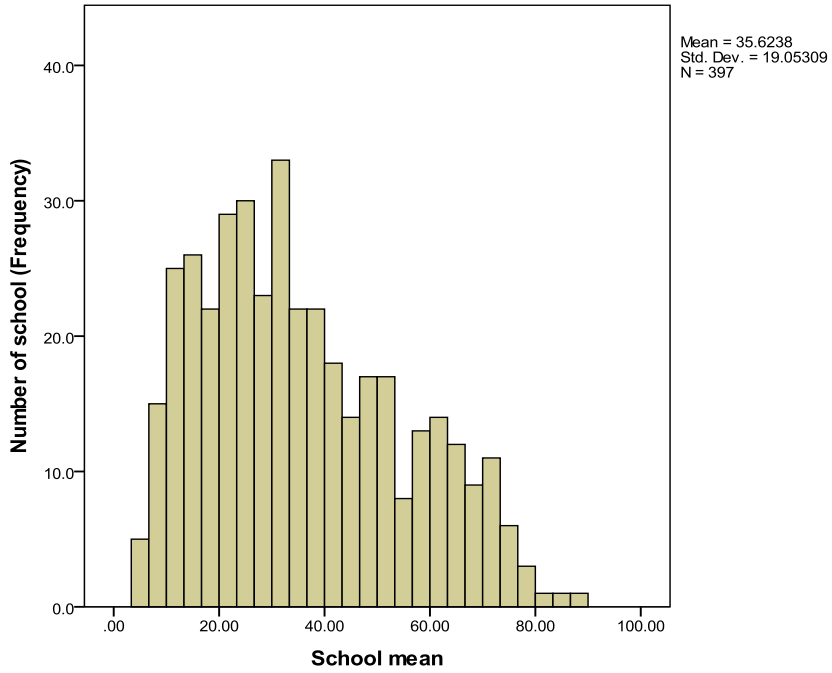
Figure 3.1.1 Distribution of overall results in Mathematics



The dataset tells that the eighth graders' population is not normally distributed, as the number of low-performing students is high.

As illustrated in figure 3.1.2 given below, the schools can be divided into two distinct categories: the high and low performing. The population on the left-hand side achieve the average of 40% of the achievement score and on the right-hand side achieve around 70% of the mean. The difference between the populations is remarkable.

Figure 3.1.2 School mean achievement score



Student Achievement by Various Content Areas in Mathematics

As per the curricular provisions, the Mathematics test includes five content areas - namely, 1) algebra, 2) geometry, 3) arithmetic, 4) sets, and 5) statistics. The number of items and the weightage they carry correspond to the weight allotted in the curriculum.

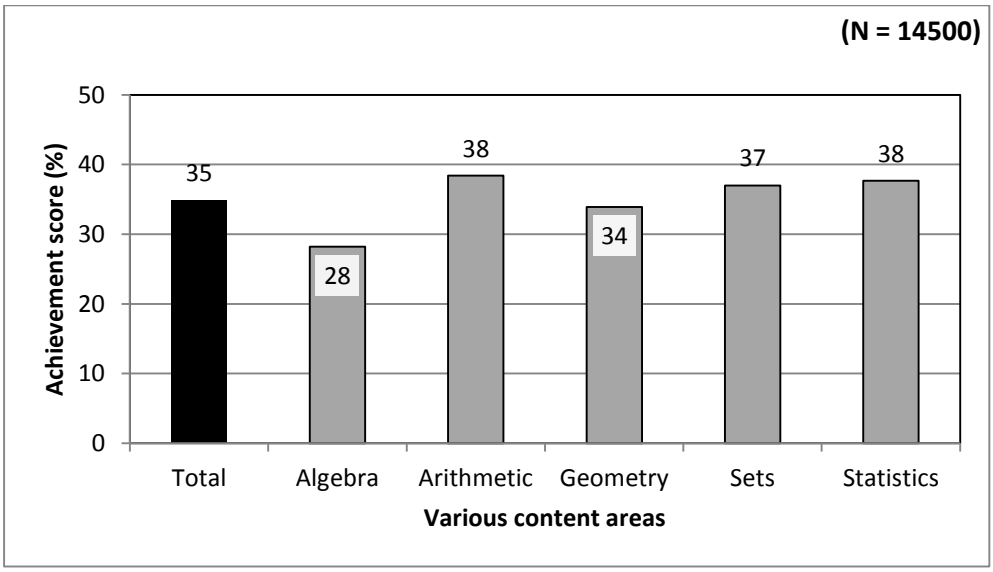
For comparability, achievement scores in all the content areas were converted into a percentage of the maximum score of each content area. Table 3.1.1 shows the students' achievement in Mathematics in five content areas.

Table 3.1.1 Achievement score in various content areas of Mathematics

Content area	Mean	SD	Min	Max
Algebra	28	24.7	0	100
Arithmetic	38	24.4	0	100
Geometry	34	26.8	0	100
Sets	37	26.0	0	100
Statistics	38	23.0	0	100
Total	35	24.2	0	100

The table shows the variations of achievements in various content areas of Mathmatics. The achievement ranges from 28 percent in Algebra to 38 percent in Arithmetic and Statistics, which is 10 percent variation. When we compare the maximum and minimum scores, the situation is the same in all content areas as the maximum score is 100 and the minimum is 0 in each content area. Figure 3.1.3 compares the variations of achievements in various content areas of Mathematics.

Figure 3.1.3 Comparison of achievement scores in various content areas of Mathematics



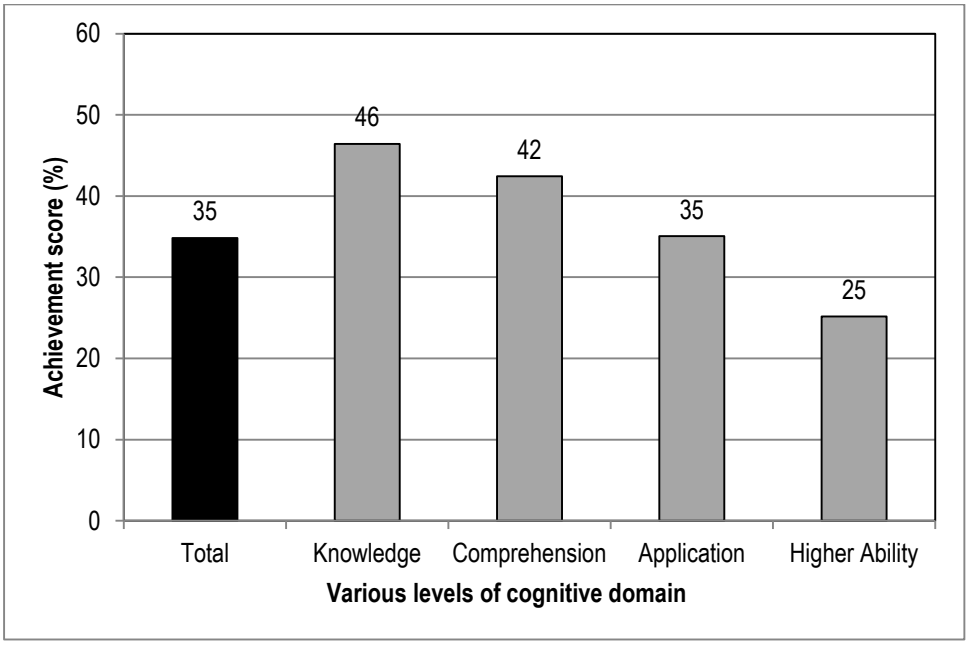
The overall national average achievement score for science is 35%. When the achievement was analysed in the five content areas, students were found performing poorly in algebra (28%), which is below the national mean (35%) by 7 percent. Similarly, achievement score in geometry is 1 percent less than the national average. The achievement scores for the remaining three content areas were found greater than the average achievement score in Mathematics where average in set is 37%, and 38% each for arithmetic and statistics.

Dataset indicates that the students are weak in algebra and geometry than in other content areas.

Achievement at Various Levels of Cognitive Domain

The mathematics test comprised of items that can be grouped under the various levels of cognitive domain according to Bloom's taxonomy (Bloom *et al.* 1956; Metfesser, Michael & Kirsner 1969). The domains are: knowledge, comprehension, application, and higher ability (reasoning/problem solving). The achievement of the students in each level of cognitive domain is shown in figure 3.1.4.

Figure 3.1.4 Comparison of achievement score at various levels of cognitive domain in Mathematics



The above figure indicates that, in general, the students' ability to solve complex problems (higher ability) is low as they achieved only 25% of the maximum score in higher ability related questions. Students are much better in recalling type of questions (46%).

While analysing further from the dataset, performance was found relatively low in higher cognitive abilities in Mathematics. For example, about 53% of the students in mathematics could not solve more than 15% of the tasks requiring the higher skills.

Student Achievement by Type of Items

There were two types of items in the test: objective and subjective. Objective items covered a wide range of content areas, which were very specific to judge because there was only one correct answer; or one explicit piece of information was needed to give the correct answer. There were some subjective items on each test version, which required a longer procedure to get full marks. Both the objective and subjective types of items were developed considering various levels of cognitive domain (knowledge, comprehension, application and higher ability) and the various difficulty levels, though the subjective items tend to be more demanding. Table 3.1.3 comprises the basic statistics of achievement in different types of items. It is clear that the students scored low in subjective items than in objective items (47%).

Table 3.1.2 Achievement by types of items

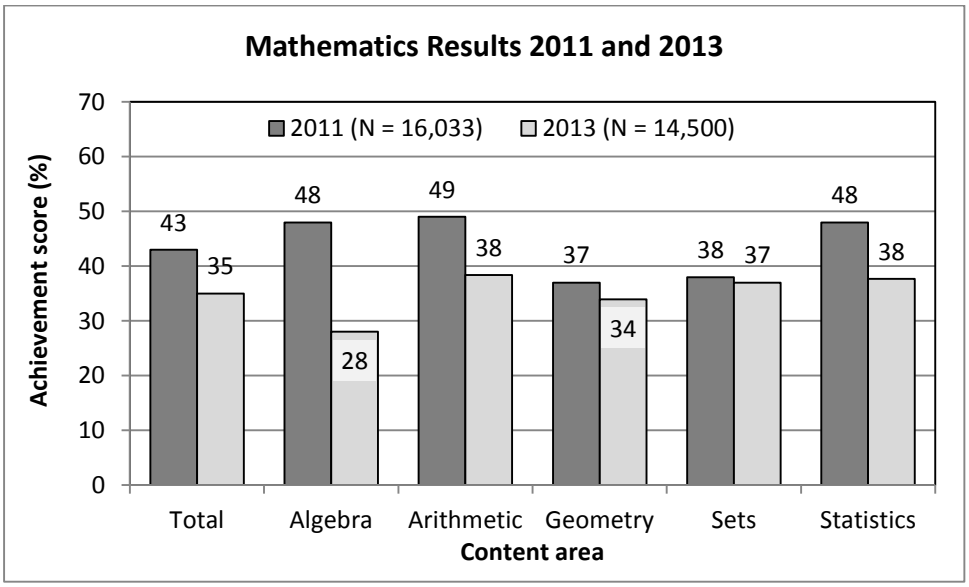
Type of items	N	Mean	SD	SE
Objective	14500	47	25.26	0.21
Subjective	14500	30	25.97	0.22

Dataset indicates that students are good at recognizing the correct answer and in very fundamental operations, such as the basic manipulation of data and numbers, and calculations in few steps. They are much weaker in reasoning, problem solving, plotting, proving the theory or formula, and constructing the shapes and figures. In many cases, the students did not even start doing the open-ended questions, hence, the low score.

Comparison of Achievement Results between NASA 2013 and 2011

Mathematics achievement in NASA 2013 is found 8 percent lower than the previous assessment (2011), which is presented in figure 3.1.5 below. Compared to the 2011 assessment, the mathematics result is comparatively lower in all the content areas.

Figure 3.1.5 Comparison of achievements between NASA 2011 and 2013 in various content areas of Mathematics



In all the content areas, the achievement level in 2013 is lower than in 2011, but in algebra, the gap is very wide.

The dataset shows that the achievement in Mathematics, instead of improving, decreased not only in the total mean but also in each of the content areas in comparison to NASA 2011.

3.2 Comparison of Achievement Results with International Standard

NASA 2013 Mathematics achievement test was made comparable with the international assessment – TIMSS. Eleven items released by TIMSS were used as linking items. Their known difficulty parameters were fixed in the calibration of the local items. Hence, the international average of $\theta = 0$ was fixed in the Nepalese datasets. When a student's ability level in NASA 2013 is zero, it corresponds to the average level of the international standards.

Figure 3.2.1 shows the comparison of the students' achievement against the international standard. In the figure, x-axis shows the content areas of Mathematics and y-axis shows the achievement of students. The middle horizontal bold line indicates the international average. When the achievement is below the average, the bars go down and when the achievement is above the international average, the bars go up from the international mean, that is, 0.

Figure 3.2.1 Comparison of Mathematics achievement with the TIMSS scale

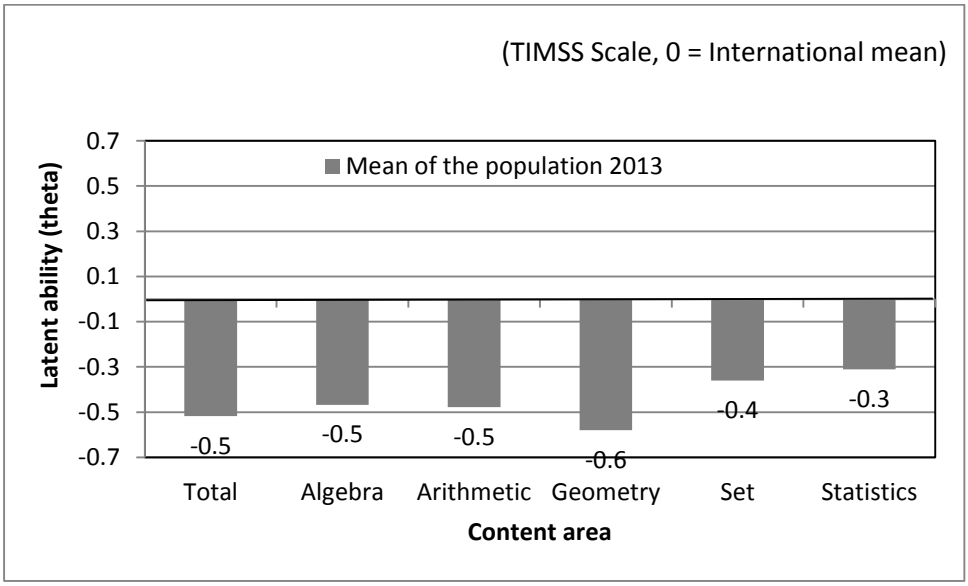


Figure 3.2.1 shows that the average achievement demonstrated by Nepali students in Mathematics is below the average achievement of TIMSS. This indicates that Nepalese students are weak in Mathematics when their achievement is compared with the international standard. The situation is the same in all content areas of mathematics as the average achievement score in each content area is below the international mean.

Comparison of NASA 2013 achievement scores with TIMSS indicates that Nepali students did not meet the international standard in Mathematics. Moreover, the situation is the same in various content areas of Mathematics.

3.3 Achievement Score by Diversity Factors

Diversity is a relative and contextual term. In the context of NASA, six diversities have been considered - namely geographical/ecological, regional, language, gender, ethnic/caste and economic diversity. NASA 2013 background information questionnaire included the questions related to these six factors of diversity related questions. However, this assessment also considered three additional comparisons. They are by districts, by school type (community/institutional) and by school location (rural/urban). These comparisons are carried out to assess the equity status of students based on achievement scores.

Student Achievement by Districts

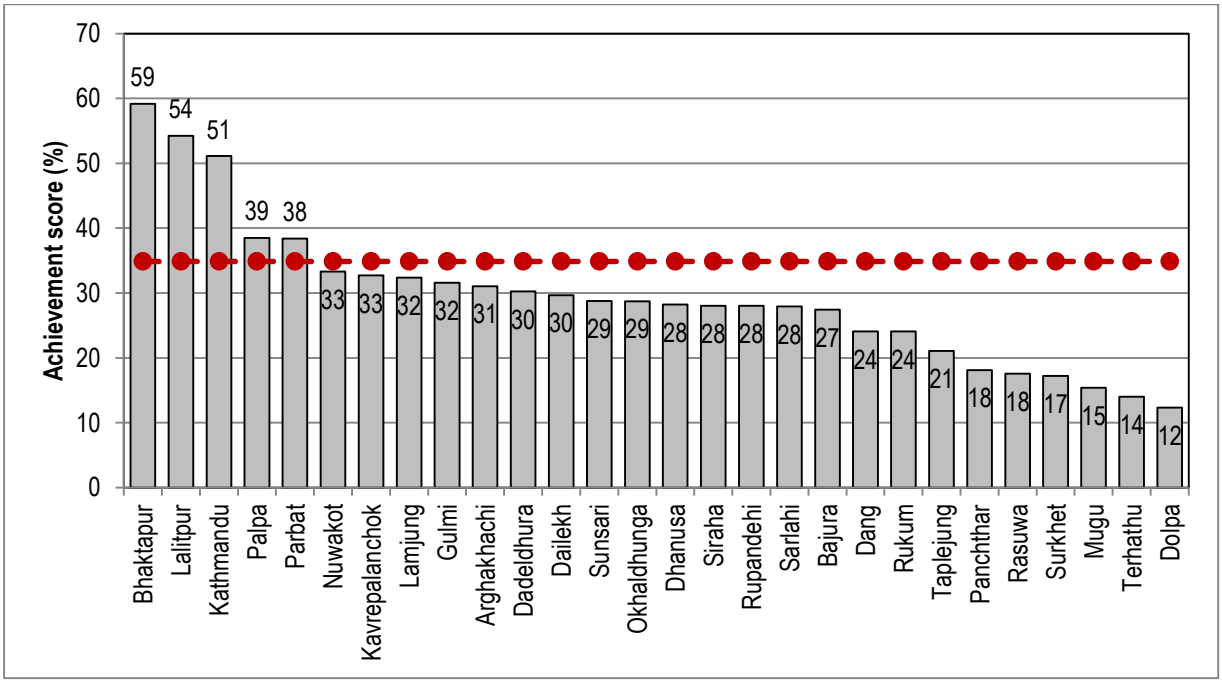
The achievement of each of the 28 sample districts is presented in table 3.3.1. The table shows mean achievement of sample districts in descending order according to the achievement scores.

Table 3.3.1 Average achievement score and standard deviation (SD) of sample districts

District	N	Mean	SD	District	N	Mean	SD
Bhaktapur	882	59	22.4	Sarlahi	322	28	17.6
Lalitpur	755	54	23.5	Rupandehi	1068	28	20.6
Kathmandu	2332	51	24.3	Siraha	556	28	22
Palpa	602	39	22.8	Dhanusa	469	28	17.1
Parbat	327	38	24.3	Bajura	288	27	15.4
Kavrepalanchok	712	33	19.4	Rukum	321	24	21.7
Nuwakot	382	33	19.8	Dang	686	24	20
Gulmi	480	32	18.7	Taplejung	216	21	17.6
Lamjung	303	32	16.9	Rasuwa	97	18	12.1
Arghakhachi	373	31	22.4	Panchathar	363	18	14.7
Dailekh	453	30	17.7	Surkhet	778	17	13.7
Dadeldhura	325	30	20.1	Mugu	104	15	17.5
Okhaldhuga	320	29	22.9	Terhathum	174	14	12
Sunsari	730	29	20.5	Dolpa	82	12	9.4
				Total	14500	35	24.2

The figure 3.3.1 compares the mean achievement of sample districts with the national mean score in Mathematics. It shows wide variations among the districts, for example the difference between the high performing district Bhaktapur (59%) and the low performing district Dolpa (12%) is 47 percent.

Figure 3.3.1 Average achievement score of sample districts in Mathematics



Of the selected districts in the sample, except for five districts, all districts' achievement score is below the national average. Out of five high performing districts, three districts are from the Kathmandu Valley and two (Palpa and Parbat) are from Western Development Region. Among the low achieving six districts three (Dolpa, Mugu and Surkhet) are from Mid-Western Development Region, two are from Eastern Development Region (Terhathum and Panchthar) and the remaining one (Rasuwa) is from Central Development Region. The data-mining tool of the SPSS software – Decision Tree Analysis (DTA) points out those lowest performing schools are mostly from Mid-Western and Eastern Regions.

The difference in achievement scores among the districts is statistically significant ($p < 0.001$). The variation explained in achievement among the districts is $\eta^2 = 0.27$; it explains 27% of the variation in achievement. Effect size is $f = 0.61$, which indicates that the difference between the lowest achieving and highest achieving districts is remarkably high.

The result indicates that there is a wide difference in achievement among districts. The results are bound to the selected 28 districts. Some districts, for example, Bhaktapur (59), Lalitpur (54), Kathmandu (51), Palpa (39) and Parat (38) have achieved higher than the average. Whereas, Dolpa (12), Terhathum (14), Mugu (15), Surkhet (17), Panchathar (18) and Rasuwa (18) have the mean score very low.

Student Achievement in Various Ecological Zones

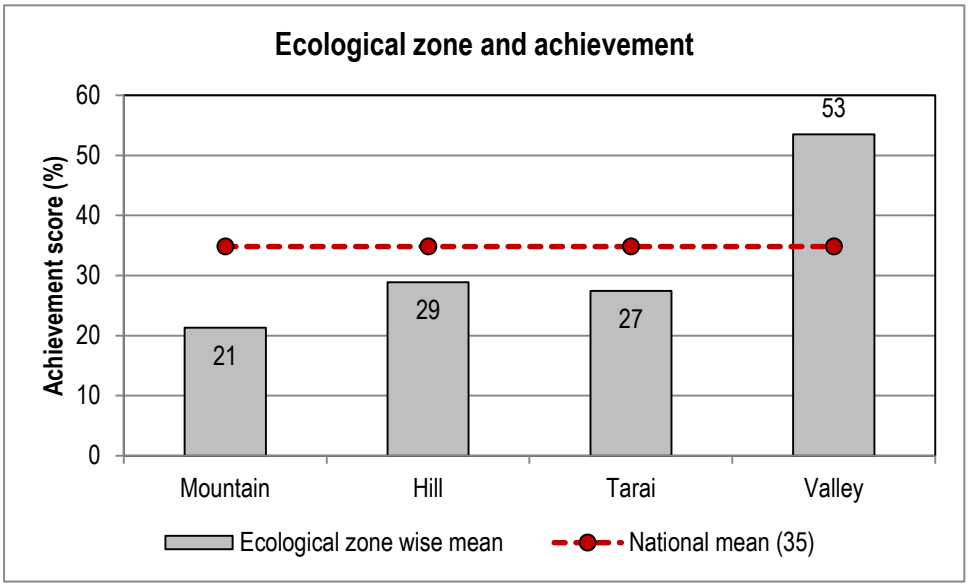
The Mountain, Hill and Tarai are three geographical zones in Nepal though the Kathmandu Valley has been considered as a special geographical area because it is the most densely populated area in the country with more opportunities than other areas. Not only from the population point of view but also due to the mixed ethnicities, weather conditions, economic activities, urbanization, as well as the dense human capacity have been taken into account while considering the Kathmandu Valley as a unique fourth geographical area in the analysis. The variation in the Ecological zones in NASA 2013 is presented in table 3.3.2. The table has also included the number of students, standard deviation, standard error and minimum and maximum scores in each of the categories

Table 3.3.2. Achievement in various ecological zones

Region	N	Mean	SD	SE	Minimum	Maximum
Mountain	787	21	16.3	0.6	0	86
Hill	5913	29	20.6	0.3	0	97
Tarai	3831	27	20.1	0.3	0	99
Valley	3969	53	23.9	0.4	0	100
Total	14500	35	24.2	0.2	0	100

Figure 3.3.2 compares the student achievements among various ecological zones as well as with the national mean score in mathematics.

Figure 3.3.2. Comparison of achievement scores by ecological zones in Mathematics



The data shows that, on average, the students from the Kathmandu Valley (53%) have outperformed the students from all the other ecological zones. The students from the Mountain areas performed the lowest (21%).

The achievement in the regions differs significantly ($p < 0.001$). In terms of Tukey's *post hoc* test, all the zones deviate from each other in a statistically significant manner at $p = 0.05$ level.. The effect size is $f = 0.54$, which shows a very wide difference between the highest and lowest performing ecological zones. Ecological zone explains 22% of the variation in the data. As a comparison, district explains 27% of the variation in achievement.

Excluding the Kathmandu Valley with the ANOVA analysis, the ecological zone explains 1% of the variation in achievement where the effect size is small ($f = 0.1$).

Dataset indicates that there is a high difference between the student performances in four ecological zones. Students in the Kathmandu Valley have outperformed (53%) the other students. The achievement is the lowest in the Mountain area (21%).

Student Achievement by Development Regions

The student achievement varies according to the developmental regions, which are divided into five as, 1) Eastern, 2) Central, 3) Western, 4) Mid-Western, and 5) Far-Western regions. Additionally, the Kathmandu

Valley is taken as the additional region though administratively it falls under the Central Developmental region. The mean achievements of the developmental regions are given in table 3.3.3.

Table 3.3.3 Student achievement scores in various developmental regions

Region	N	Mean	SD	SE	Minimum	Maximum
Eastern	2359	25	20.3	0.4	0	97
Central	1982	30	18.7	0.4	0	99
Western	3153	32	21.5	0.4	0	97
Mid-Western	2424	22	18.3	0.4	0	96
Far-Western	613	29	18.1	0.7	0	96
Valley	3969	53	23.9	0.4	0	100
Total	14500	35	24.2	0.2	0	100

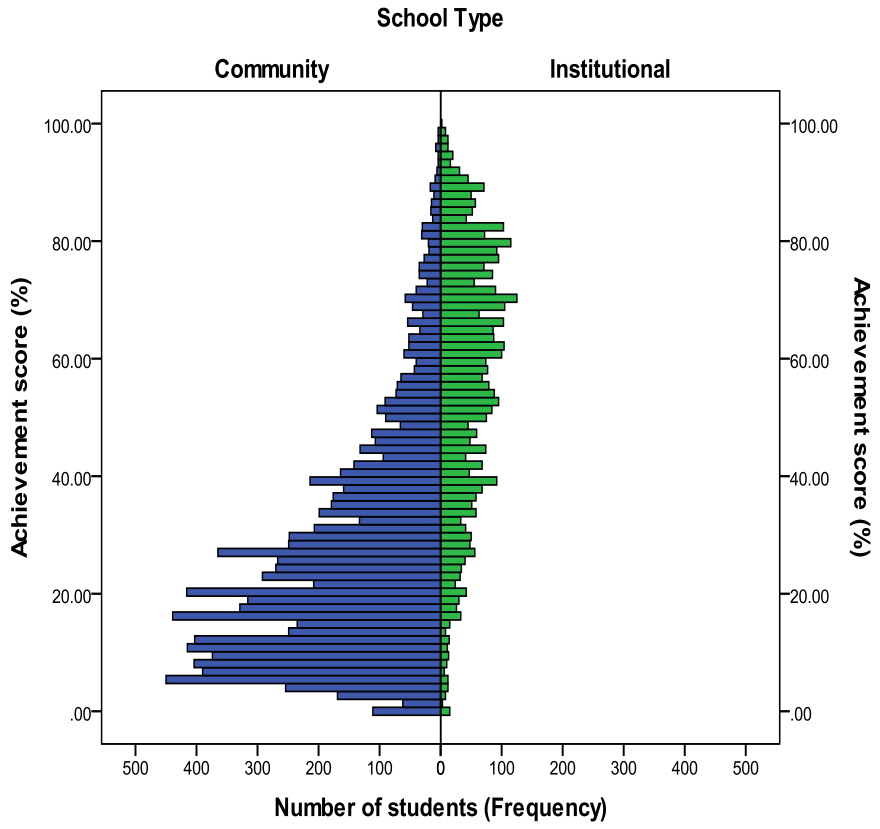
The best performance is found in the Kathmandu Valley (53%). All other regions have performed below the national average. The performance is the lowest particularly in the Mid-Western (22%) and Eastern (25%) regions. The difference between the regions is statistically significant ($p < 0.001$) and Tukey's *Post hoc* test shows that all the developmental regions differ from each other significantly with at least at 5% significance level except for central and far-western regions. The effect size is also found high ($f = 0.57$); and it explains 24% of the variation in student achievement. However, if the Kathmandu Valley is excluded from the analysis, the effect size will be medium ($f = 0.2$) and the development regions explain only 4% of the variation in achievement.

The dataset reveals that there is wide inequity among the developmental regions for children's opportunities to reach an adequate level in Mathematics. The difference between the lowest performing region (Mid-Western, 22%) and the highest performing region (the Kathmandu Valley, 53%) is remarkable, that is more than 30 percent. However, if the Kathmandu Valley is excluded from the analysis, the difference will be 10 percent. Except for Kathmandu Valley, all other ecological zones and development regions are poor in Mathematics.

Student Achievement by School Type

Plotting the community and institutional schools and their mean achievements in a graph as in the figure 3.3.3, the community school students on the left hand side are distinctly distributed at least in three kinds of populations (groups). There is a remarkably high number of the students achieving below 30% of the score. The right hand side distribution shows the student population of institutional schools, which is shifted towards better performing, levels. However, in institutional schools, there are some students obtaining very low marks in Mathematics. This indicates that the students in institutional schools are also categorised as low to high performers though majority of them are high performers.

Figure 3.3.3 Distribution of community and institutional schools based on their achievement in Mathematics



All the schools are categorized into community and institutional (usually called private) schools. The difference in achievement of the two types of schools is presented in table 3.3.4.

Table 3.3.4 Student achievement according to school type

Type of school	N	Mean	SD	SE
Community	10,332	26	19.2	0.2
Institutional	3,934	57	21.7	0.3
Total	14,266	35	24.2	0.2

The average performance of the students from institutional schools is 57% whereas it is 26% in community schools, that is, the difference is 31 percent, which is remarkable. This shows wide difference in the level of achievement between community and institutional schools in which the difference is statistically significant ($p < 0.001$) and the effect size is high ($f = 0.69$). It explains 33% of the student variation ($\eta^2 = 0.325$). Figure 3.1.3 illustrates that the deviation within the community schools is remarkable, ranging from 10% to 50%. On the other hand, most institutional/private schools from the sample are performing high. However, the number of the students in the institutional schools is small compared to the community schools.

The dataset indicates that the students from the institutional schools have outperformed the students from the community schools. However, it is difficult to identify the effects of the socio-economic status of parents and school process.

Student Achievement in Rural and Urban Schools

One of the strata of sampling in NASA 2013 was school location. The schools were categorized into rural and urban groups. The achievements of the students from rural and urban schools are disaggregated in table 3.3.5.

Table 3.3.5 Student achievement in rural and urban schools

Type of school	N	Mean	SD	SE
Rural	10,333	28	20.8	0.2
Urban	3,958	52	24.3	0.4
Total	14,291	35	24.2	0.2

The achievement level of the students in the urban schools (52%) is remarkably higher than that in rural schools (28%). The difference in average score is significant ($p < 0.001$) and the effect size is high ($f = 0.48$). Location of schools explain 19% variation in student achievement ($\eta^2 = 0.19$).

Data shows a huge disparity between urban and rural students. The difference is remarkably high.

Student Achievement in Various Language Groups

In the context of Nepal, student achievement has been found somehow affected by the language spoken at their homes i.e., the mother tongue. In many cases, mother tongue reflects the ethnic background, and hence it may be taken as a possible reason for achievement differences in mathematics.

Based on the total data, 28% of the eighth graders speak a language other than Nepali as their first language. These "other" languages are quite fragmented; the largest groups in the student dataset are Tamang (4.2%), Tharu (2.3%), Newari (2%) and Magar (1.9%). Excluding Nepali, total students are categorized into nine language groups, among which 18.9% of the students fall into the group "others". Because the languages are very fragmented and the Nepali speakers are the majority of the students, for the purpose of statistical analysis, all the other languages were grouped as "non-Nepali speakers". The results are presented in tables 3.3.6 and 3.3.7 and further illustrated figure 3.3.4.

Table 3.3.6 Achievement of Nepali and non-Nepali language speaking groups

Language group	N	Mean	SD	SE
Nepali	10453	37	24.9	0.2
Non-Nepali	4047	28	20.9	0.3

The difference between the mean achievements of students having their home language Nepali and Non-Nepali is 9 percent. However, students from Newari community have outperformed (47%) the other language speaking groups, including Nepali-speaking group (37%). On contrary, language groups, such as Rai (20%), Sherpa (22%), Tharu (22%), Limbu (23%), Magar (27%) and others (27%) have performed very low (see table 3.3.7).

Table 3.3.7 Student achievement by linguistic and ethnic backguound

Language/Etnicity[1]	N	Mean	SD	SE	Minimum	Maximum
Newari	282	47	23.9	1.4	3	100
Gurung	78	32	13.8	1.6	4	72
Tamang	587	32	19.5	0.8	0	99
Magar	260	27	18.3	1.1	0	85
Limbu	133	23	16.0	1.4	0	76
Tharu	328	22	17.1	0.9	0	89
Sherpa	53	22	21.5	3.0	0	77
Rai	47	20	14.1	2.1	4	65
Others	2268	27	20.9	0.4	0	99
Nepali	10453	37	24.9	0.2	4	72

1) The language groups with less than 10 students have not been considered for analysis separately.

The difference between the language groups is statistically significant ($p < 0.001$). However, the effect size is medium ($f = 0.21$); the language group explains 4% of the variation in the data ($\eta^2 = 0.044$). Excluding Nepali and the "other" minority language groups, the effect size is high ($f = 0.43$), indicating a remarkable difference between the highest performing group (i.e., Newari, 47%) and the lowest performing group (i.e., Rai, 20%).

Achievement of the Students of Various Ethnic/Caste Groups

Education in Nepal has been influenced in several ways by the legacy of the historical caste system which still remains in the mind-set of people in Nepalese society. Historically, the Brahmans and Chhetris have been heavily involved in education, but Dalits, for example, have remained outside the educational system. Hence, modern society has made lots of efforts to make the education accessible for all children regardless of the caste and ethnicity. The recent National Population Census (2011) shows that the enrolment of Hill Dalits has increased remarkably at the primary level of schooling but their number at the secondary level and higher education is still very small (CBS, 2012). The achievement results of the students from various ethnic/caste groups are presented in table 3.3.8 and further illustrated in figure 3.3.5.

Table 3.3.8 Student achievement by ethnic/castes backgurond

Caste	N	Mean	SD	SE	Minimum	Maximum
Brahman/Chhetri	5330	40	25.3	0.3	0	100
Janajati	5051	36	23.7	0.3	0	99
Dalit	1428	24	17.6	0.5	0	89
Madhesi	1302	29	21.7	0.6	0	99
Minorities	38	28	19.3	3.1	4	92
Others	698	37	25.7	1.0	0	100
Total	13847	35	24.2	0.2	0	100

Dalits' achievement score in Mathematics is lower (24%) than the achievement of other castes/ethnic groups. The difference between Brahmans/Chhetris (40%) and Dalit (24%) is 16 percent. The overall difference among the groups is statistically significant ($p < 0.001$), but the effect size is moderate ($f = 0.2$). Categorization of students according to their ethnic/caste background explains 4% of the variation in data ($\eta^2 = 0.04$).

Except in the Kathmandu Valley (41%), the performance of Dalit students is low across the ecological zones and development regions.

Table 3.3.9 Dalit students' achievement in various ecological zones and developmental regions

	Eastern	Central	Western	Mid-Western	Far-Western	Valley	Total
Mountain	14			10	24		20
Hill	20	28	26	20	21		23
Tarai	23	27	23	14			22
Valley						41	41
Total	22	27	25	19	23	41	24

Dataset indicates that Dalit students' performance score is lower than the performance of other ethnic/caste groups.

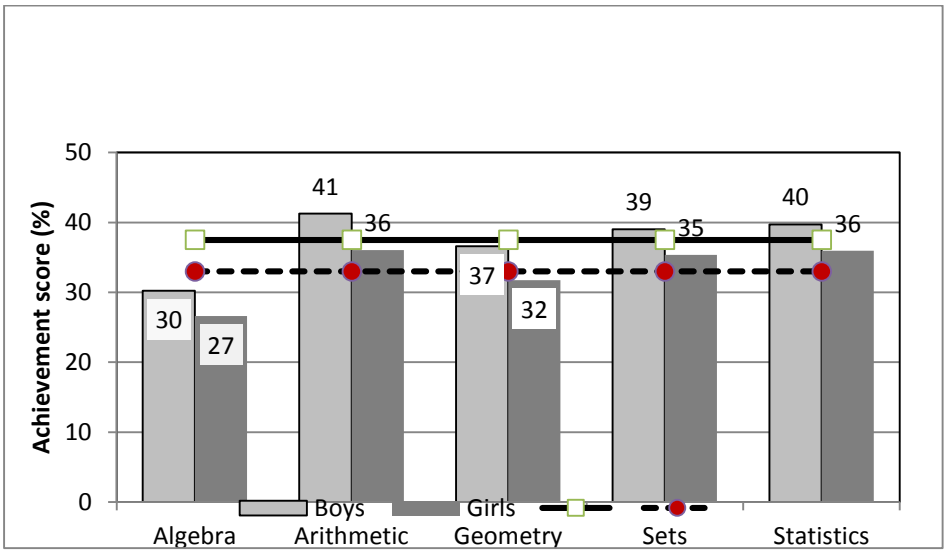
Gender and Student Achievement

In previous assessments, there was a significant difference in achievement between boys and girls (see ERO, 2011; ESD, 2008). Similar situation is found in Mathematics in this assessment as there is 5 percent difference in achievement between boys and girls. Basic achievement results of boys and girls are presented in table 3.3.10. Figure 3.3.4 further compares the achievement of girls and boys in various content areas. The figure shows that in each content area boys have outperformed girls.

Table 3.3.10 Student achievement of boys and girls

Sex	N	Mean	SD	SE	Minimum	Maximum
Boys	6837	38	24.2	0.3	0	100
Girls	7547	33	24.0	0.3	0	100
Total	14384	35	24.2			

Figure 3.3.4 Comparison of achievement between boys and girls in different content areas of Mathematics

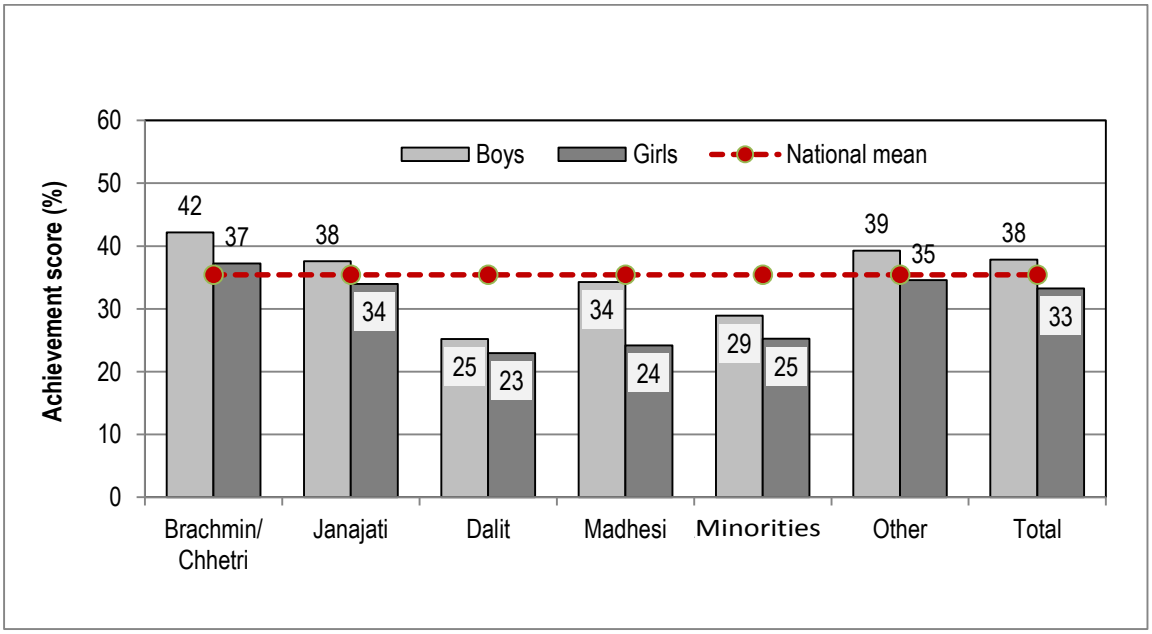


In NASA 2013, there is statistically significant ($p < 0.001$) difference in the achievement between boys (38%) and girls (33%) in Mathematics. It is noteworthy that in all content areas boys have outperformed the girls. However, the effect size is small ($f = 0.1$), indicating that the difference is not remarkable. Gender explains only 1% variation in achievement score.

Gender and ethnicity/caste

There are gender differences in the achievement of various ethnic/caste groups in NASA 2013 in Mathematics. For example, the difference in achievements between boys and girls is the highest in Madhesi communities, that is, 10 percent. Similarly, in the Brahmins/Chhetris and "others" groups, the difference 5 percent (see figure 3.3.5). The difference is statistically significant at $p < 0.001$. The effect size is medium ($f = 0.21$), which explains 4% variation in achievement ($\eta^2 = 0.041$).

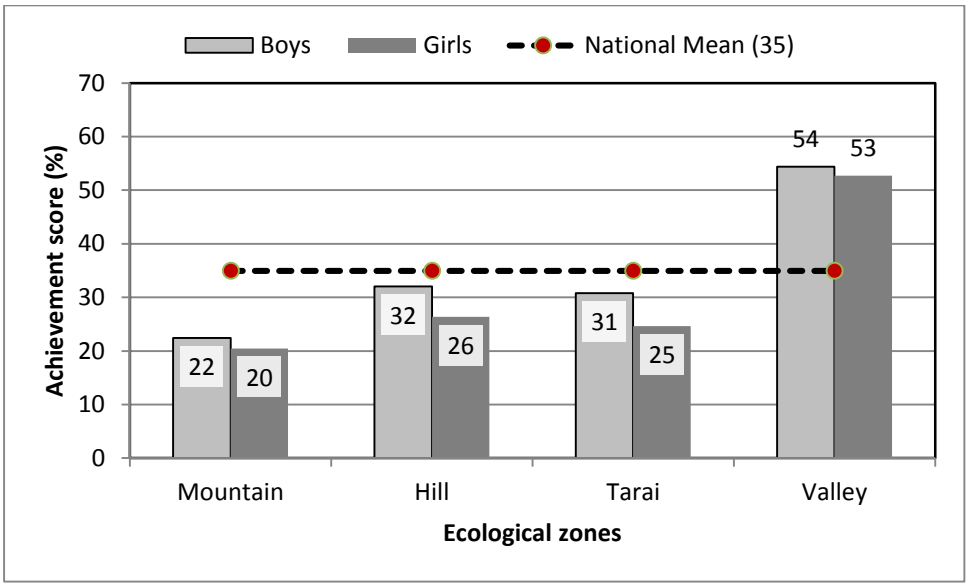
Figure 3.3.5 Comparison of achievement between boys and girls in various ethnic/caste groups in Mathematics



Gender and ecological zone

The achievement of girls and boys in mathematics differs significantly across ecological belts ($p < 0.001$). In all ecological regions, boys have outperformed girls but in the Kathmandu Valley, there is no statistically significant difference between boys and girls, though boys are somehow better than girls(see figure 3.3.6).

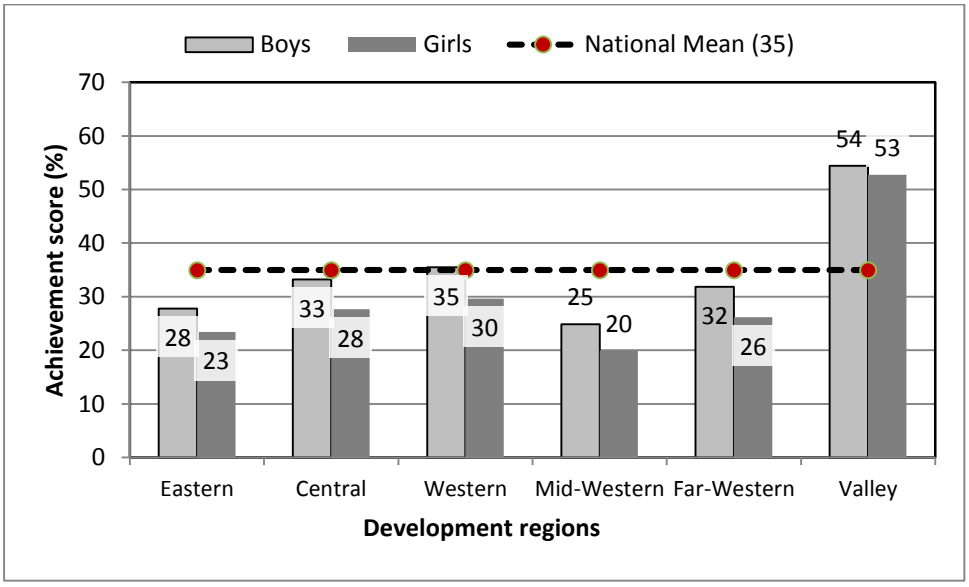
Figure 3.3.6 Comparison of achievement between boys and girls in various ecological zones in Mathematics



Gender and developmental region

The difference in achievement between the boys and girls is the widest in the Far-Western region, which is 6 percent. In the Central, Western and Mid-Western regions, it is 5 percent whereas only a small (that is 1%) difference is found in the Kathmandu Valley. The differences are statistically significant ($p < 0.001$) except for the Kathmandu Valley. Comparison of achievement of boys and girls in various development regions is illustrated in figure 3.3.7.

Figure 3.3.7 Comparison of achievement between boys and girls in various development regions in Mathematics



Dataset shows that boys are outperforming girls in all content areas of mathematics. Differences are not wide though they are significant. Gender explains only 0.1% variation in achievement, that is, a small effect size. Differences are the widest in Western and Far-Western regions. Only in the Kathmandu Valley, there is narrow difference in achievement between boys and girls.

3.4 Selected Explanatory Factors and Achievement

Several factors have already been handled in the previous section such as geographical factors including districts, ecological zone and developmental region; school related factors including school type and school location. Some individual related factors were also handled, such as home language, caste/ethnicity and gender.

This section deals with socio-economic status (SES) of the students' families, paid work beyond the school time, students' attitude towards Mathematics as a school subject, age of the student, and support provided for the studies as the main family and individual related factors. Similarly, in this section school and teacher-related factors, the availability of textbooks, homework given by the teacher and selected activities in school are also handled.

Achievement of Students from Various Socio-economic Status (SES)

The variables indicating the socioeconomic status were grouped into seven categories. The seven categories of SES are: mother's education, father's education, mother's occupation, father's occupation, home possessions , home accessories, and type of school attended. Finally, the SES is estimated based on seven indicators related to the economic, educational, and occupational background of the family (see, section 2.8). In this section, the education of the parents is further elaborated, so that the literacy of the parents is analysed in relation to achievement in Mathematics.

Several SES-related variables were analysed by using a data-mining tool of SPSS– the Decision Tree Analysis (DTA). The method was very effective in finding the cut-offs of the predicting variable, such as mother's education, and classifying the factor into several groups which differ statistically in the most significant way from each other in relation to student achievement.

Parents' education

In the background questionnaire of NASA 2013, parents' education is divided into the following seven categories: 1) illiterate, 2) literate, 3) grade 10, 4) SLC, 5) Certificate level, 6) Bachelor's level, 7) Master's level and above.

DTA classifies mother's education into seven groups with statistically significant differences in students' achievement levels. The data shows that students having illiterate mothers obtained only 28% of the maximum score on average, which is the least average score in comparison to the students having other categories of mothers in terms of their education. The achievement of students has increased along with the increase in mothers' educational level from literate to Bachelor's degree. However, the achievement of students having mothers with the qualification of Master's level and above has been lower compared to the achievement of those students having mother with Bachelor's degree. For example, when mothers are Bachelor's degree holders, their children obtained more than 30 percent scores compared to the illiterate mothers. Only the students having illiterate or just literate mothers are performing below the national average. The figures 3.4.1 and 3.4.2 below explain these results. In each group, the number of mothers is high enough to make a credible prediction. The difference between each group is statistically significant ($p < 0.001$). Mother's education explains 13% of variation in achievement ($\eta^2 = 0.127$), indicating a medium effect size ($f = 0.38$).

Figure 3.4.1 DTA of mother's education and students' achievement in Mathematics

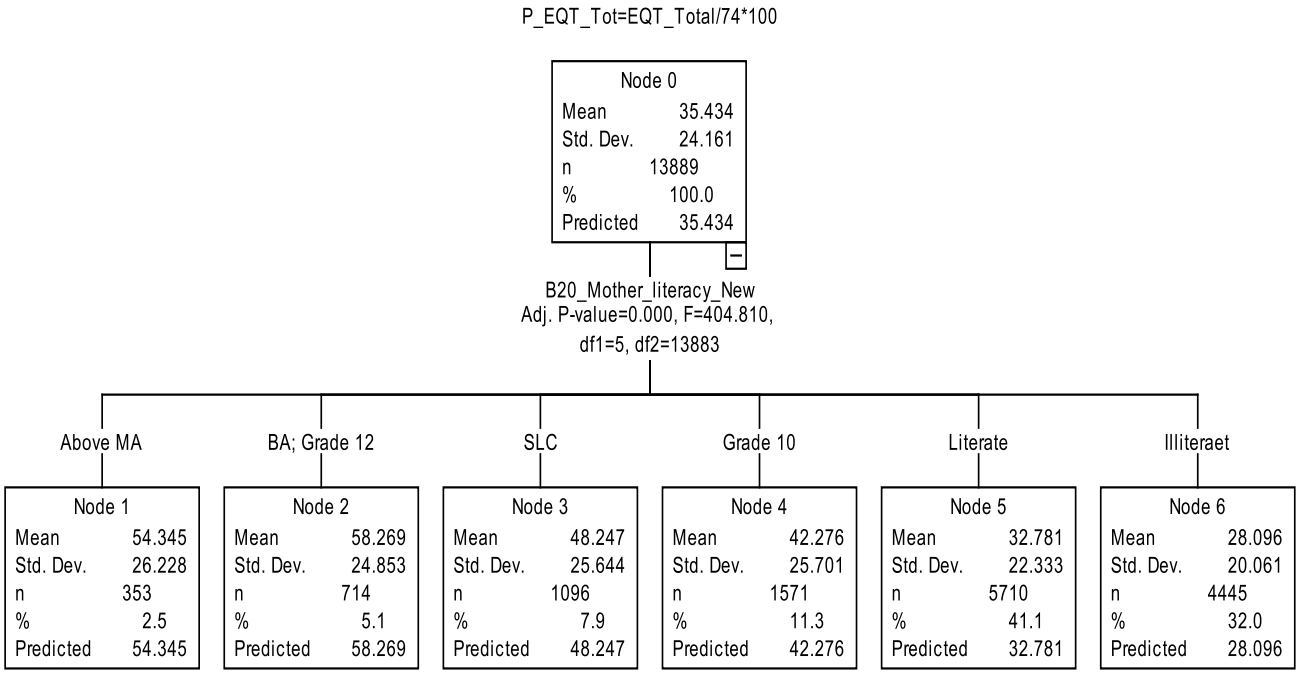
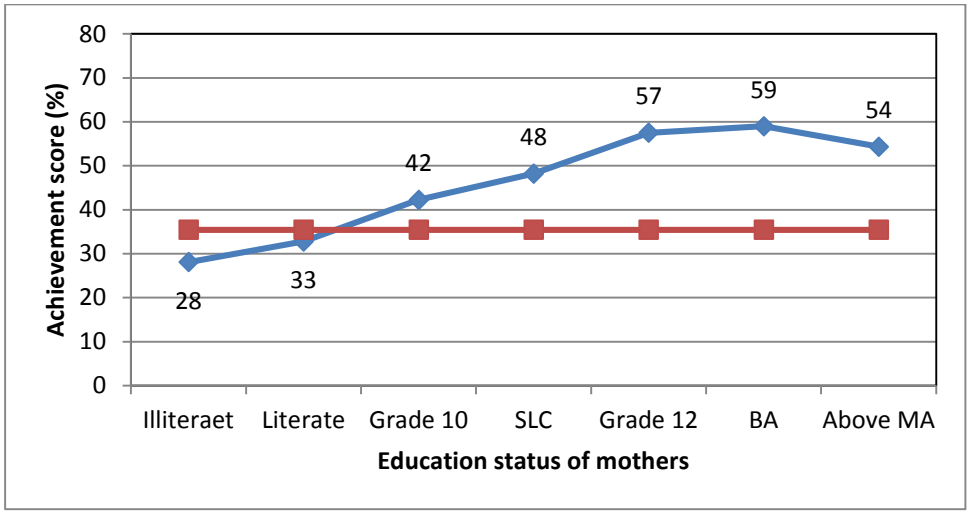


Figure 3.4.2 Relation between mother's education and students' achievement in Mathematic



In similar vein, the DTA categorizes father's education into seven groups, they are, illiterate, literate, grade 10, and SLC, grade 12, and Bachelor's, Master's and above (see, figures 3.4.3 and 3.4.4). As in the case of mothers, children with illitrate fathers have obtained lowest marks in comparison to other categories of fathers. The difference in achievement between different groups is statistically significant at $p < 0.001$ level. When compared with mother's education, children of literate mothers obtained higher score than the children of literate fathers. The children of literate fathers have achieved 25 percent more than the children of illiterate fathers, whereas in the case of mothers the difference is 28 percent. The same is found for SLC passed mothers (48%) and fathers (41%). However, achievement is higher (56%) when the father is master's degree. Father's education explains 14% variation in data ($\eta^2 = 0.14$), indicating a high effect size ($f = 0.41$).

Figure 3.4.3 DTA of father's education and students' achievement in Mathematics

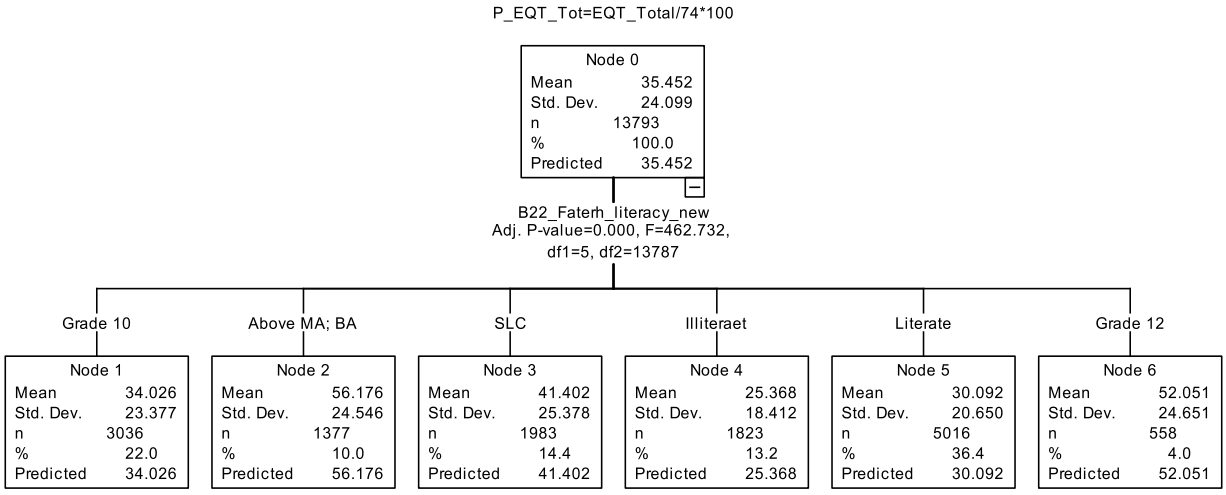
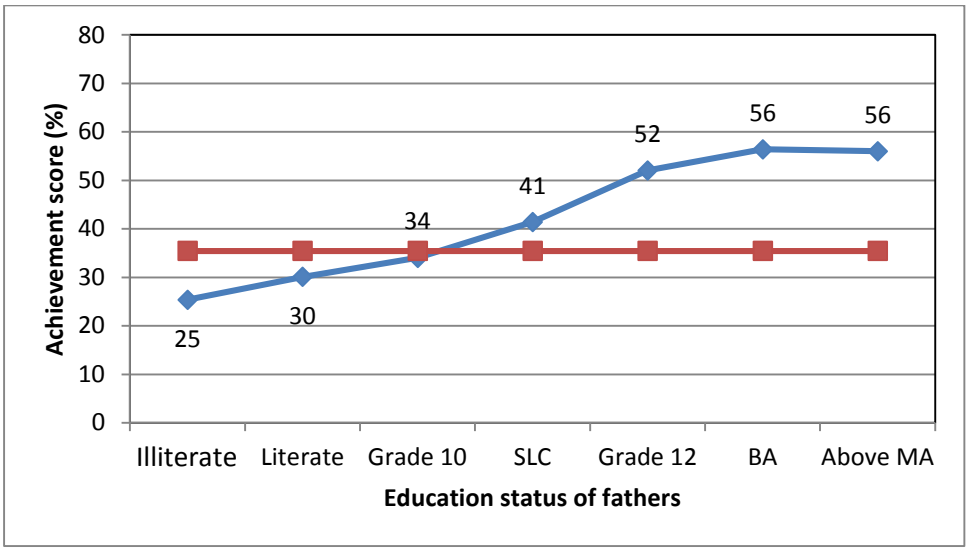


Figure 3.4.4 Relation between father's education and students' achievement in Mathematics



Mothers' and fathers' education status shows more or less similar result in students' achievement, i.e. the higher the education status of parents, the better the performance of students and vice-versa. Children of mothers' having Bachelor's degree performed better than the other children. However, children of fathers having Bachelor's or Master's degree have obtained obtain the same level of higher achievement. It is evident from the result that the educational status of parents and students' achievements are positively correlated, that is, the higher the parents' education, the better the result of their children.

Result reveals that the educational level of the parents predicts the children's future achievement level in Mathematics. When both parents are illiterate, achievement of their children becomes very low compared to the children of other parents.

Parents' occupation and student achievement

The occupations of parents were grouped into the following eight categories: 1) agriculture and household chore, 2) only household chore, 3) work abroad, 4) teaching 5) others jobs/service, 6) business, 7) daily wage and 8) work in others home (maid). The DTA was used to find statistically the most deviating occupation groups related to student achievement (see fig. 3.4.5). The difference among groups are statistically significant ($p < 0.001$). Student achievement is the lowest i.e., 29 percent, when the mother's occupation background is agriculture, household chore or daily wage, whereas the children of mothers having service occupation have scored the highest (i.e., 54 percent). It shows that either economic or

intellectual ability or both at home give positive impact on children's achievement. Mothers' occupation shows 12% variation in data ($\eta^2 = 0.121$) with a medium effect size ($f = 0.37$).

Figure 3.4.5 DTA of mother's occupation and students' achievement in Mathematics

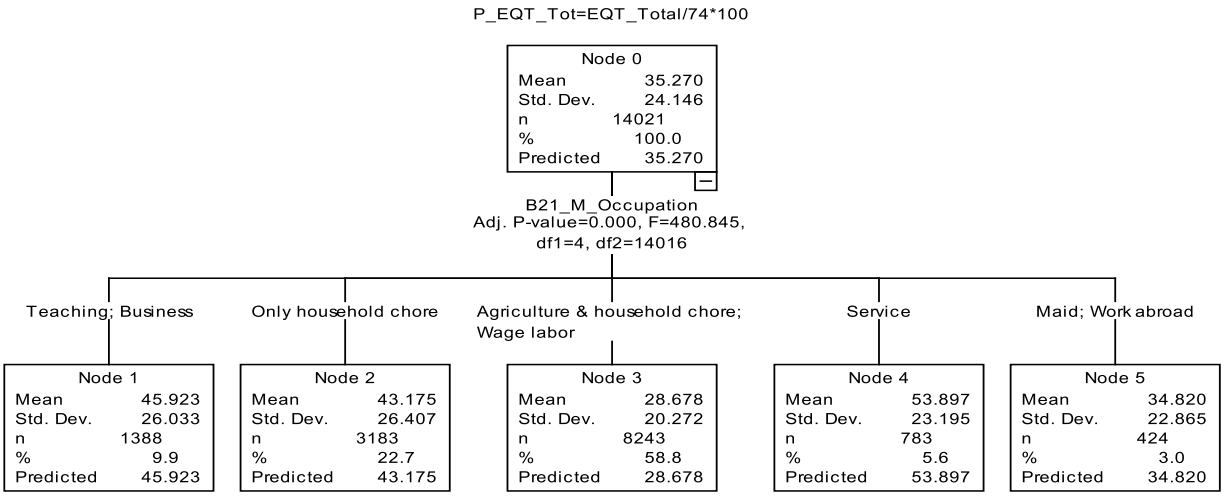
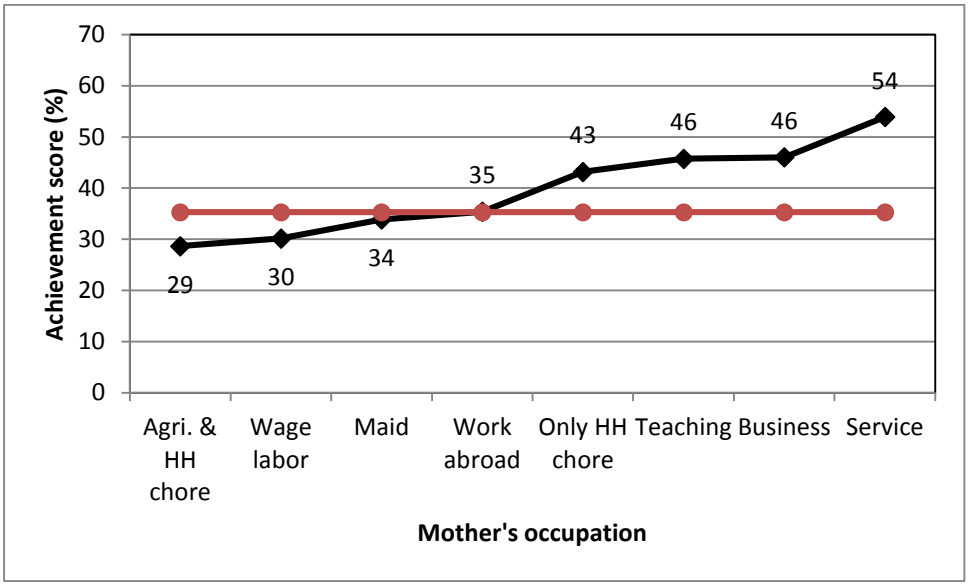


Figure 3.4.6 Comparison of mothers' occupation and students' achievement in Mathematics



Similarly, DTA categorizes fathers' occupation into six groups. These six categories in ascending order according to children's achievements are: 1) only household chore, 2) agriculture and household chore, 3) wage labour or maid, 4) work abroad, 5) business or teaching, and 6) service. If the father is service holder the achievement of his children is higher, (i.e., 48 percent), whereas if the father is involved only in

household chores, the achievement of his children is remarkably low (i.e., 23 percent). The difference is statistically significant ($p < 0.001$). Fathers' occupation shows 14% variation in achievement ($\eta^2 = 0.136$) with a high effect size ($f = 0.4$).

Figure 3.4.7 DTA of fathers' occupation and students' achievement in Mathematics

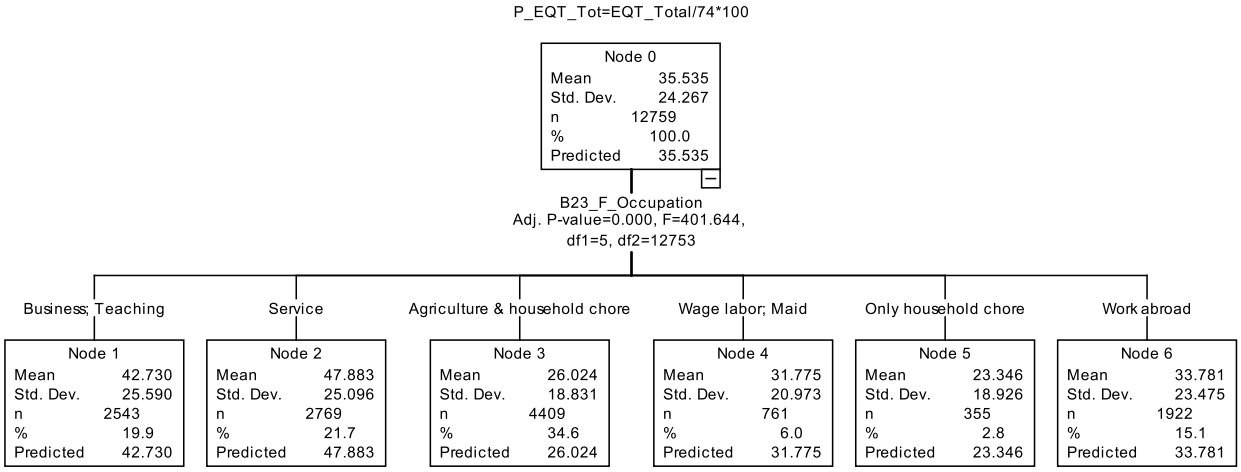
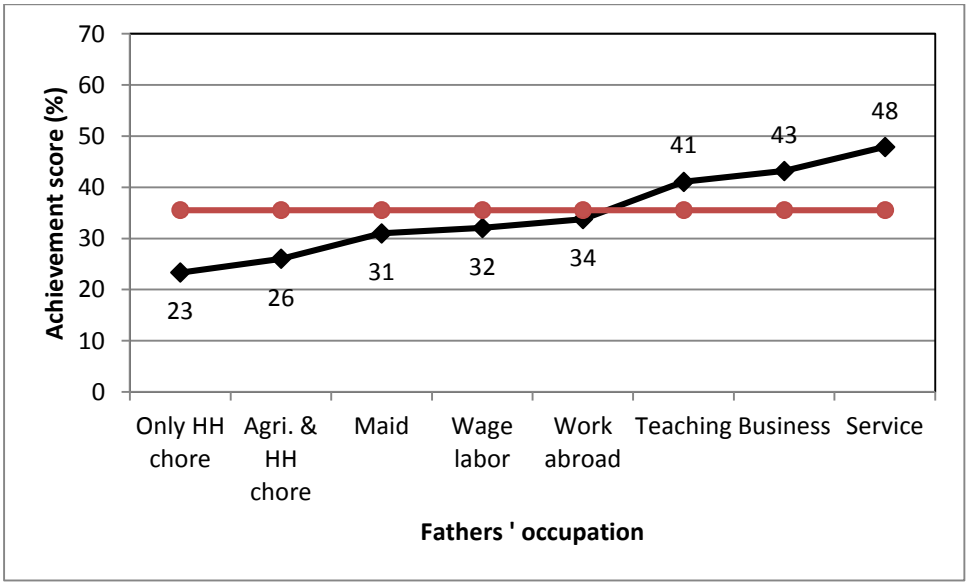


Figure 3.4.8 Comparison of fathers' occupation and students' achievement in Mathematics



Students' achievement is low when both parents are involved in agriculture and household chores, and it is the highest when the parents are involved in service.

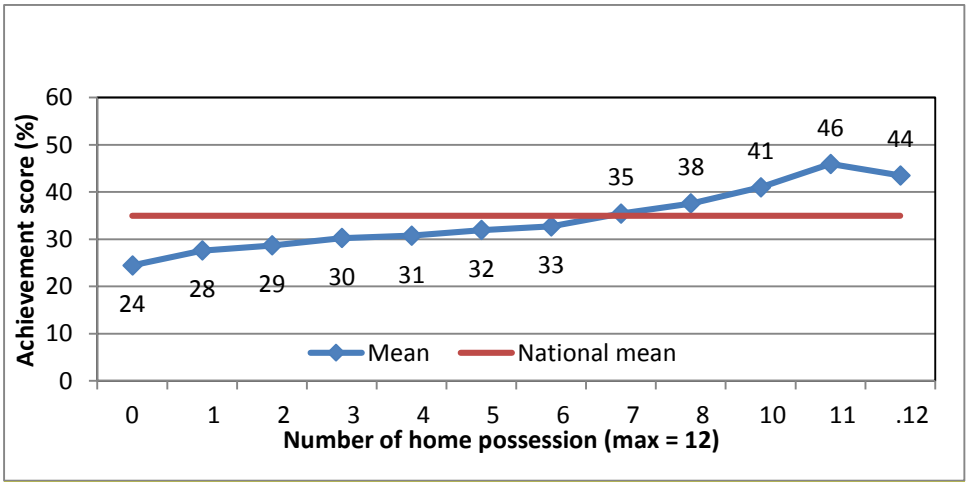
The dataset explains that if the parents are involved in agriculture or household chores, the students' achievement in Mathematics is significantly lower than the achievement of students from other occupational groups.

Home possessions and accessories

There were two kinds of home possessions defined in the background information questionnaire distributed to the students. One is related to the facilities that are supportive for the students' study at home. These facilities are, for example, a table for study, a separate room for student, a peaceful place for study, a computer for school work, software for computer assisted learning, internet facilities, literary magazines, access to classical literature, poetry books, or artistic things like pictures, dictionary and other books that support children for their study. Another type of home possession includes different types of normal home accessories, such as the number of mobile phones, televisions, computers, and radios.

There were 12 questions in the background questionnaire filled by the students related to home possessions. Each scored one if the student had access to the possession. Adding these items up, the maximum score is 12, indicating that the students have the access to all of the possessions. The result shows that higher access of home possessions provides high score to the students and hence the lower the score the fewer the possessions at home. Figure 3.4.9 shows relation between home possessions and achievement level where the achievement level of the student goes up when there is access to home possessions. Pearson product moment correlation coefficient between the achievement level and the factor (r = 0.193) is statistically significant ($p < 0.001$) though it is not very high, and the effect size is moderate (d = 0.43).

Figure 3.4.9 Relation between home possessions and achievement in Mathematics



The figure shows the relation between the home possessions and achievement in Mathematics. It is seen from the figure that when the number of accessories increase, the achievement also increases. In the background questionnaire, the question asked to the students was "how many of the following accessories do you have in your family?" The options were 0, 1, 2 and 3 or more. After dichotomizing the items individually by using meaningful cut-offs found with ANOVA and DTA, (see table 3.4.1), all the four indicators were added. The maximum score was four, indicating that the students possessed at home a set of all the accessories.

Table 3.4.1 Dichotomizing the indicators for home accessories

Accessory	cut-off for 1	cut-off for 0
Mobile phone	3, 4	0 to 2 missing
Television	1 to 3	0 missing
Computer	1 to 3	0 missing
Radio	3, 4	0 to 2 missing

The relation between the availability of home assessories and achievement is presented in the following figure.

Figure 3.4.10 Relation between home accessories and achievement in Mathematics

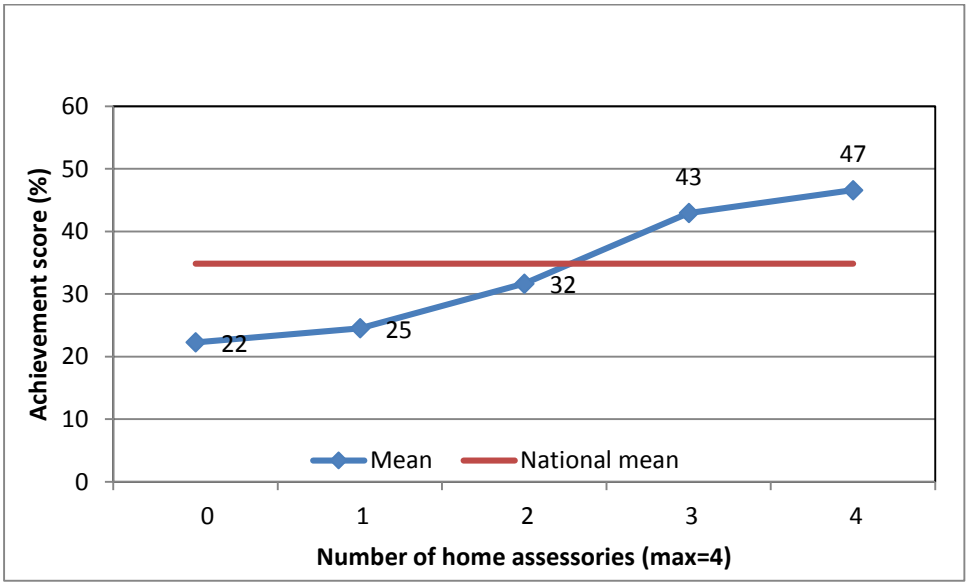


Figure 3.4.10 clarifies how the number of home accessories affects student achievement. Students' achievement is 22% if none of the accessories is available and it is 47% if all of them are available. Availability of all the stated facilities indicates the high SES of the family and therefore it could be linked with SES of the family. Correlation between home accessories and achievement is $r = 0.333$ ($p < 0.001$) which ,is positive, but not very high. The effect size of $d = 0.82$ indicates a high size of difference between the groups. The difference between the lowest group with no mobile phones, television, computer and radio (22%), and the highest group with all these accessories (47%) is remarkable. It should be noted that if radio is removed from the accessories, 64% of those students who have the access to all three accessories are from the Kathmandu Valley.

Data shows that when the children have 0 to 6 home possessions out of the 12, the achievement level is lower than the national average and those having more than six home possessions the achievement is higher than the national average. The average achievement score of students from the families with 11 possessions is the highest (i.e., 44%). The same pattern shows with home accessories as if there is no or up to two accessories, the results are very poor; and when there are three or more accessories, the results are remarkably high.

SES and achievement

The socioeconomic status was analysed based on seven indicators, which were all first dichotomized. The seven variables indicating socio-economic status are: mother's education, father's education, mother's occupation, father's occupation, home possessions, home accessories, and type of school students are studying. Score on each of the variables were added and the total score of each variable was changed into percentage of the maximum score (PSES). PSES represents the percentage of SES the student possesses, 100 percent denotes the student has the highest SES measured with these variables, that is, all the seven indicators of SES are positive, and 0 refers to the lowest possible SES, that is, all the seven indicators of SES are negative. The analysis of the PSES by using Univariate GLM (that is the Regression modelling) shows that there is a strong relation between SES and achievement. Figure 3.4.11 below presents the relationship between SES of the students and the achievement.

Figure 3.4.11 Relation between the SES and achievement in Mathematics

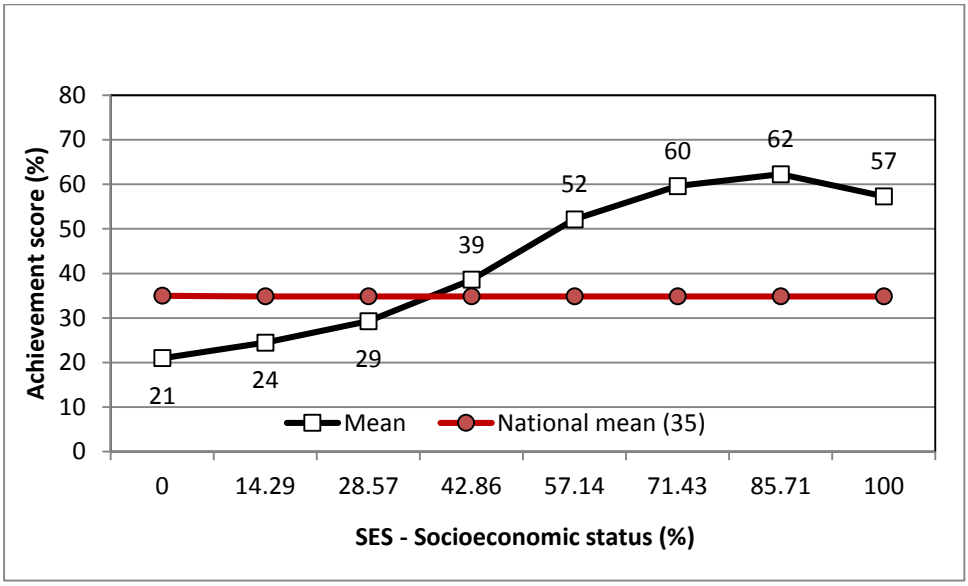
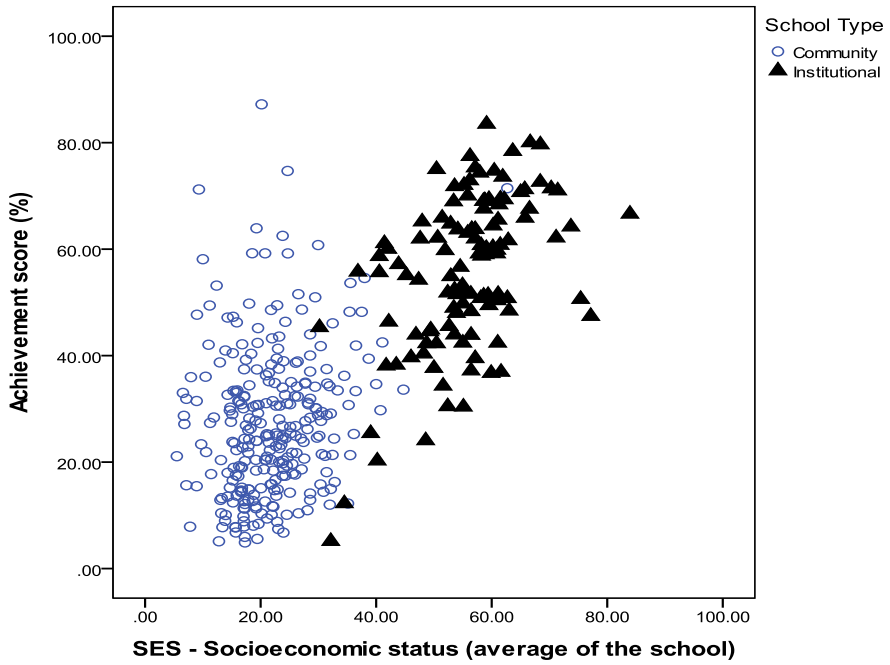


Figure 3.4.11 shows a positive relationship between SES and the achievement; the correlation between the variables is $r = 0.526$ which is a significant association ($p < 0.001$). The differences between the SES groups are statistically significant ($p < 0.001$) and the effect size is high ($f = 0.64$); that is, the highest and lowest groups differ from each other remarkably. SES explains 29% of the student variation in data ($\eta^2 = 0.291$). The DTA suggests that three highest groups should be merged, as there are not many students in the highest SES groups. After combining three highest SES groups, the mean score of the students of the merged group is 60.

The dataset suggests that the socioeconomic status plays a vital role in student achievement. The difference in achievement between the lowest and highest SES groups is remarkable.

Analysing further with the scatter diagram based on SES and achievement, figure 3.4.12 shows that two types of schools (community schools in circle and institutional schools in triangle) fall into two groups. The institutional schools, with relatively higher SES, are concentrated more on relatively high performing group whereas community schools, with relatively lower SES, fall from very high performing to very low performing group with more concentration towards low performing group.

Figure 3.4.12 Distribution of achievement by socio-economic status and type of schools* in *Mathematics



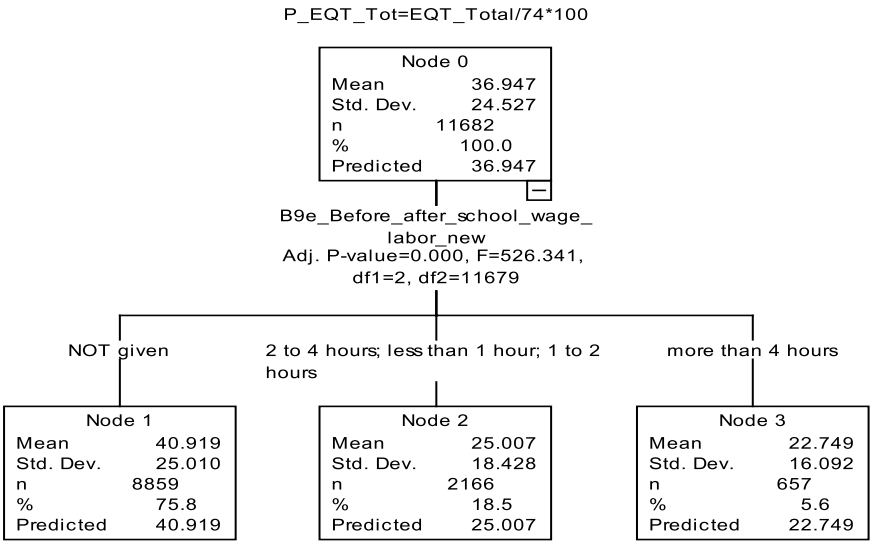
The dataset reveals that the students in the institutional schools, with relatively high SES, concentrated more towards high performing group and the students in the community schools, with relatively low SES, concentrated more towards low performing group.

Working Beyond School Hour and Achievement

There were questions about student background in the questionnaire related to their activities beyond the school hours. The questions were related to working before/after the school hours in a paid job and involvement in household chores. The values of each variable are divided into five categories and marked on a five point scale. The scale was 1 point for working not at all, 2 points for the work of less than 1 hour per day, 3 points for the work of 1 to 2 hours per day, 4 points for the work of 2 to 4 hours per day, and 5 points for the work of more than 4 hours per day. DTA shows that when the children are not engaged in paid work at all, their results are notably above the national average (41) (see figure 3.4.13). If the students work for paid job, even less than an hour to four hours, the results are lower (25%) than the national average. Further, if they were working more than 4 hours, the results were remarkably low (23%). The differences are statistically significant ($p < 0.001$) though the effect size is moderate (f = 0.4) as most of the children do not need to work in a paid job. Working beyond the school hours indicates that the family is poor and

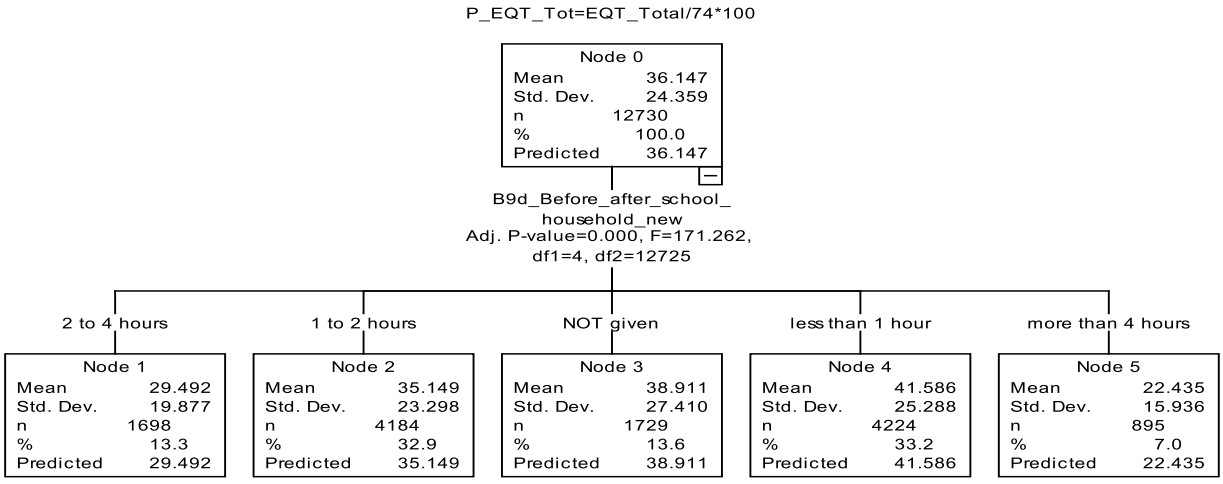
extra earning is needed. When the student needs to work more than 4 hours per day there is no time or energy to be involved in school homework.

Figure 3.4.13 DTA of paid work and achievement in Mathematics



The DTA shows that when the children spend more than 2 hours per day in household chores, the results are poor (see fig. 3.4.14). However, when the amount of time spent in household chores is less than one hour per day, the achievement level is higher than the average (42%). Differences are significant ($p < 0.001$) though the effect size is moderate ($f = 0.23$) as around 18 percent children participate in household work for more than 2 hours per day.

Figure 3.4.14 DTA of household work and achievement in Mathematics



The dataset shows that paid work of the students beyond the school hours reduces the achievement. Getting engaged in unpaid household work up to one hour gives positive effect in student achievement, whereas when the children need to work for more than 2 hours per day either paid or unpaid, the achievement level reduces remarkably.

Student Attitude towards the subject of study and Achievement

The attitude about the subject tells us what the students think about mathematics and its usefulness in their daily life and future utility. The correlation between Mathematics achievement and attitude toward Mathematics is widely studied, and it has been found that there is a relationship between the attitude of the students and achievement though the connection is not always clear (see, for example Metsämuuronen 2012a; 2012b; House & Telese, 2008; Shen & Tam, 2008; Kadijevich, 2006; 2008).

In NASA 2013, the same shortened version of Fennema–Sherman Attitude Scales (FSAS, Fennema & Sherman, 1976) used in several international comparisons, like in TIMSS and PISA studies, was used to identify the relation between attitude towards the subject and their achievement. The original scale included nine dimensions, but in these international comparisons, only three were used with four items on each and two negative items on each of the first two dimensions. The names of the factors can be "Liking Math", "Self-Efficacy in Math", and "Experiencing utility in Math" (compare naming in, e.g., Kadijevich, 2006; 2008). Factor analysis was used to identify the factors of the responses in FSAS and the negative items were reversed to make the whole test unidirectional. As in several countries of Asia, the expected factor structure cannot be found in Nepal (for a deconstruction of the test scales, see Metsämuuronen, 2012a; 2012b). Hence, only the total score is used to show the relationship between attitude and achievement. The relation between the attitude, which is divided into deciles of somehow equal number of students, and achievement score, is shown in figure 3.4.15.

Figure 3.4.15 Relation between students' attitude and achievement in Mathematics

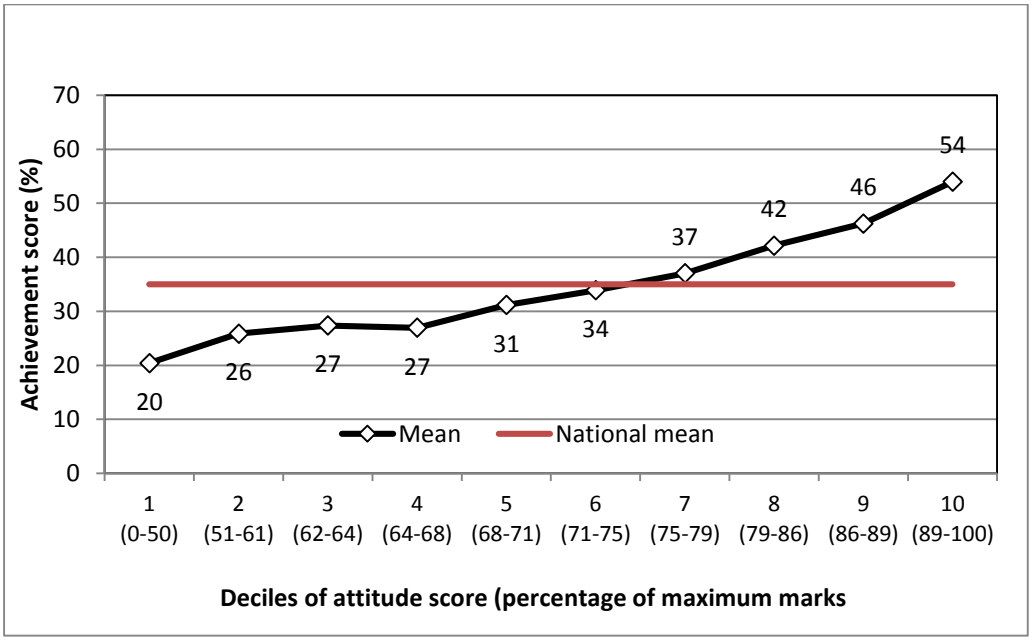


Figure 3.4.15 shows that the difference between the lowest attitude group (20%) and highest attitude group (54%) is remarkable. However, the real difference in achievement is found only in the two highest attitude groups, that is 46% and 54%. The correlation between positive attitude towards Mathematics and achievement is $r = 0.374$ ($p < 0.001$); the effect size is high ($d = 0.95$).

Data reveals that the more positive the attitude is towards mathematics the higher is the achievement. The data also supports the fact that positive attitude influences the positive achievement but not the other way round.

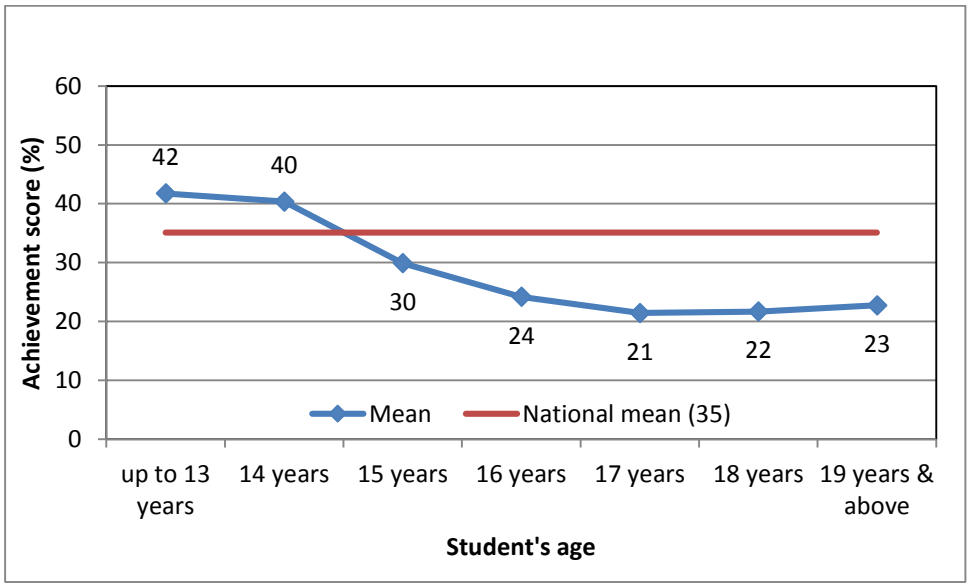
Age of the Student and Achievement

Although official age for grade eight students is 13 year in Nepalese context, the age of the students attending grade eight varies widely. Some students have mentioned their age below thirteen years and some above 20. The age of all the students below 13 were encoded as '13 years', and all students above 19 were encoded as '19 years and above'. The descriptive statistics of the mean in each year are presented in table 3.4.2 and further visualized in figure 3.4.16.

Table 3.4.2 Descriptive statistics of the students' achievement in different age groups

Age	N	Mean	SD
13 years	3111	42	26.1
14 years	4882	40	24.9
15 years	3407	30	21.2
16 years	1530	24	17.4
17 years	507	21	16.3
18 years	209	22	16.7
19 yea	150	23	16.8
Total	13796	35	24.2

Figure 3.4.16 Relation between the age of student and achievement in Mathematics



Data explains that the best achievers are those students who are at the proper age for grade eight, that is, the students of 13 to 14 years. The achievement level is remarkably low when the students are above the age of 15 year. It clearly shows that students with higher age have weaker results. Students of higher age are those who either started school much later than they should have, or repeated the classes. Correlation between the variables is - 0.237 ($p < 0.001$) and effect size is found high ($d = 0.64$). The ANOVA hints that the age explains 8.7% variation of the achievement level.

Dataset indicates that the highest performance is found in the students studying with their correct age group, that is, at the age of 13 and 14 years. Otherwise the achievement decreases as the age of the student increases.

Support for Study to the Student and Achievement

The relation between the support provided for study and achievement was analysed based on the responses to the question: *"who supports you when you do not understand what you have studied or felt difficult?"* In the question, students were asked to select only one option. This assessment did not consider the case where there were more than one supporters. The descriptive statistics of the supporters are given in table 3.4.3.

Table 3.4.3 Descriptive statistics about the support to the students

Who supports	N	Mean	SD
No one	793	36	25.2
Brother/Sister	6006	35	24.2
Father	1578	35	26.6
Teacher	3489	33	23.8
Mother	776	34	26.9
Tuition	2276	38	23.4

Support in mathematics is necessary for students to achieve better than average score in an assessment. There is about 2 percent difference between the students who did not receive any kind of support and those who received (private) tuition. Interestingly, the data shows that those students obtained more score who studied independently than those who received support from their family members. However, students obtained low score among the groups who are dependent on teachers (33%).

The result shows positive correlation between private tuition and student achievement as the achievement of the students having private tuition is higher than the students having family members' support. However, students studying independently without taking any support from family members and teachers obtained relatively better score than those who had received support from them.

Availability of Textbook and Student Achievement

The data shows that there were some students who did not have Mathematics textbook even up to the end of academic session. Table 3.4.4 shows the descriptive statistics of the availability of the textbook of Mathematics and the achievement (mean).

Table 3.4.4 Availability of textbook of mathematics and the achievement

Availability of Mathematics textbook	N	Mean	SD
Yes	13479	36	24.2
No	760	24	21.8
Total	14239	35	24.2

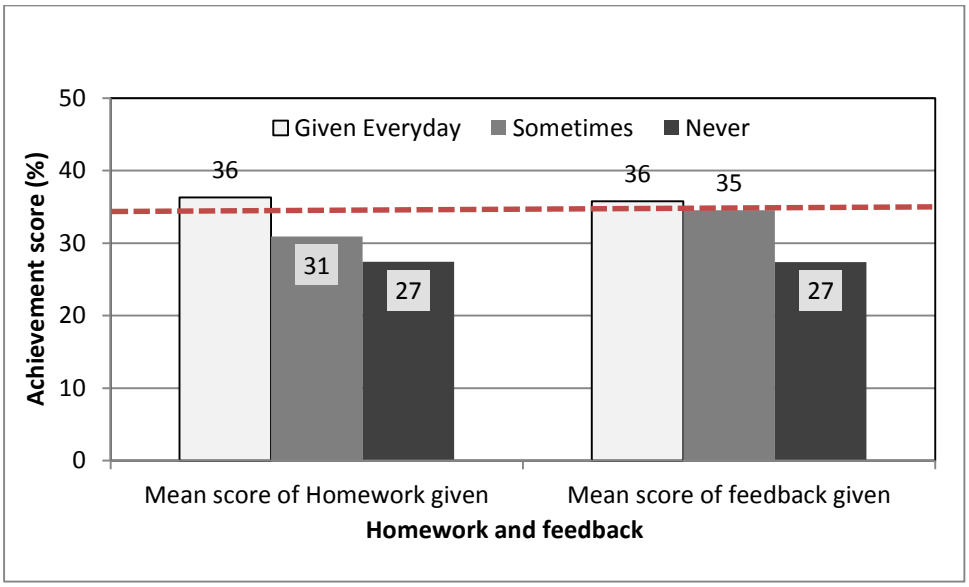
Out of 14,239 students who responded to the question, 5% of the students did not receive a Mathematics textbook even up to the end of the academic session. The relation between the availability of textbook and achievement is significant ($p < 0.001$) though the effect size is small ($f = 0.11$) due to the small group size. The difference in achievement is quite big, that is 12 percent point of the score.

Data shows that 5% of the students did not receive textbook in Mathematics even up to the end of the academic session. The achievement level of the students without textbook is significantly lower than those who have access to the textbook.

Homework and Achievement

Homework is considered as one of the ways to enhance learning which can be used as drill, exercise and an evaluation tool as well. When homework is regularly checked and feedback is provided, it is likely to boost achievement levels. Statistics related to homework given and feedback provided is presented in figure 3.4.17.

Figure 3.4.17 Effect of homework in students' achievement in Mathematics



If teacher assigns homework and provides feedback to the students regularly then the students' achievement is higher (36%). If homework is not regularly assigned and feedback is not provided to them, the achievement score is below the national average (31%). When home assignment and feedback are never given then the average achievement score is remarkably low (27%). The differences are statistically significant ($p < 0.001$). Those groups with no homework assigned are, however, very small and hence, the effect size is small ($\eta^2 = 0.012$ and 0.002).

Dataset gives the evidence that if the teacher assigns homework and provides feedback to the students regularly, the achievement has been higher than without assigning homework or not providing feedback.

Positive and Negative Activities in the School and Student Achievement

The activities of students and teachers determine the learning environment of a school. Bullying, for example, is one of the hindering incidents for the students in school that may affect learning. In the students' background questionnaire, several school related activities were asked to the students. Here, bullying is handled as one of the negative indicators and students' impressions of school and teacher activities are taken as the examples of positive indicators.

Bullying at school

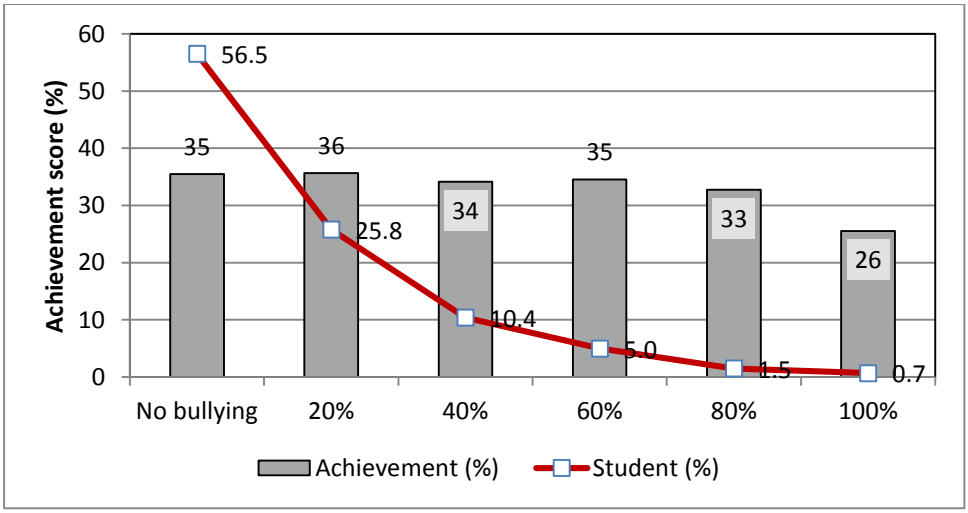
Bullying is one of the problems in schools as it worsens the learning environment for the students. International Studies like TIMSS and PISA emphasize to identify such indicators. In NASA 2013, student

background questionnaire, five questions indicate the varieties of bullying that tend to happen in schools. All the incidences were stemmed by the phrase "*which of the following activities happened in your school last month?*" The students' responses are presented in table 3.4.5 and further visualized in figure 3.4.18. 'No (%)' indicates the percentage of the students' response of no such activity happened in the school and 'Yes (%)' indicates the percentage of the students who experienced the particular type of bullying within the last month. Alone, 24.7% of the students mentioned that, during the last month, something of their own was stolen, which is an alarming sign of the system.

Table 3.4.5 Bullying and the achievement

Type of Bullying	No (%)	Yes (%)
Something of mine was stolen	75.3	24.7
I was made fun of or called names	84.3	15.7
I was hit or hurt by other student(s)	85.7	14.3
I was made to do things I didn't want to do by other students	88.9	11.1
Fellow students kept outside without involving me in activities	90.6	9.4

Figure 3.4.18 Effect of bullying in the achievement in Mathematics



The sum of all five items that is, total of 100 percent, is considered an indicator of highest bullying. Figure 3.4.18 shows the extent of bullying with the percentage of the bullying and achievement of the students in each category of bullying. The line graph shows that the percentage of students encountered various levels of bullying. If only one activity of bullying is reported, it is categorized as 20% bullying, whereas if all five activities are reported, it is categorized as 100% bullying. Data shows that 56.5% of the students did not encounter any kind of bullying during the last month. One can infer from this the remaining 43.5% have

encountered at least one type of bullying, which is a remarkable number of students. About 2.2% of the students are experiencing a severe kind of bullying, which is the sum of the students who encountered 80% and 100% bullying. It is found, however, that learning outcomes are remarkably low only with 0.7% of the students who have encountered extreme bullying including all five types of harassments (26%). Students who did not feel bullying and students who encountered extreme bullying of all five kinds have almost 10 percent point achievement gap although a very small number of the students reported all kinds of bullying (n = 95). However, the difference is statistically significant ($p < 0.001$) though the effect size is small ($f = 0.04$). Though the extreme case of severe bullying is rare, bullying is found quite common in schools.

The dataset reveals that a large number of the students (43.1%) have encountered bullying in schools within the last month. Though the phenomenon does not have a great effect except in the group of extremely bullied students, all possible efforts should be put to root out the phenomenon from schools.

Positive activities at school

The activities that can boost the learning and achievement of students are categorized as positive activities. Such positive activities at the school were asked from the students in two sets of questions listed in table 3.4.6. The table shows that the responses of the students in all four categories, which are in the 4-point rating scale, anchored to fully agree and fully disagree.

Table 3.4.6 Students' response towards teacher and school-related activities in schools

Teacher and Students activities	Respondents in %			
	Fully agree	Partially agree	Partially disagree	Fully disagree
Students get along well with most teachers	75.7	19.7	3.0	1.6
Most teachers are interested in student's well-being	86.0	10.2	2.2	1.6
Most of the teachers really listen to what I have to say	59.0	32.3	6.2	2.4
If I need extra help, I will receive it from my teacher	79.1	16.6	2.8	1.5
Most of my teachers treat me fairly	57.3	23.1	8.1	11.6
I like to come and stay in school	90.0	7.0	1.5	1.5
Students in my school try to do their best	77.0	19.0	2.8	1.3
Teacher in the school care about the students	80.1	15.4	3.1	1.4
Teacher wants the students to do their best	90.6	6.0	1.5	1.9
Average	77.2	16.6	3.5	2.8

It was further analysed by recoding the variables into two categories, that is, 1 for agree and 2 for disagree. Furthermore, the sum of nine indicators is converted into the percentage of maximum score to analyse the

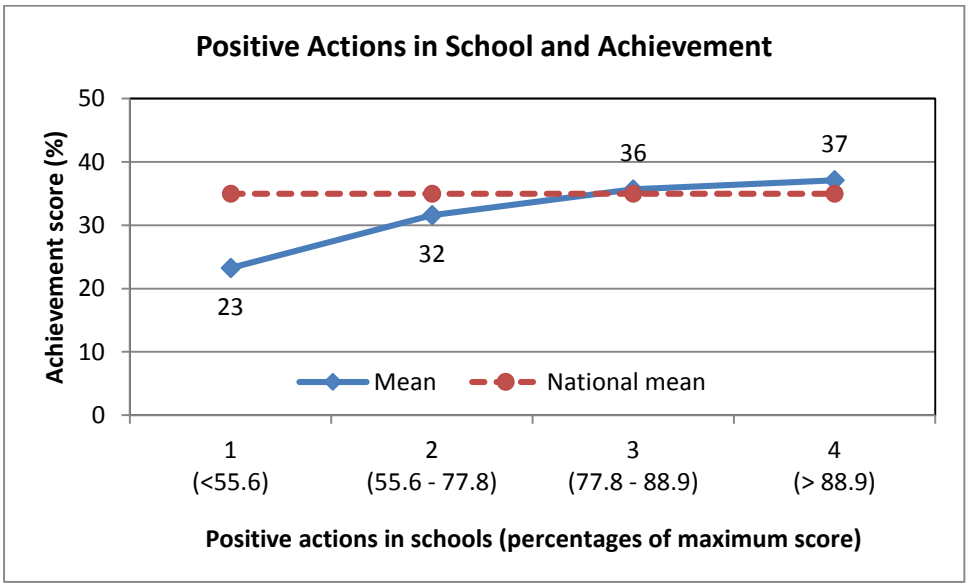
level of positive activities and its relation with achievement. The overall result is that the feeling of the positive actions in the school relates positively with the student achievement. The correlation between the sum of nine positive activities and achievement is positive ($r = 0.166$, $p < 0.001$).

DTA finds four attitude groups in the indicator. These boundaries and descriptive statistics are presented in table 3.4.7 and further illustrated in figure 3.4.19.

Table 3.4.7 Students' response towards teacher and school-related activities in the schools

% of total score	N	Achievement	SD
0 -- 55.6	1149	23	20.8
55.6 -- 77.8	1769	32	24.6
77.8 -- 88.9	3209	36	24.2
88.9 -- 100	8164	37	24.0
Total	14291	35	24.2

Figure 3.4.19 Relation between positive actions at school and achievement in Mathematics



The data shows that there is a positive connection between the students' feeling about the school activities and achievement. After dividing the indicator into four groups based on DTA, the differences between the groups are statistically significant ($p < 0.001$); however, the effect size is small ($f = 0.17$). The learning achievement is higher than the average only when the students are extremely positive towards school and

teachers' behaviour. However, the difference between the most positive and the most negative groups' achievement is notable, that is 11 percent.

Dataset records that when students feel that the actions of the teachers and the schools are ultimately good, the results are better than average. At the other extreme, if they feel that such actions are negative, the results are far below the average.

3.5 Summary of Findings

The main findings on **Mathematics** subject at grade eight in NASA 2013 are as follows:

Basic Results

- The average achievement score in Mathematics is 35%, and the achievement in Mathematics is not normally distributed as there were three populations: low-performing, medium-performing and high-performing students, but the larger number of students is concentrated towards the low performing area.
- The learning achievement is poorer in Algebra (27%) and Geometry (34%) than in other content areas.
- Students' ability to solve complex problems in Mathematics is low. Students are good in fundamental operations, the basic manipulation of data and numbers, and calculations with few steps. They are much weaker in reasoning, problem solving, plotting, proving the theory or formula, and constructing the shape and figures.
- In all content areas, the results are lower in NASA 2013 than NASA 2011.

Equity Indicators

- Boys are slightly out-performing girls in all content areas of Mathematics and the differences are significant.
- The results are far better in urban areas than in the rural area, and the difference is remarkable. Similarly, the results in 23 districts out of 28 sample districts are below the national average (35%). The Kathmandu Valley has outperformed all the other districts.
- Of the Developmental Region, the difference between the lowest performing region (Mid-Western region) and the highest performing region (Kathmandu Valley) is remarkable, that is, more than 30 percent.

- The students in the institutional schools have performed better in comparison to community schools. However, variation in achievement scores between the community schools is remarkable.
- There are wide differences in achievement among the language and ethnic groups, for example, students from Newar community performed better in Mathematics (47%). On the other hand students from Rai (20%), Sherpa (22%), Tharu (22%), Limbu (23%), and Magar (27%) communities scored lower in Mathematics.
- As a whole, the socioeconomic status (SES) has played a strong role in the achievement in Mathematics.

Selected explanatory factors

- The educational level of the parents predicts the children's future achievement level in Mathematics. Especially challenging is the situation when both parents are illiterate.
- If the father, mother or both are from an agriculture or related occupation, the students' achievement in Mathematics is significantly lower than the other occupational groups.
- The difference in Mathematical achievement between the lowest and highest SES groups is remarkable.
- When children need to work more than 4 hours per day (either paid or unpaid) in the household work, the achievement level is remarkably low (23%). Paid work of any duration has negative effect on student achievement, but engaging students in household chores for a short time (upto 1 hour per day) shows positive effect on student achievement.
- Around 5% of the students lacked the textbook in Mathematics; achievement of those students without having textbook is lower in comparison to the students having textbook.
- Students in a large number (43.1%) have encountered bullying in the school within the last one month of the test. Bullying has negatively affected the achivement of students in Mathematics; as the higher the bullying at school to a student, the lower the achievement of the student.

Chapter 4: Analysis of Student Achievement in the Nepali Language

Nepali language is one of the compulsory subjects in the school curriculum as it is assigned six out of 40 credit hours a week. This assessment is based on the learning outcomes/ curricular competencies as set in approved curriculum for grade eight students in Nepali language subject. Overall, the objective of grade eight Nepali is to develop basic language skills and creativity in students.

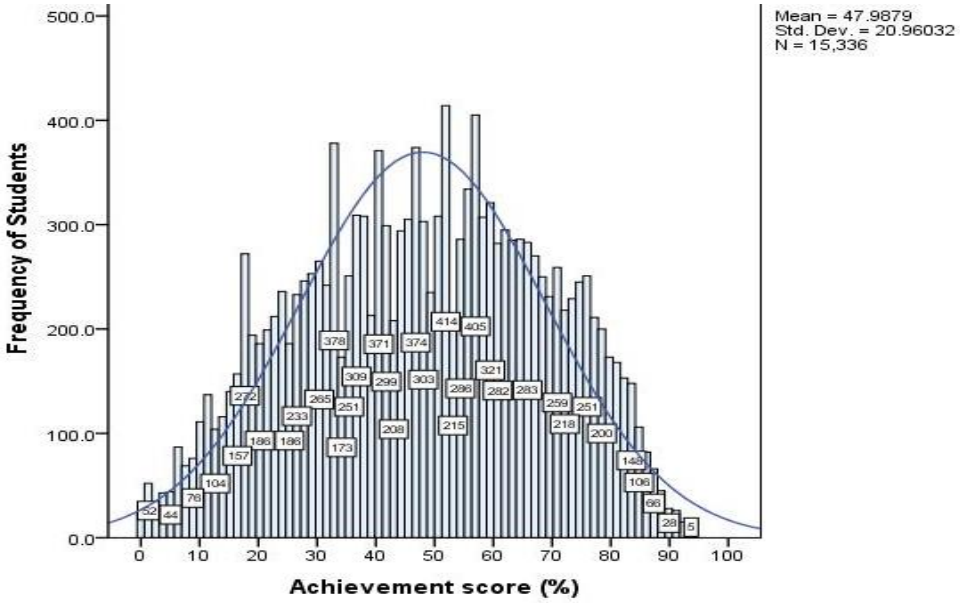
This chapter presents the analysis of assessment results of grade eight Nepali language subject. As the basic result in Nepali language, this chapter presents the overall distribution of achievement scores. It further presents achievements in various content areas, various levels of cognitive domain and objective and subjective type of items. The second section of this chapter presents the comparison of NASA 2013 results in Nepali language with NASA 2011. Results with various diversity factors, including district, ecological zone, development region, school type, school location, language at home, and caste/ethnicity are included in section three of this chapter. Section four deals with the selected explanatory factors about students' achievement in grade eight Nepali language. The factors that explain achievement in Nepali subject are: parents' education and occupation, home possessions and accessories, SES, working beyond school hour, age of the student, support for study, availability of textbook, homework given and checked/ feedback provided, activities at school (both negative and positive). The final section summarizes the findings related to the achievement in Nepali language.

4.1 Basic Achievement Results in Nepali

Overall Distribution of Achievement Scores

Achievement score is usually distributed normally in a large sample in the students' achievement study. As shown in the figure 4.1.1, the achievement score in Nepali language subject is distributed normally.

Figure 4.1.1. Distribution of overall achievement scores in the Nepali language

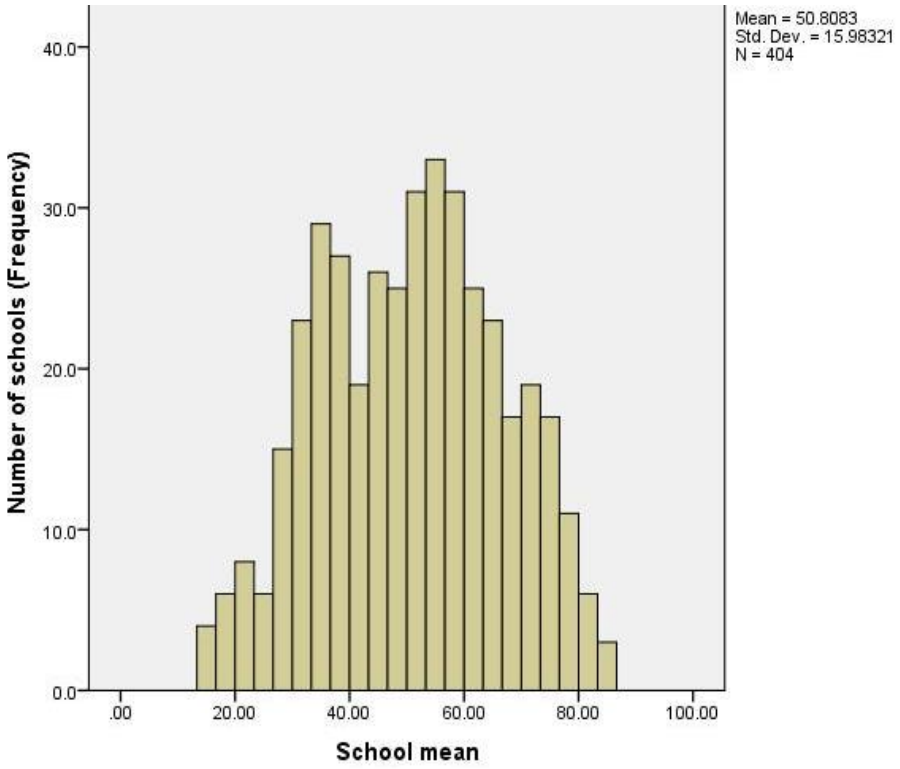


The sample size for Nepali subject (15,336 students) was big enough to form a normal distribution. In the final score of Nepali subject, the majority of the students lie slightly in the medium-performing part of the distribution. Negligible number of students falls in the group of achieving more than 90% whereas significant number of student falls in the group achieving below 10%.

The dataset indicates that the grade eight population in Nepali is normally distributed.

As illustrated in figure 4.1.2 given below, the schools can be divided into two categories – namely, the high performing and low performing.

Figure 4.1.2 Distribution of schools' mean achievement scores in the Nepali language



Students' achievement score of the sample schools as presented in figure 4.1.2 shows that there are two categories of schools, namely high performing and low performing. The maximum average score of schools on the left hand side is about 35%, whereas it is about 70% for the schools at the right-hand side. It indicates a remarkable difference in achievement score between the high and low performing schools.

Student Achievement in Various Content Areas of the Nepali Language

As per the curriculum, the Nepali language test includes four content areas, namely, 1) *Reading*, 2) *Writing*, 3) *Grammar*, and 4) *Vocabulary*. The weightage in four content areas of the assessment were proportionally equal to the weightage allocated in the curriculum.

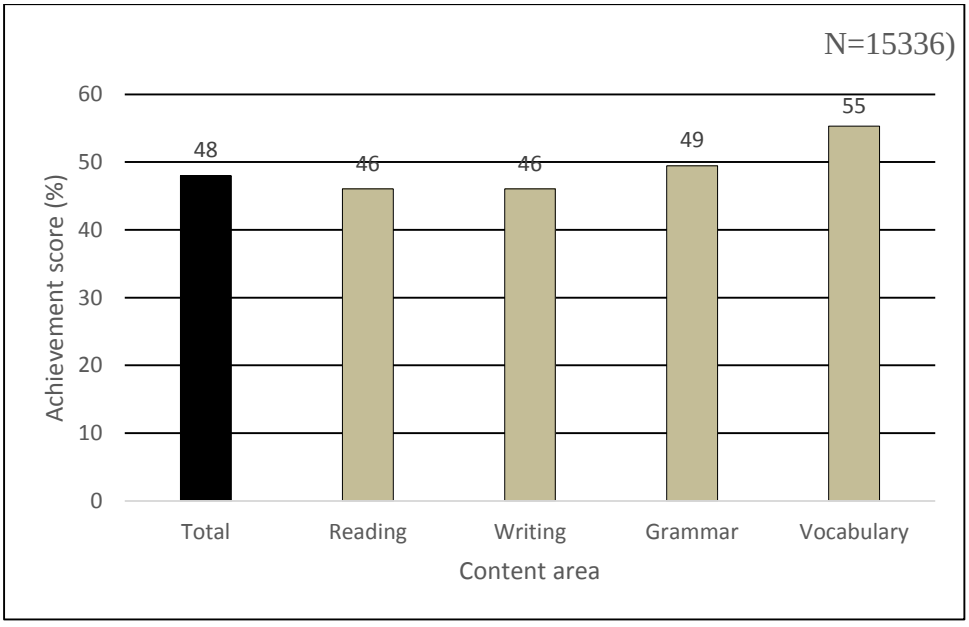
For comparability, the achievement in all of the content areas were converted into percentage. Table 4.1.1 shows the students' achievement in Nepali in total and the achievement level in each of the four content areas. Figure 4.1.3 further illustrate the comparison of achievement scores in various content areas.

Table 4.1.1 Achievement in various content areas of the Nepali language

Content areas	Mean	SD	Min	Max
Reading	46	24.5	0	100
Writing	46	22.0	0	100
Grammar	49	21.4	0	100
Vocabulary	55	26.1	0	100
Nepali total	48	21.0	0	100

The table shows the variations of achievements in various content areas of Nepali. The achievement ranges from 46 percent in reading and writing to 55 percent in vocabulary, which is 9 percent variation. Comparing the maximum and minimum scores, the situation is the same in all content areas, as the maximum score is 100 and the minimum is 0 in each of the content areas. Figure 4.1.3 compares the variation of achievement in various content areas.

Figure 4.1.3 Comparison of achievement scores in various content areas of the Nepali language



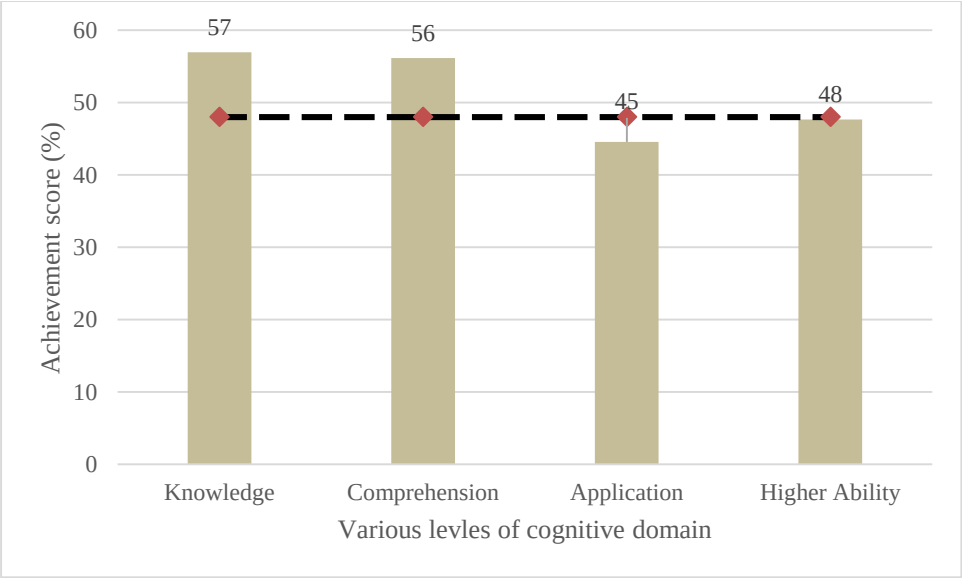
The overall national average achievement score for Nepali language is 48. Of the various content areas, students are found weakest in reading and writing (46), but in vocabulary (55) and grammar (49) the performance is better than the national average.

Dataset indicates that the learning achievement in Nepali language is the highest in vocabulary and the lowest in reading and writing.

Achievement in Various Levels of Cognitive Domain

The Nepali language test items are categorized into various levels of cognitive domain according to Bloom's taxonomy (Bloom *et al.*, 1956; Metfesser, Michael & Kirsner, 1969). The categories were knowledge, comprehension, application and higher ability (reasoning/problem solving/creativity). The achievement of students for each level is shown in figure 4.1.4.

Figure 4.1.4 ***Comparison of achievement score in various levels of cognitive domain in the Nepali language***



The score in application level items is below the national average, whereas the scores obtained in both the comprehension and knowledge level items are above the national average. One of the remarkable result is that students obtained 3 percent higher score in higher ability items than the application level items. However, student performed better in items of lower level of cognitive domain than higher level.

Going through the dataset, it is found that a large number of students were able to solve only 15% or less problems on higher ability. Ten percent of the students could solve just less than 15% of the problems requiring the higher cognitive abilities. Similarly, only 11% students were able to solve all the application type of items.

The dataset shows that the students' ability to solve application and higher ability items is quite low (45% and 47%) in comparison to comprehension (56%) and knowledge (57%), that is, students are much better in the recall type of questions than in analytical questions.

Student Achievement by Type of Items

There were two types of items in the test: objective and subjective. Objective items covered a wide range of content areas and were very specific to judge because there was only one correct answer or one explicit piece of information required to get the correct answer in a question. There were some subjective items in each test version, which require a longer procedure to get the full marks. Both the objective and subjective types of items were based on various levels of cognitive domain, which include knowledge, comprehension, application, and higher ability, covering various difficulty levels. Table 4.1.2 presents the basic statistics on item type and achievement.

Table 4.1.2 Achievement by type of items

Type of items	Mean	SD	Min	Max
Objective	58	22.0	0	100
Subjective	45	23.5	0	98

Students' mean score in subjective items is much lower (45%) than that of objective items (58%). Most of the objective items were of knowledge, comprehension and application type whereas subjective items were mostly application and higher ability type.

Dataset clearly explains that the students are performing well in recognizing the correct answer and in recalling simple facts from the texts, comprehending the basic information from paragraph, table, chart, and a few steps of logical thinking. They are weaker in writing free texts or letters, or preparing synthesis and abstracts from a text. In many cases, the students attempted the open-ended task like free writing, problem solving and analysis, but the skills were not high enough for obtaining highest marks.

4.2 Comparison of NASA 2013 Achievement Results with NASA 2011

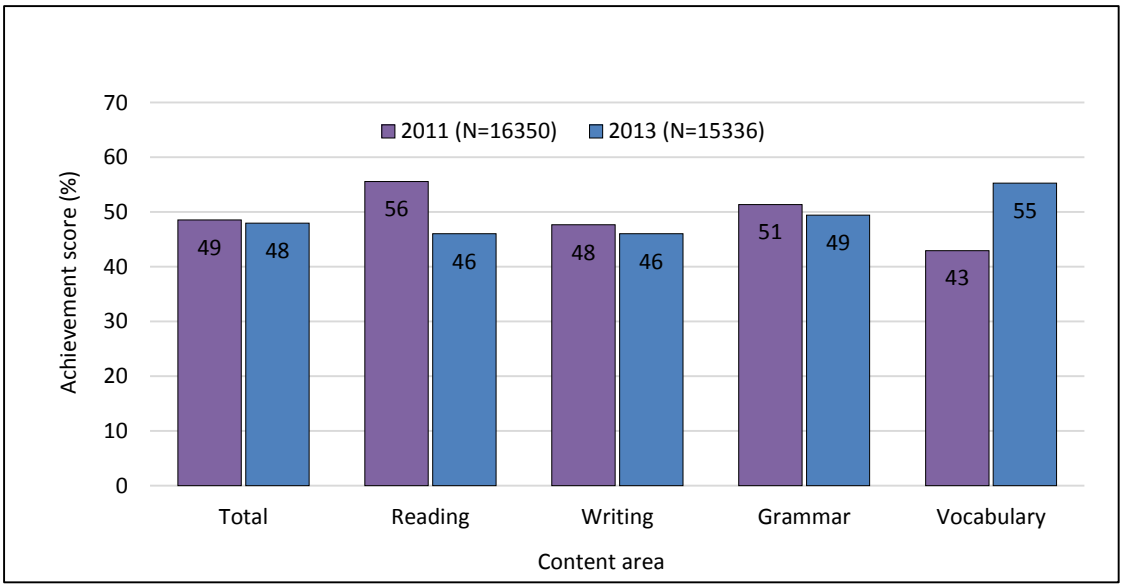
The results of NASA 2013 and 2011 are as presented in table 4.2.1 for the purpose of comparison. The achievement in the year 2013 is found just 1 percent low (round figure) than the earlier one.

Table 4.2.1 comparing the achievement in 2011 and 2013

Year	*N*	*Mean*	*SD*	*SE*
2011	*16350*	*48.6*	*19.3*	*0.15*
2013	*15336*	*47.9*	*21.0*	*0.17*

The result in Nepali as a whole and the achievement in various content areas between NASA 2013 and NASA 2011 are presented in figure 4.2.1.

Figure 4.2.1 Comparing results of NASA 2011 and 2013 in various content areas of the Nepali language



As shown in figure 3.2.1, the achievement level in 2013 is slightly lower than that of 2011. Except in the vocabulary, achievement in other content areas in NASA 2013 is lower than that of NASA 2011. However, direct comparison is difficult as the sample students and districts were not the same as previous one.

The dataset indicates that the achievement score of students in Nepali has not been improved.

4.3 Achievement Scores by Diversity Factors

Diversity is a relative and contextual term. In the context of NASA, six diversity factors have been considered - namely geographical/ecological, regional, language, gender, ethnic/caste and economic diversity. NASA 2013 background information questionnaire included questions related to above six diversities. However, this assessment also considered three additional comparisons. They are by districts,

by school type (community/institutional) and by school location (rural/urban). These comparisons are carried out to assess the equity status of students based on achievement scores.

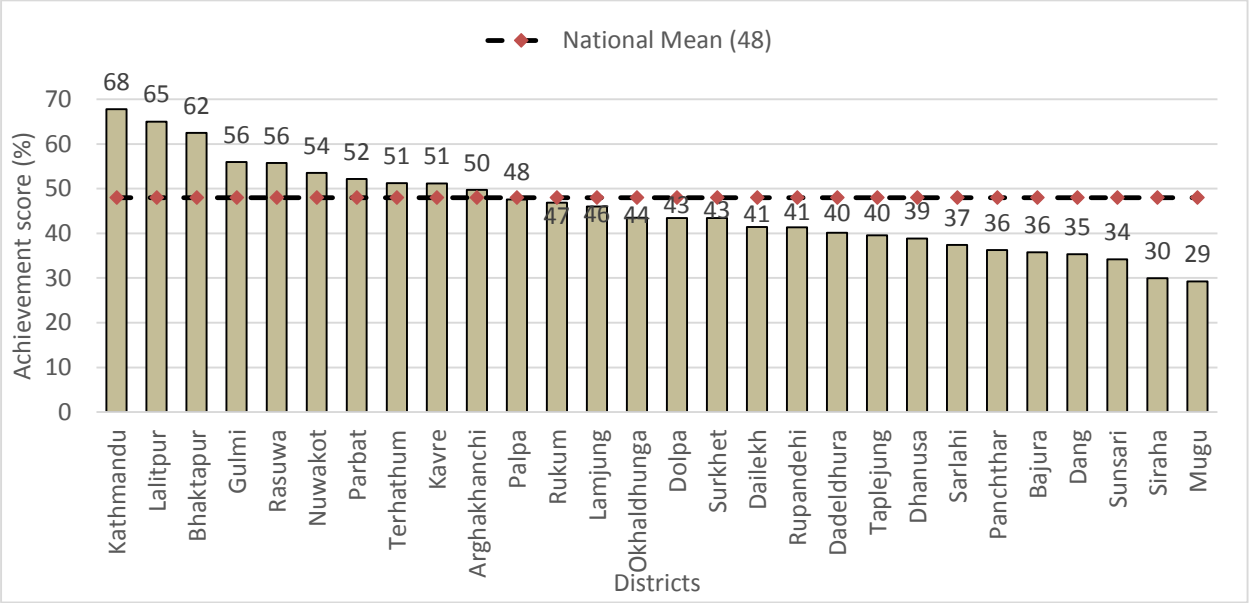
Student Achievement by Districts

Out of 75 districts, 28 were randomly selected to represent the ecological zones and developmental regions. The achievement of students by district is presented in table 4.3.1 and in figure 4.3.1. The table shows the achievement in descending order according to the achievement. The mean represents the average achievement percentage of the particular district.

Table 4.3.1 Average achievement score in sample districts

Districts	N	Mean	SD	Districts	N	Mean	SD
Kathmandu	2230	68	15.7	Dolpa	86	43	18.9
Lalitpur	914	65	15.2	Surkhet	744	43	16.9
Bhaktapur	438	62	17.6	Dailekh	497	41	18.6
Gulmi	508	56	14.1	Rupandehi	1210	41	18.3
Rasuwa	101	56	13.3	Dadeldhura	394	40	14.4
Nuwakot	536	54	15.3	Taplejung	294	40	16.2
Parbat	297	52	17.2	Dhanusa	438	39	21.7
Terhathum	144	51	15.3	Sarlahi	356	37	20.7
Kavre	811	51	17.6	Panchthar	595	36	17.4
Arghakhanchi	304	50	17.4	Bajura	339	36	19.0
Palpa	577	48	15.9	Dang	715	35	18.3
Rukum	376	47	18.4	Sunsari	1142	34	17.9
Lamjung	344	46	15.7	Siraha	516	30	19.6
Okhaldhunga	310	44	18.6	Mugu	120	29	16.3
			Total		15336	48	21.0

Figure 4.3.1 Comparison of average achievement of sample districts



Above data shows that students' achievement was very low in Mugu (29) from Mid-western development region, Siraha (30) and Sunsari (34) from Eastern development region, Dang (35) from Mid-western development region and Bajura (36) from Far-western development region. The best performing three districts Kathmandu (68), Lalitpur (65) and Bhaktapur (62) come from the Kathmandu Valley. The difference was very wide (39%) between low performer (Mugu) to high performer (Kathmandu).

The difference in achievement scores between low and high performing district is statistically significant ($p < 0.001$). The district explains 32% of the variation in achievement ($\eta^2 = 0.326$). Effect size is high ($f = 0.70$), indicating that the difference between the lowest performing and highest performing district is significant.

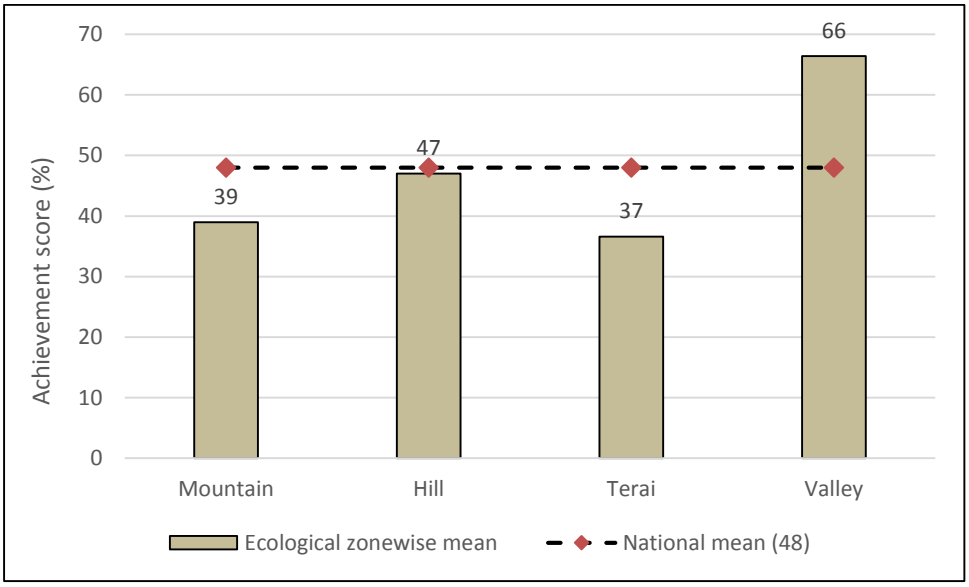
The dataset indicates that there is a wide difference in achievement in Nepali among the districts. The results are bound to the randomly selected 28 districts. The results in Mugu (29%), Siraha (30%), Sunsari (34%), Dang (35%), Bajura (36%), Panchthar (36%), Sarlahi (37%), and Dhanusa (39%) are comparatively low.

Student Achievement by Ecological Zone

The Mountain, Hill and Tarai are three ecological zones in Nepal though the Kathmandu Valley can be taken as a special geographical area because of being the most densely populated area in the country with more opportunities than other areas. From not only the population point of view, but also the factors of

mixed ethnicities, weather conditions, economic activities, urbanization as well as the dense human capacity have been considered in the case of the Kathmandu Valley as a unique fourth geographical area in the analysis. The variation in the ecological zones in NASA 2013 is presented in the figure 4.3.2.

Figure 4.3.2 Comparison of achievement of various ecological zones in the Nepali language



The above data shows that, on average, the students from the Kathmandu Valley (66%) have outperformed the students from the rest ecological zones whereas the students from the Tarai area performed the lowest (37%). It is noteworthy that some of the lowest performing districts, such as Siraha (30%), Sunsari (34%), Dang (35%), and Sarlahi (37%) are from Tarai.

The achievement across the regions differs significantly ($p < 0.001$) as the Tukey's *post hoc* test tells that all the zones deviate from each other in a statistically significant manner at $p < 0.001$ level. The effect size f = 0.62 shows great difference between the highest and lowest performing ecological zones. Ecological zone explains 27% of the variation in the data.

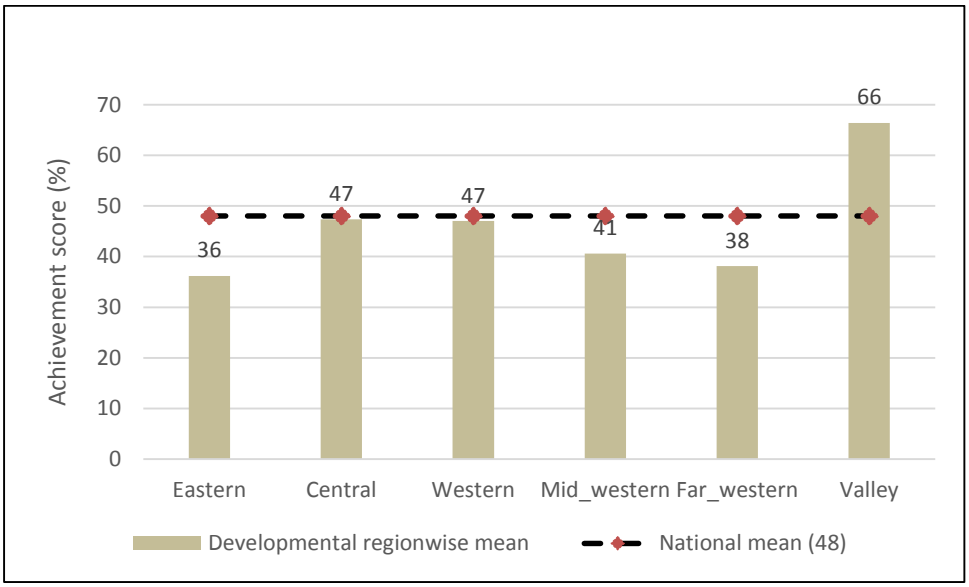
Dataset indicates that there is a wide difference in the students' performance among the four ecological zones. Students in the Kathmandu Valley outperform the other students. The achievement is the lowest in Tarai region.

Student Achievement by Developmental Region

Students' achievement varies according to the developmental regions, which are divided into Eastern, Central, Western, Mid-Western, and Far-Western regions. Additionally, the Kathmandu Valley has been

considered as the additional region though administratively it falls under the Central Developmental Region. The mean achievement in each development region is given in figure 4.4.3.

Figure 4.3.3 Comparison of the achievement of various developmental regions in the Nepali language



The above data shows that the highest achievement is found in the Kathmandu Valley (66%), which is 19 percent higher than the region having nearest achievement level (47%) in the Western region and Central region (47%). The achievement is the lowest particularly in the Eastern region (36%), while the achievement in the Far-western (38%) and Mid-western (41%) regions are higher than the Eastern region. The Central and Western regions are found close to the national mean. The difference between the regions is statistically significant ($p < 0.001$) as Tukey's *post hoc* test shows that excluding Central, Far western and Western regions other development regions differ from each other significantly with at least at the 1% significant level. Development regions show 27% of the variation in data and its effect size is high ($f = 0.62$).

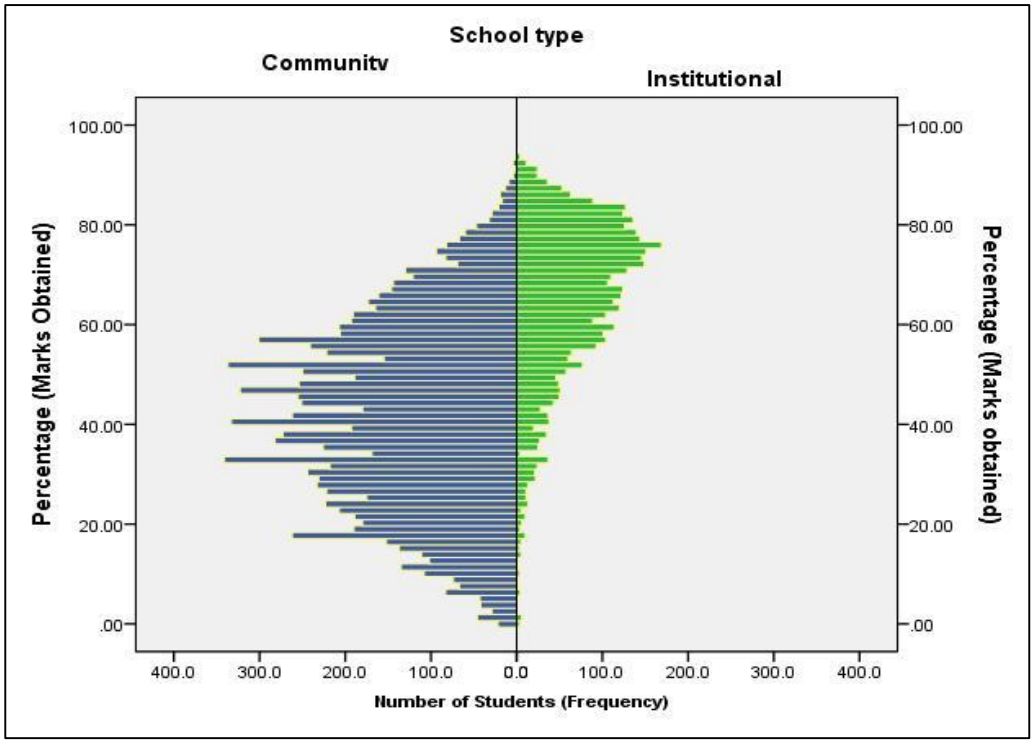
The dataset indicates that the difference between the lowest performing region (Eastern, 36%) and high (Western and Central, 47%) is significant, whereas, the highest performing region-the Kathmandu Valley (66%) is remarkably higher by 30 percent.

Student Achievement by School Type

One way of categorizing schools of Nepal is based on ownership and management. On this basis there are two types of schools: community and institutional. Community schools are generally

funded by the government and managed by the community, and generally known as public schools, and the institutional schools are generally known as private schools. Overall distribution of results in community and institutional schools is given in the figure 4.3.4.

Figure 4.3.4 Distribution of the students' mean score in community and institutional schools



The left hand side distribution shows the community school students and the institutional school students on the right hand side. There are a number of the students in the community schools getting equally high marks as in the institutional schools. Figure 4.3.4 explains that the students in the community schools vary from low performers to the high performers whereas the students from the institutional schools are concentrated more on relatively high performing group.

The difference between the achievement of students from community and institutional schools are presented in table 4.3.2.

Table 4.3.2 Student achievement by type of school

Type of school	N	Mean	SD	SE
Community	11252	42	18.9	0.18
Institutional	4084	65	16.3	0.25
Total	15336	48	21.0	0.17

The achievement gap between the community and institutional schools is remarkably big. The average achievement score for institutional schools is 65% whereas for community schools it is only 42%, a difference of 23 percent. The difference is statistically significant ($p < 0.001$) and the effect size is high ($f = 0.56$), showing a wide difference between the community and institutional schools. Category of the students at the community and institutional schools explains 24% of the variation in student achievement ($\eta^2 = 0.24$). The variation within the community schools is also remarkable, ranging from 20% to 80%, but the variation is relatively smaller in most private schools (see table 4.3.2).

The dataset indicates that, on an average, the students in the institutional schools outperform the students in community schools.

Student Achievement by Location of Schools

One of the strata of sampling in NASA 2013 was school location. The schools were categorized into rural and urban schools. The achievement of students in rural and urban schools is presented in table 4.3.3.

Table 4.3.3 Location of school and student achievement

Type of school	N	Mean	SD	SE
Rural	11932	45	20.6	0.19
Urban	3404	57	19.4	0.33
Total	15336	48	21.0	0.17

The achievement level of the students from urban schools (57%) is remarkably higher than that from rural schools (45%). It tells that the institutional schools and schools from the Kathmandu Valley contributed to raise the achievement level of the urban schools. Excluding the schools from the Kathmandu Valley, the mean achievement of rural and urban schools is found 41% and 48% respectively. The difference in average score is significant ($p < 0.001$) and the effect size is moderate ($f = 0.25$), and the location of schools explains 5% of variation in student achievement ($\eta^2 = 0.058$).

The above data indicates that the students in the urban schools obtained 12 percent point more score than the students in the rural areas.

Language at Home and Student Achievement

Dataset shows that 28.5% of the eighth graders speak a language other than Nepali as their first language. These "other" languages are quite fragmented; the largest groups in the student dataset are Tamang (4.9%), Tharu (2.7%) and Newari (2.3%). After categorizing the languages into ten groups (excluding Nepali), the remaining 15.2% of the students who are classified into the group "Others" for the convenience of statistical analysis, all the other language speakers were grouped into "non-Nepali". The results are presented in tables 4.3.4 and 4.3.5 and further illustrated in figure 4.3.6.

Table 4.3.4 Home language and Student achievement

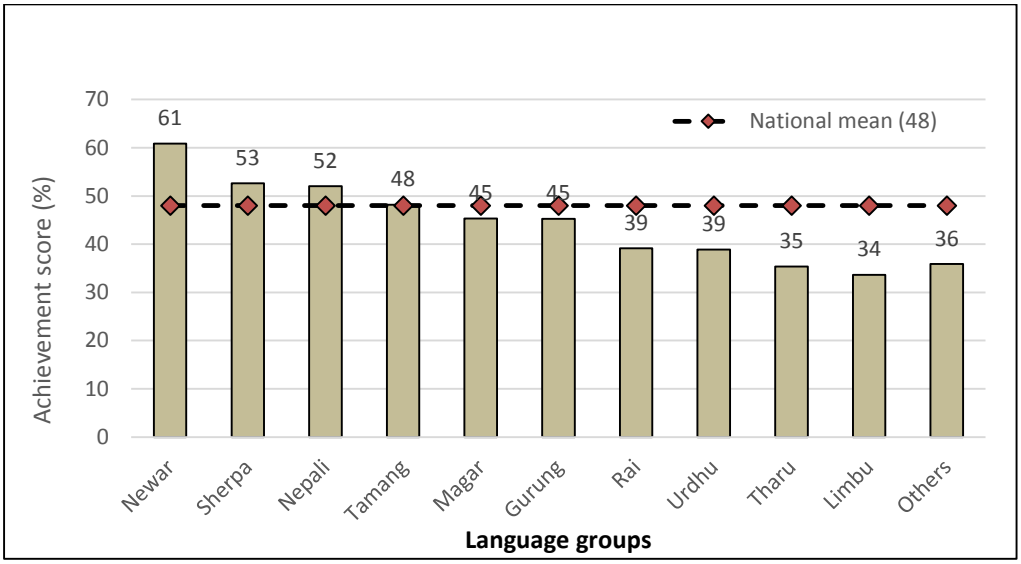
Language group	N	Mean	SD	SE
Nepali	10394	52	20.2	0.19
Non-Nepali	4370	41	19.3	0.29

When combining all the minor language groups as "Non-Nepali", there is a notable difference between the language groups. However, based on the table 4.3.5 it is evident that the students from Newari (61%) and Sherpa (53%) community have performed higher comparing the students from Nepali speaking community (52%). On the other hand, the students from Limbu (34%), Tharu (35%), "Others" (36%), and Urdu (39%) background performed much lower comparing the students from Nepali language speaking communities.

Table 4.3.5 Student achievement of various language groups

Language group	N	Mean	SD	SE	Minimum	Maximum
Newari	344	61	15.6	0.84	11	92
Sherpa	39	53	20.5	3.28	18	86
Nepali	10394	52	20.2	0.20	0	94
Tamang	718	48	16.6	0.62	0	91
Magar	311	45	16.4	0.93	0	87
Gurung	78	45	14.8	1.68	18	81
Rai	30	39	21.8	3.99	1	81
Urdu	11	39	17.1	5.15	0	62
Tharu	407	35	17.4	0.86	0	84
Limbu	187	34	16.9	1.24	1	81
Others	2244	36	18.7	0.40	0	90

Figure 4.3.5 Comparison of achievement of various language groups in the Nepali language



The difference between the language groups is statistically significant ($p < 0.001$), but the effect size is medium ($f = 0.34$) because the minority groups are small and the division into smaller language groups explains somewhat 10% of the variation in the data ($\eta^2 = 0.105$). When analysing only the minority languages by excluding the Nepali language groups and the group "Others", the effect size is high ($f = 0.40$) indicating a remarkable difference between the highest performing group (Newar, 61%) and the lowest performing group (Limbu, 34%). One of the notable results from the above figure is that in Nepali subject the students from the community having home language Newari and Sherpa have outperformed the students from Nepali speaking communities; however, the results are different in other language groups as their performance is lower than that of students having Nepali as home language. In this context, it is difficult to predict the result in the Nepali language based on the different language speaking groups.

Ethnicity/Caste and Student Achievement

Nepali education system has been influenced in several ways by the caste system. There are caste/ethnicity based variations in the participation in and achievement of education in Nepal. Historically, the Brahmans and Chhetris have been heavily involved in education, but Dalits, for example, have practically remained outside or participated less in the educational system. Hence, the government has made lots of efforts to make the education possible and accessible for all children regardless of their caste and ethnicity. The recent National Population Census 2011 shows that the enrolment of Hill Dalits has increased remarkably at the lower level of schooling but their number at the secondary and higher education is still very small (CBS,

2012). However, still there are differences in the achievement of various caste and ethnic groups, which is presented in table 4.3.6.

Table 4.3.6 Student achievement by ethnic/caste background

Caste	N	Mean	SD	SE	Minimum	Maximum
Brahman/Chhetri	5690	54	20.1	0.27	0	94
Janjati	4982	49	19.8	0.28	0	94
Dalit	1493	42	18.7	0.49	0	92
Madhesi	1396	37	20.2	0.54	0	89
Minorities	57	38	18.9	2.50	4	85
Others	1050	47	21.3	0.66	0	92

The above data shows that the Madhesi students are performing the lowest (37%) in Nepali language. The overall difference between the groups is statistically significant ($p < 0.001$) and the effect size is medium ($f = 0.27$). Caste/ethnic background explains 6.8% variation in achievement ($\eta^2 = 0.068$). Apart from this, the mean of the various caste/ethnic groups differ from each other statistically at 5% significance level ($p < 0.05$).

A positive sign from equity viewpoint is that the Dalit students have performed remarkably better than the national mean (48%) in the Central Mountain (51%) and Western Hill (50%) (see table 4.3.7). However, the results are much lower than the average in other regions and very poor in Eastern and Mid-Western Tarai area (29%). The number of students in certain strata is small, and hence it may be wise not to make the too strong implications of the results. However, in the Kathmandu Valley Dalits performed exceptionally high (63%).

Table 4.3.7 Dalit students' achievement in different ecological and development regions

	Eastern	Central	Western	Mid-Western	Far-Western	Valley
Mountain	43	51	-	35	36	
Hill	42	47	50	42	35	
Tarai	29	39	37	29	-	
Valley						63

Figure 4.3.7 Madhesi students' achievement in various content areas of the Nepali language

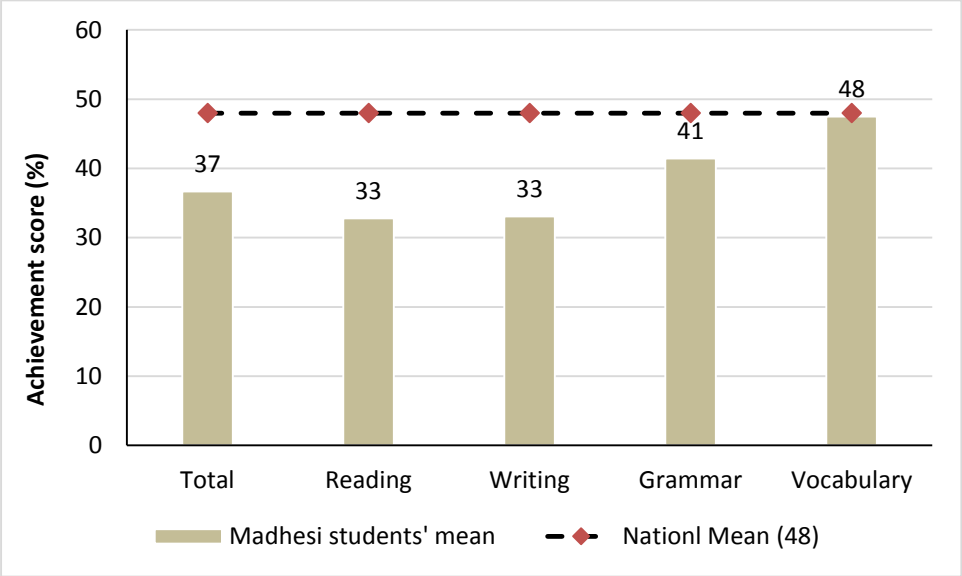


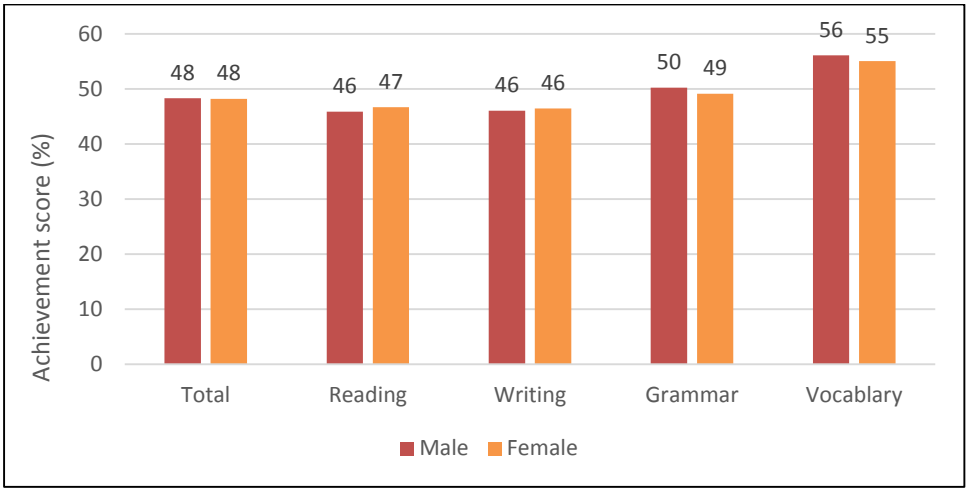
Figure 4.3.7 shows that the Madhesi students have the same mean as the national mean in vocabulary, whereas, performance in reading and writing is found poor compared to the national mean. That is, a notable percentage of Madhesi students have not achieved the required level of performance in reading, writing and grammar in Nepali language, but they are good in vocabulary.

Dataset indicates that the Madhesi students' performance in Nepali language is lower than the national average except in vocabulary.

Achievement by Gender

Basic achievement results of boys and girls are presented in figure 4.3.8. This figure further compares the achievements of girls and boys in various content areas. The figure shows that remarkable variations are not found in various content areas of Nepali language.

Figure 4.3.8 Comparison of achievement of boys and girls in different content areas

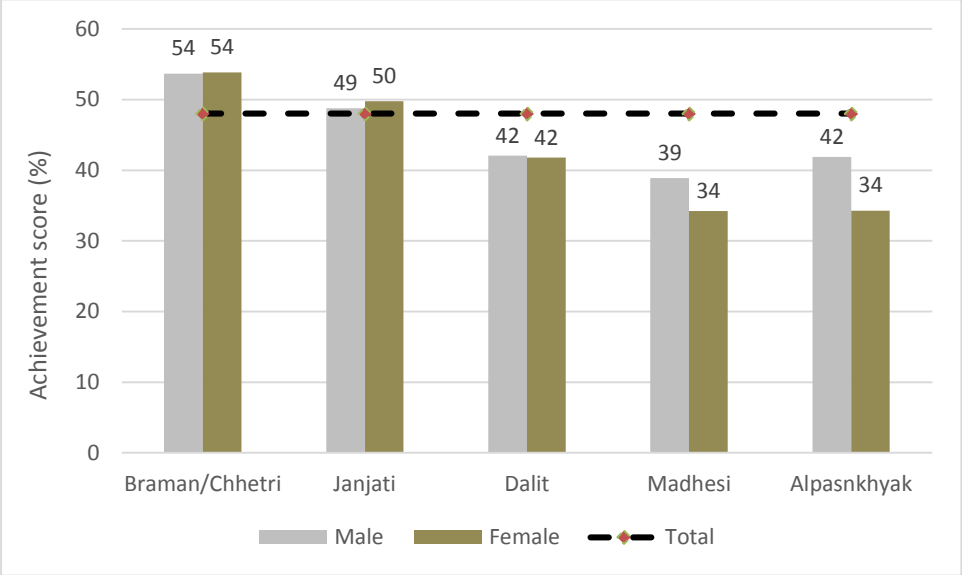


There is no statistically significant difference between boys (48%) and girls (48%) ($p > 0.05$) in the total score though difference is found in the area of reading. Girls are better performers (47%) in comparison to boys (46%) which is not statistically significant ($p = 0.001$). In grammar and vocabulary, boys performed better than girls. However, the difference is not significant in vocabulary but significant in grammar ($p < 0.001$), and the effect sizes is negligible ($d = 0.004$).

Gender, caste/ethnicity and achievement

The difference in achievement in Nepali subject between boys and girls is the highest among Madhesis (difference is 5 percent points) where boys have outperformed girls. Tukey's post hoc test shows that the differences are statistically significant at $p < 0.01$ level. The Madhesi boys outperformed girls only in the Central and Western regions, but in the Far-Western region and the Kathmandu Valley girls are remarkably better than boys by 19 and 14 percent respectively (see figure 4.3.9).

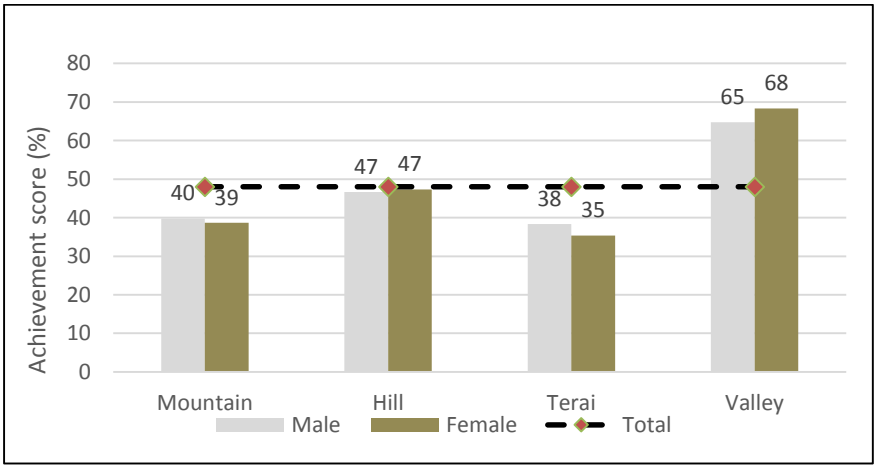
Figure 4.3.9 Comparison of achievement of boys and girls in various castes/ethnicities in the Nepali language



Gender, ecological zone and achievement

The achievement of girls and boys differs significantly across the ecological belts. Namely, in Mountain and Tarai regions boys have outperformed girls. However, in the Kathmandu Valley girls performed better than boys. In case of Hill, there is no difference between boys' and girls' performance. Similarly, there is no statistically significant difference between boys and girls in Mountain whereas it is statistically significant in Tarai and the Kathmandu Valley at ($p<0.001$) (see figure 4.3.10).

Figure 4.3.10 Comparison of achievement by gender in various ecological zones in the Nepali language

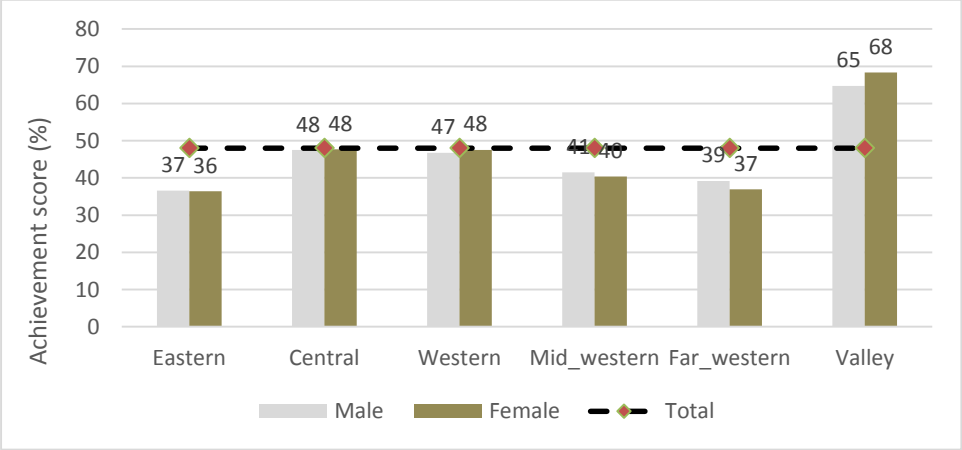


When it comes to the Ecological zones, the differences between boys and girls are very small as the effect size is also small (f = 0.05).

Gender, developmental region and achievement

Notable differences are not found among the development regions in terms of gender. The difference between boys and girls is found somehow wider in the Kathmandu Valley (3 percent) and Far-western (2 percent) (see, figure 4.3.11).

Figure 4.3.11 Comparison of achievement of boys and girls in various developmental regions in the Nepali language



Dataset indicates that boys and girls perform equally in general. However, there are some variations among different communities and regions.

4.4 Selected Explanatory Factors and Achievement

Several factors have already been discussed in section 4.3, for example geographical factors such as districts, ecological zone, and developmental region; school-related factors such as school type and school location; some individual factors related to the students, such as home language, caste/ethnicity and gender. This section deals with socio-economic status (SES) of the students' families, paid work after school, students' attitude towards Nepali language as a school subject, age of the student, and support provided for the studies as the main family and individual related factors. As a sample of deepening school and teacher-related factors, the availability of textbooks, assigning homework and providing feedback by the teacher, and selected activities in the school are considered. Many other factors could be dealt as covered by the background questionnaire.

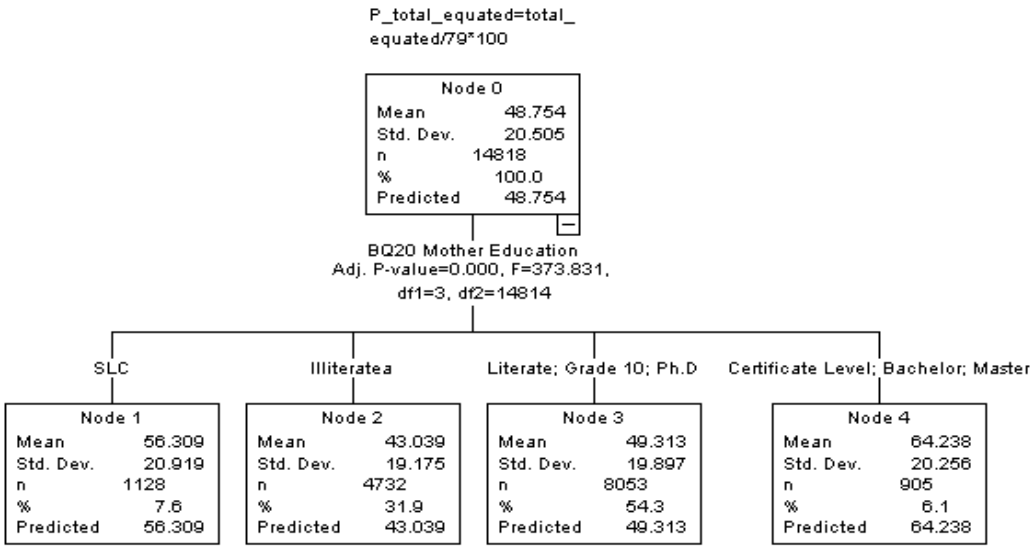
Parents' Education and Occupation and Student Achievement

The variables indicating the socioeconomic status were categorized into parents' education, parents' occupation, home possessions, home accessories (mobile phones, radio, TV available in the students' home), and type of school attended. Finally, SES is explained based on seven indicators related to economic, educational, and occupational background of the parents. In this section, education of the parents is further elaborated and the literacy status of the parents is analysed. Several SES related variables were analysed using a data-mining tool of SPSS-the Decision Tree Analysis (DTA). This method is used to find the cut-offs of the predicting variable, i.e., in case of mother's education it classifies the factors into several groups, as it differs statistically in the most significant way from each other in relation to student achievement.

Parents' education

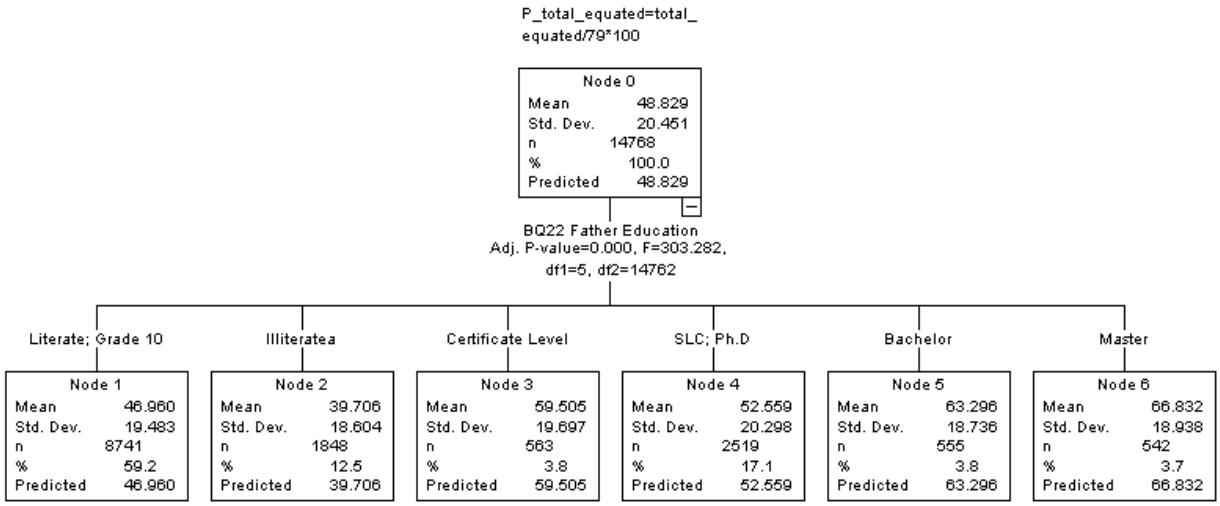
In this assessment, questions related to parents' educational status were included in the background questionnaire and is divided into seven categories as follows: 1) illiterate, 2) literate, 3) grade 10, 4) SLC, 5) Certificate level, 6) Bachelor level, 7) Master level and above. DTA classifies mother's education into four groups with statistically significant differences in students' achievement levels. Figure 4.4.1 illustrates these categories and their values are: illiterate mothers (43%), literate mothers (49%), SLC (56%) and mothers' having certificate, bachelor or master degree (64%). In each group, the number of mothers is high enough to make a credible prediction. The difference between each group is statistically significant ($p < 0.001$).

Figure 4.4.1 DTA of mother's education and students' achievement in the Nepali language



Similarly, as in figure 4.4.2, DTA divides father's education into similar six categories, as in the case of mother. These categories are illiterate (39%), literate (46%), SLC (52%), Certificate (59%), Bachelor (63%) and Master (66%).

Figure 4.4.2 DTA of father's education and students' achievement in the Nepali language



From the figure, it can be concluded that students of illiterate father achieved lowest and the achievement of students increases as the fathers' education increases up to Bachelor degree. The positive relation of parent's education with student achievement is illustrated in figure 4.4. Mother's education explains about 7% (η^2 = 0.070) and father's education 9% (η^2 = 0.093) of the variation, however, the strong relationship was found with the mother's education. DTA shows that when both parents are illiterate, students achieved very low (39%), on the other hand they achieved highest (59%) when mother's education is SLC and father education is certificate or higher level.

Dataset indicates that the educational level of parents affects the children's achievement level. Achievement of the students is more influencing if both parents or mother is at least SLC passed, or if father has the qualification of Bachelor's level or higher.

Figure 4.4.3 Relation between parents' education and student achievement in the Nepali language

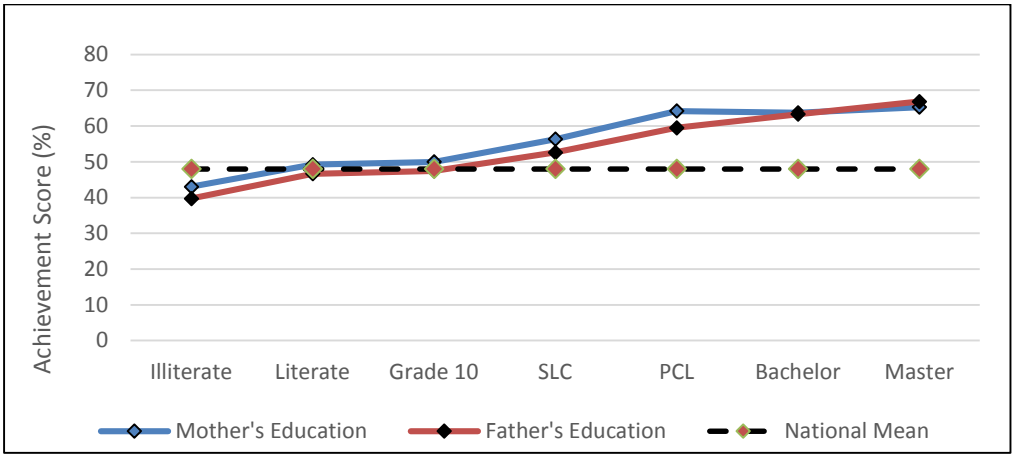
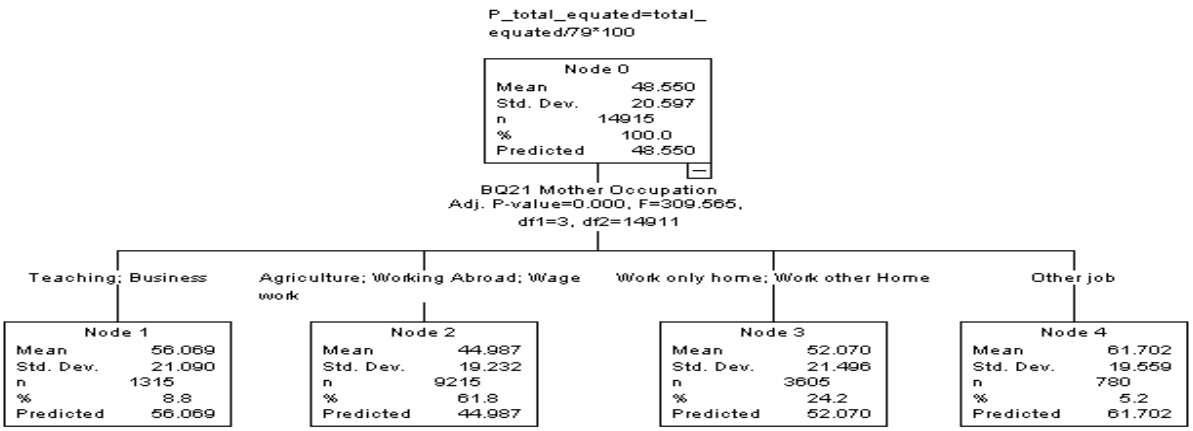


Figure 4.4.3 shows that father's and mother's level of education has impacts on their children's achievement. Students' performance is found below the average national mean if their parents are either illiterate or just literate. Thus, the relation is: the higher the parents' education, the better the students' achievement.

Parents' occupation

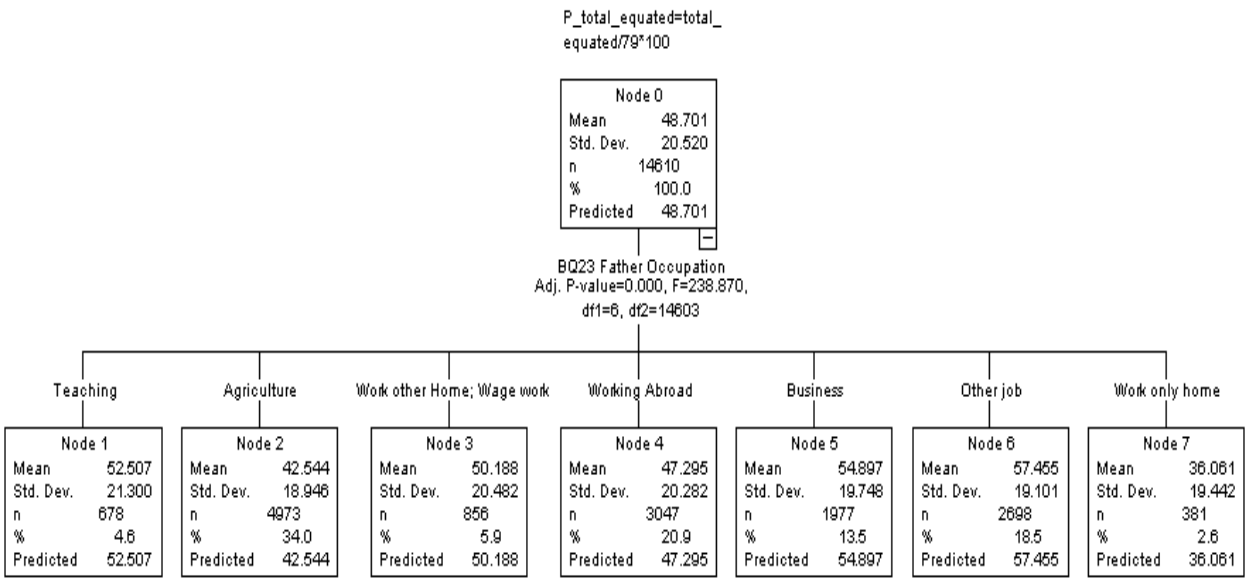
The occupation of parents was categorized into eight groups as follows: 1) agriculture, 2) teaching, 3) service (government, non-government and private sector job), 4) business, 5) working abroad, 6) work at others' home, 7) daily wages work, and 8) work only at home. While comparing the mean using ANOVA, student achievement is the lowest when the mother's occupational background is agriculture (44%). It is statistically significant and lower than the mother involved in teaching (55%), other job (61%), business (56%) and work only at home (52%). The result is presented in figure 4.4.4 as classified by DTA.

Figure 4.4.4 DTA of mother's occupation and students' achievement in the Nepali language



In a similar manner as with the parents' education, DTA was used to find the statistically most deviating groups related to student achievement. This analysis found four significant nodes. The lowest achievement is in the group where the mother comes from agriculture, working abroad and daily wages (45%) (see figure 4.4.4). Significantly higher achievement is seen when the mother is having other jobs (61%) and teaching or business (56%). The highest achievement is in the group "other jobs" (61%). The "Other" occupations are generally defined as government and non-government services, and private jobs. However, the result indicates that the mothers who are employed in any kind of job or business have been found beneficial for the achievement of the students. The difference is found statistically significant at ($p < 0.001$) for each group.

Figure 4.4.5 DTA of father's occupation and students' achievement in the Nepali language



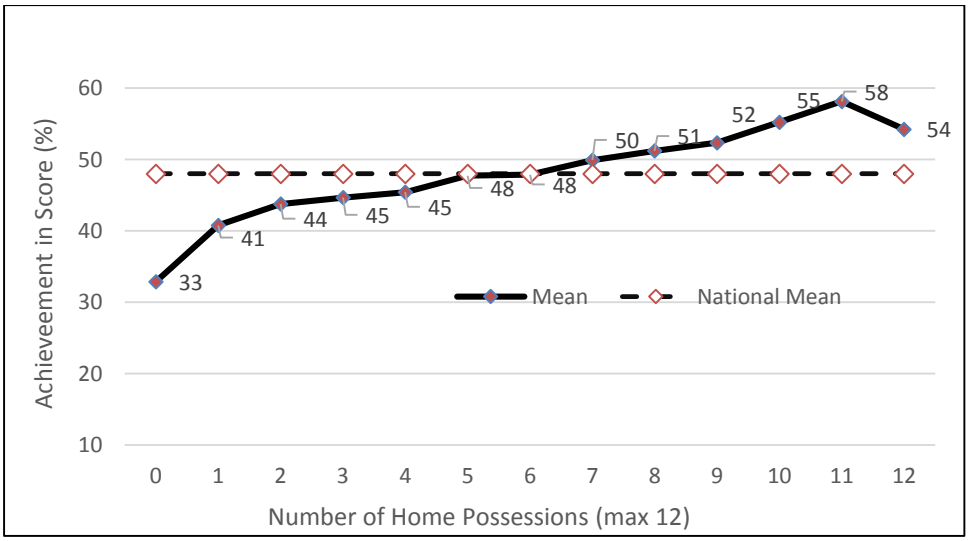
In case of father's occupation, the main division is whether the father works in agriculture (42%) or not. If the father is in a teaching profession, the students perform better (52%) than the children of fathers working in agriculture. Similarly, if the parents are from other job, like the mother, students perform better than other groups (see figure 4.4.5). These seven groups are created by DTA, whereby the difference between each group is found statistically significant at $p < 0.001$.

Home possessions and accessories

This factor is taken into account assuming whether the facilities and resources available at home have some effects on the achievement. There were two kinds of home possessions defined in the background information questionnaire for the students. One is related to the facilities that are supportive for the students' study at home. For example, whether they have a table for study, a separate room for them, a peaceful place for study, a computer for schoolwork, software for the computer assisted learning, internet facilities, literary magazines, access to classical literature, poetry books, or artistic things like pictures, dictionary and other books. Other types of home possession include different types of normal home accessories such as the number of mobile phones, televisions, computers and radios.

Related to home possessions, there were 12 questions in the background questionnaire for students. Each question was scored 1 if the student had an access to the possession. Adding these items up, the maximum score was 12 indicating that the student had access to all of the possessions and hence lower the score means fewer the possessions at home. Figure 4.4.6 shows the positive correlation between the home possessions and achievement level, except in the case of the highest category (i.e., all the home possessions). Pearson Product Moment correlation coefficient between the achievement level and the factor (r = 0.19) is statistically significant ($p < 0.001$) though the value is small.

Figure 4.4.6 Relation between the home possessions and achievement in the Nepali language



For calculating the SES value regarding home possession, the cut-off for the factors was set on four possessions. If the students possessed 5 items or more as mentioned in background questionnaire, the

student was given 1, otherwise 0. The same pattern – the more accessories, the better results can also be found with home accessories, as in figure 4.4.6. The question in the background questionnaire was set differently compared with home possessions. Regarding the accessories the question asked was: "How *many* of the following accessories do you have in your family?" The question was accompanied by the options 0 to 3 (or more). Results are similar in other accessories except for the radio i.e., there is no significant correlation between the availability of radio and student achievement. Therefore, the availability of the home accessories is dichotomized in three groups. After dichotomizing the items individually by using meaningful cut-offs found with ANOVA and DTA (and maximizing the differences in achievement level, see table 4.4.1), all the three indicators were summed, though the missing is not included. The maximum score was 3 indicating that the students possessed a set of all the accessories.

Table 4.4.1 dichotomizing the indicators for home accessories

Accessory	cut-off for 1	cut-off for 0
Mobile phone	2, 3	0 and 1
Television	1–3	0
Computer	1–3	0

Figure 4.4.7 Availability of home accessories and achievement in the Nepali language

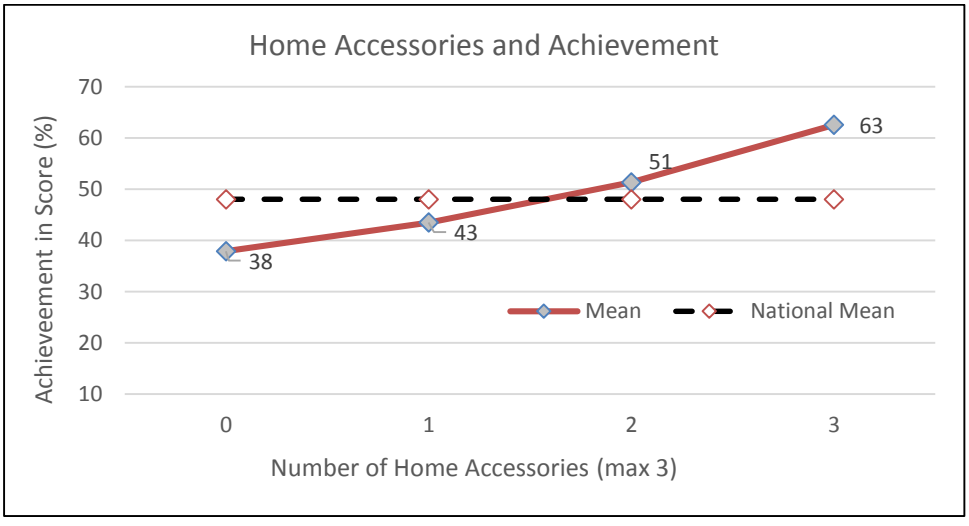


Figure 4.4.7 clarifies that when the number of home accessories increases, students' achievement also increases ranging from 38% (if none of them are available) to 63% (if any three of them are available).

Availability of the stated facilities indicates the high SES of the family. Therefore, the correlation between home accessories and achievement is found strongly positive ($r = 0.39$), ($p < 0.001$) and the effect size is medium.

Data shows that when children have very few home possessions (i.e. 0 to 2 out of the 12), the achievement level is found low (< 44%), which is statistically significant than the achievement of students of family having more than two home possessions (> 48%). When the family is having ten to eleven possessions, the average score is found higher compared with the national average. The same is true for home accessories, that is, when none or only one accessory indicator out of three is met, the results are lower than the average (38–43%) and when there are two or more met, the results are remarkably higher (> 51%). If three or four indicators are met, the results are the highest (63%).

The dataset indicates that either economic or intellectual ability or both at home support children to increase their educational achievement. If the students come from high SES family (having a large number of home possessions and accessories), their performance is found remarkably higher than the performance of other students..

SES and achievement

The value of socio-economic status was calculated based on seven indicators, which, were dichotomized first of all. The variables mother's education, father's education, mother's occupation, father's occupation, home possessions, home accessories, and the type of school were summed as SES and changed into the percentage of the maximum score (PSES). Deeper description of the transformations for SES has been given in chapter 2. The PSES represents the percentage of SES of the student's family; 100 means that the student has the highest possible SES where all the seven indicators of SES are positive, and 0 refers to the lowest possible SES where all the seven indicators of SES are negative. The analysis of PSES by using Univariate GLM (the Regression modelling) shows the strong relation between SES and student achievement. Figure 4.4.8 presents the relationship between SES of the students and their achievement.

Figure 4.4.8 Relation between the SES and achievement in the Nepali language

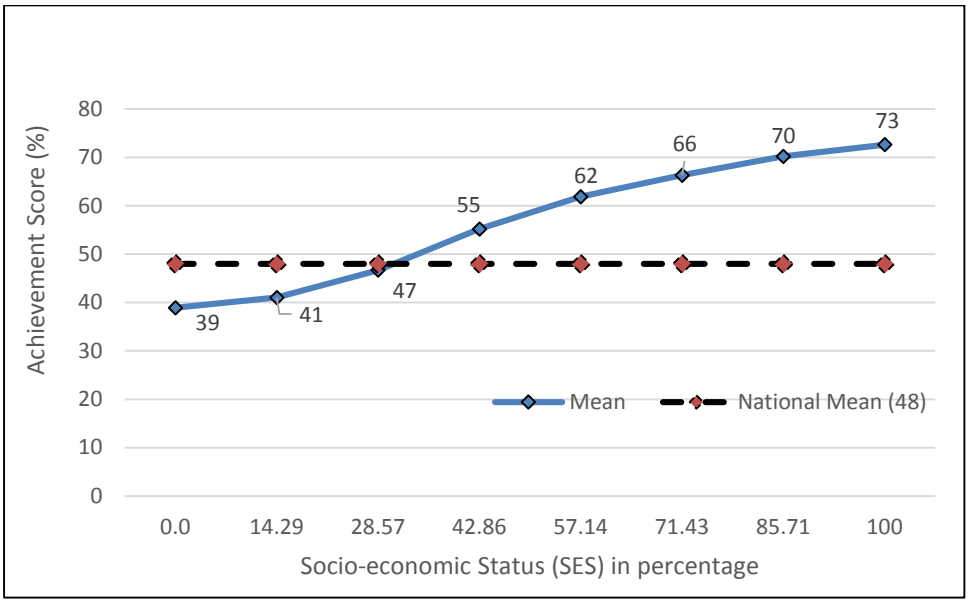
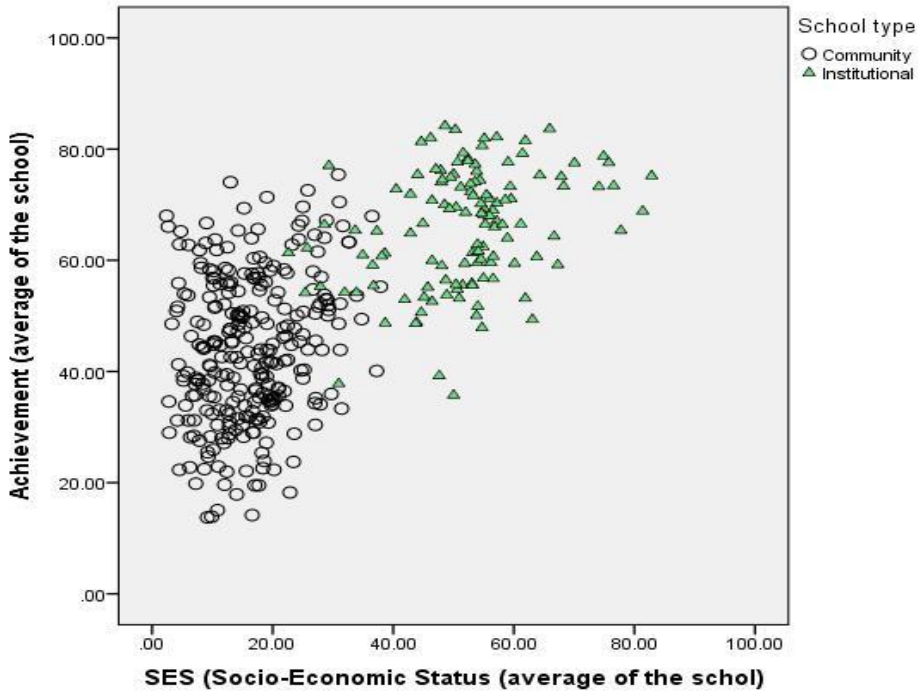


Figure 4.4.8 shows a positive relationship between SES and achievement as the correlation between the variables is $r = 0.461$ which is statistically significant ($p < 0.001$). The differences between the SES groups are statistically significant ($p < 0.001$) and the effect size is high ($f = 0.53$). SES explains about 21% of the variation in achievement ($\eta^2 = 0.219$).

The dataset indicates that the socioeconomic status plays a vital role in student achievement. The difference between the lowest and highest SES groups is remarkable (34 percent). Especially challenging is the situation in the families where the father or both parents are illiterate or both of them work in agriculture.

While plotting the school in terms of mean score and SES in a scatter diagram as in figure 4.4.9, community and institutional schools (community school in circle and institutional schools in triangle) fall into two distinct groups. Most of the institutional schools which have high SES value are among the relatively high performers, whereas the community schools even having low SES value belong to the range from low performing to high performing.

Figure 4.4.9 Achievement of schools with various levels of SES in the Nepali language



The dataset indicates that most of the students in the institutional schools have high SES, and they have achieved relatively higher score in comparison to the students from community schools. Further, the students from the community schools form two kinds of groups: high-performing and low-performing schools. Although the variations in achievement among institutional as well as community schools are wide, community schools are concentrated more on the range of 20% to 60% score. Institutional schools are concentrated more on the range of 40 to 80% of score.

Working Beyond School Hour and Achievement

Seven questions were incorporated in the background questionnaire of the students' activities beyond the school time. The values of the variables are divided into five categories: 0 (no at all), 1 (less than 1 hour per day), 2 (1 to 2 hours per day), 3 (2 to 4 hours per day), and 4 (more than 4 hours per day). The GLM Univariate (regression modelling) indicates that cut-off for example, whether or not the students work in a paid capacity. The relationship is negative when students are engaged in various activities like playing and chatting with friends, playing games, paid job, and entertainment. On the contrary, watching TV 1 to 2 hours a day has positive effect on students' achievement, whereas more than two hours spending on watching TV or not watching TV negatively affected the students' achievement (see figure 4.4.10).

Figure: 4.4.10 Relationship between watching TV and achievement in the Nepali language

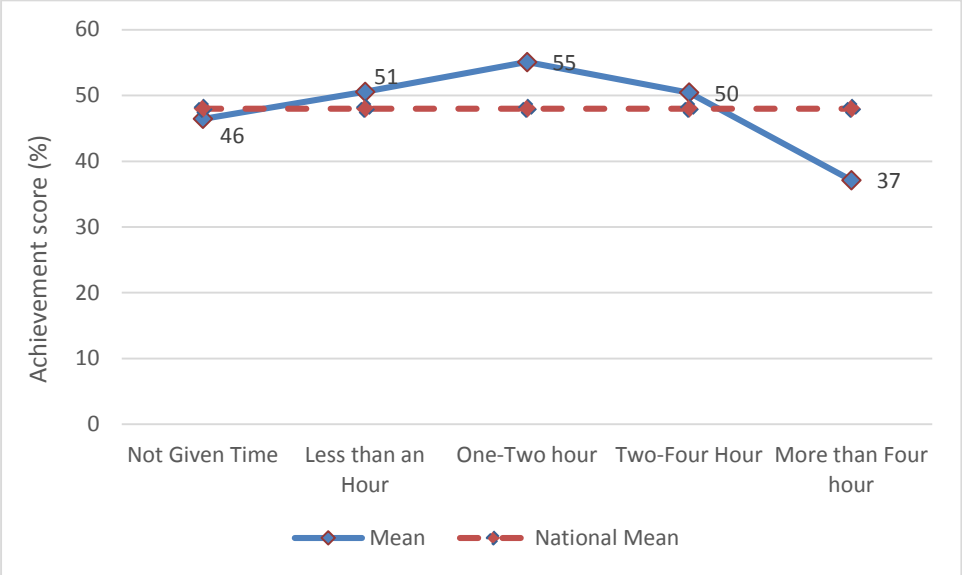
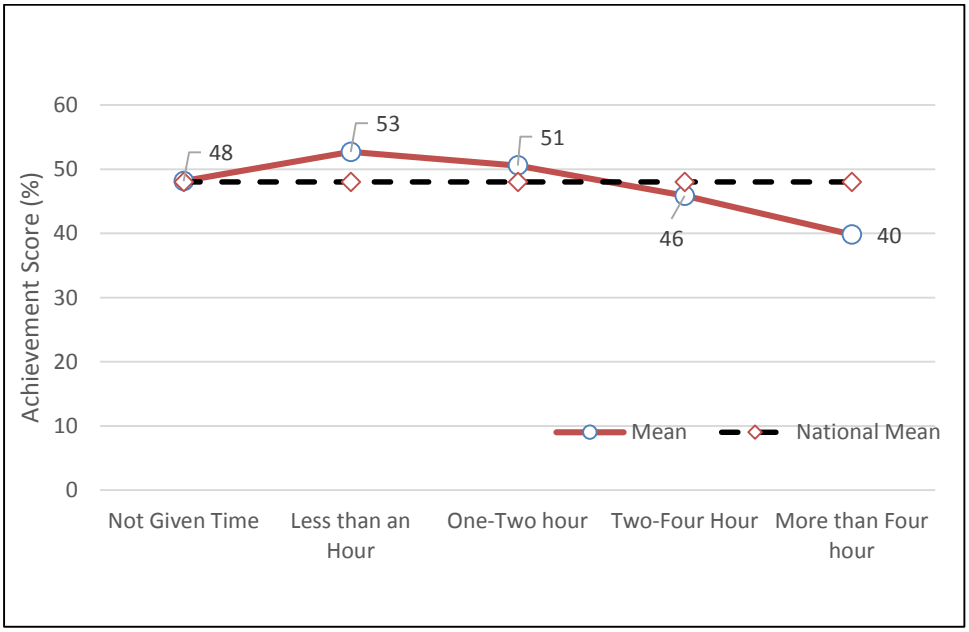
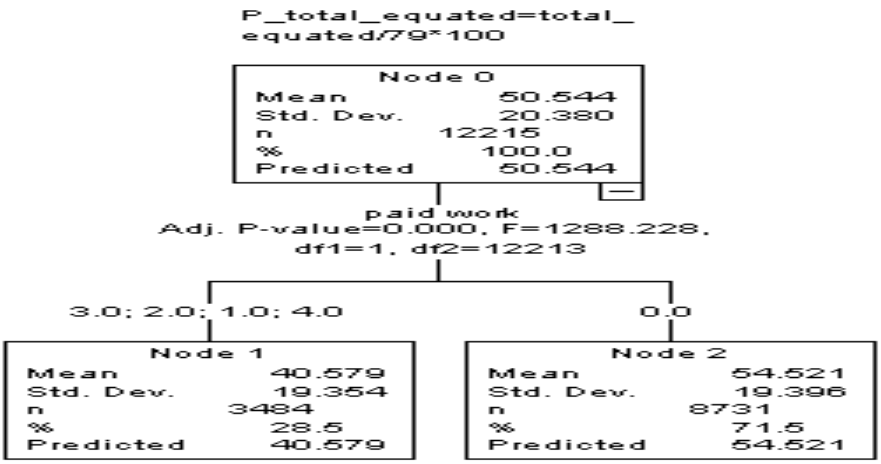


Figure 4.4.11 Relationship between household chores and achievement in the Nepali language



The above figure shows that when children do not work at all at home, the results are lower than the national average (48%). If the students are involved in household chores less than an hour or one to two hours, the results are higher than national average. If students are involved more than two hours in household chores, the results are remarkably lower than the national average. The differences are statistically significant ($p < 0.001$) though the effect size is medium ($f = 0.18$) as out of the total students 21 percent students were engaged in household chores for more than two hours.

Figure 4.4.12 DTA of paid work and achievement in the Nepali language



The DTA shows that when the children have no paid work at all, the results are notably above the national average (55%) (see figure 4.4.12). If the students are working on a paid job, even less than one hour, the results are remarkably lower than the average (41%). The differences are statistically significant ($p < 0.001$) though the effect size is medium (0.33), as large number of the children were not engaged in paid work. Working before/after school time indicates that the family is poor and the extra earnings is needed.

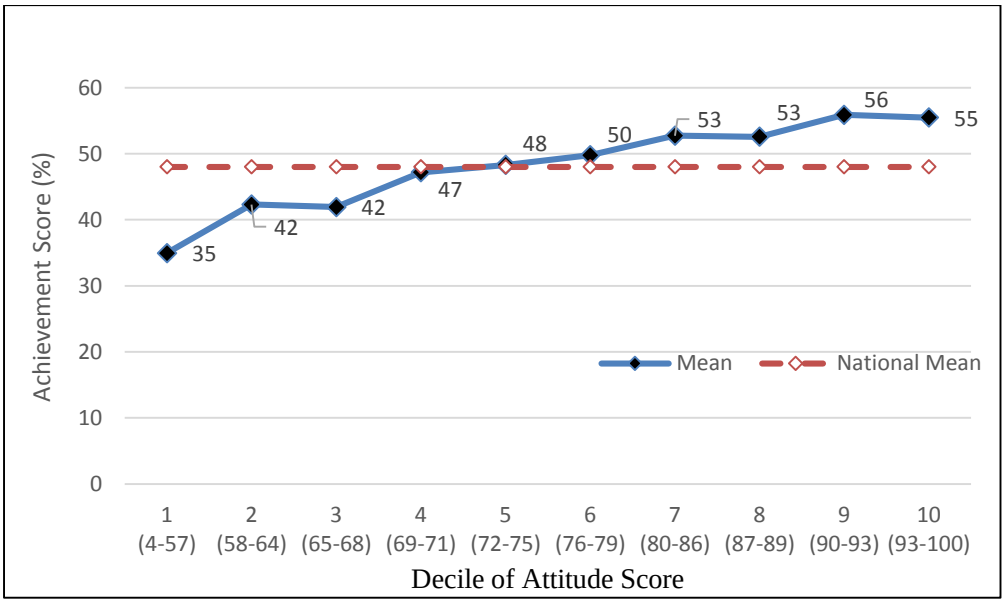
The dataset indicates that working in a paid job or working in unpaid household work for four hours per day noticeably reduces the students' achievement. However, a decent level of household chores that is up to two hours per day supports the learning for the students.

Student Attitude Towards the Subject of Study and Achievement

In the context of Nepali language assessment, the attitude has been explained as what the students think about Nepali language subject and its usefulness in their daily life and future. There is a more or less relationship between the attitude of the students and their achievement. The correlation between achievement and attitude towards the subject is widely studied, though the connection is not always clear, (see, for example Metsämuuronen 2012a; 2012b; House & Telese, 2008; Shen & Tam, 2008; Kadijevich, 2006; 2008). In NASA 2013, the same shortened version of Fennema–Sherman Attitude Scales (FSAS, Fennema & Sherman, 1976), as used in several international comparisons like TIMSS and PISA studies are used. The original scale contains nine dimensions but in these international comparisons only three were used with four items on each dimension and two negative items on each of the first two dimensions (see the detail in chapter 2). The names of the factors can be “Liking Nepali”, “Self-Efficacy in Nepali”, and

"Experiencing Utility in Nepali" (compare naming in, e.g., Kadijevich, 2006; 2008). Factor analysis was used to identify the factors of the responses in FSAS and the negative items were reversed to make the whole test unidirectional. As in several countries in Asia, the expected factor structure cannot be found in Nepal (for a deconstruction of the test scales, see Metsämuuronen, 2012a; 2012b). Hence, only the total score is used to show the connection of attitudes and achievement. The relation between the attitude as divided into ten groups with somehow an equal number of the students, that is, deciles and the achievement score is shown in figure 4.4.13.

Figure 4.4.13 Relation between attitude and achievement in the Nepali language



There is clear positive correlation between the student' attitude and their achievement (r = 0.31). The connection is moderately high (f = 0.33); the division of attitude into ten groups explains the achievement level at somewhat 9% (η^2 = 0.09) variation. The difference between the lowest and highest attitude group is 21 percent. The connection is clear though it is not known whether the positive attitude is a consequence of high achievement or the other way round.

Data indicates that positive attitude towards the Nepali subject is positively correlated with the achievement.

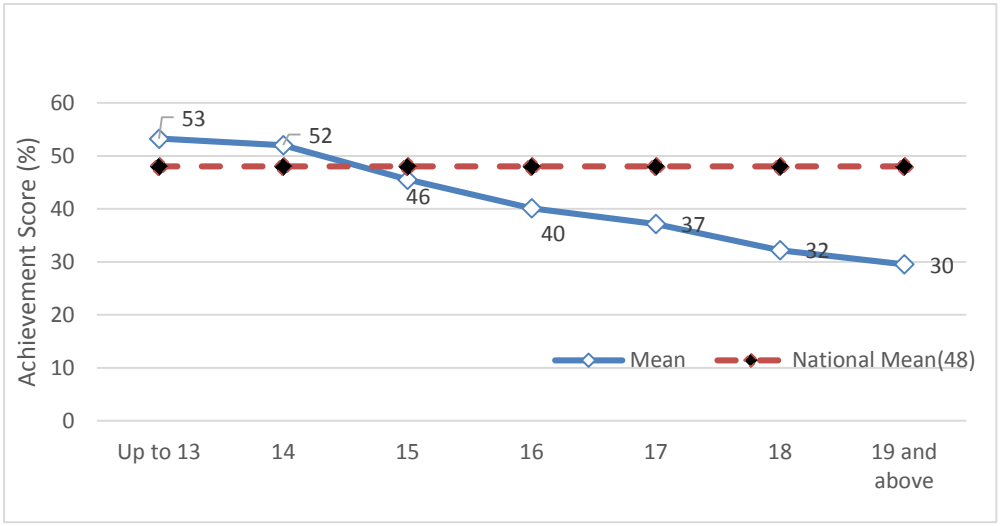
Age of the Students and their Achievement

The age of the students attending grade eight varies widely, although the official age for grade eight students is 13 in the Nepalese context. Some students have mentioned their age below thirteen years and some above 20. All the students aged below 13 were encoded as '13 years', and all those above 19 were encoded as '19 years and above'. The descriptive statistics of the mean in each year are presented in table 4.4.2 and further visualized in figure 4.4.14.

Table 4.4.2 Descriptive statistics of the students' achievement in different age groups

Age	N	Mean	SD
13 years	3532	53	20.5
14 years	5115	52	20.6
15 years	3610	46	20.0
16 years	1587	40	18.3
17 years	565	37	17.6
18 years	238	32	16.9
19 years & above	103	30	17.1
Total	14750	48	20.8

Figure 4.4.14 Relation between the age of the students and achievement in the Nepali language



It is evident that the best achievers are those students who are at the proper age for grade eight studies, i.e., 13 and 14 years old, scoring 53% and 52% respectively. The higher age the students means either the

students have started their study much later, or they have repeated the class. The achievement level is remarkably lower than the average when the students are of age 16 or higher. Correlation between the variables is negative (i.e., - 0.26 at $p < 0.001$) indicating moderate effect size ($f = 0.28$). The ANOVA hints that the age explains 7% ($\eta^2 = 0.07$) of variation in results. The difference in achievement between proper age and over age students is up to 23 percent.

Dataset indicates that the students studying at proper age have obtained the highest achievement score and the achievement decreases as the age increases.

Support to the Students for Study and their Achievement

The relation between the support provided to the students for study and their achievement was analysed based on the information gather on the question: *"Who supports you when you do not understand what you have read?"*. In the question, there was only one option chosen by the respondent in this study, that is, only one supporter was considered. The descriptive statistics of the supporters are given in table 4.4.3.

Table 4.4.3 Descriptive statistics of support to the students and achievement level

Support from	N	Mean	SD
No one	628	53	19.7
Tuition	2115	51	19.2
Brother/Sister	6410	50	20.0
Mother	551	48	23.5
Father	1462	47	21.4
Teacher	3311	44	20.6

Support is necessary for the students to gain better than average achievement. Mother's support at home raises the achievement level (48%), which is less than the achievement of the students having support from brothers and sisters (50%). It is interesting to note that those who were supported by their teacher obtained lower score (44%) against those who did not get any support (53%). Those who took tuition performed higher than those supported by other (51%).

The dataset indicates that the support provided by the mother, through tuition and by brothers and sisters has raised the achievement level more than the support provided by father and teacher.

Availability of Textbook and Student Achievement

There were some students who did not have the Nepali textbook up to the end of the academic session. Table 4.4.4 shows the descriptive statistics of availability of the Nepali textbook and the achievement.

Table 4.4.4 Availability of Nepali textbook and the achievement

Availability of Nepali textbook	N	Mean	SD
Yes	14215	49	20.5
No	414	35	20.2
Total	14629	49	20.6

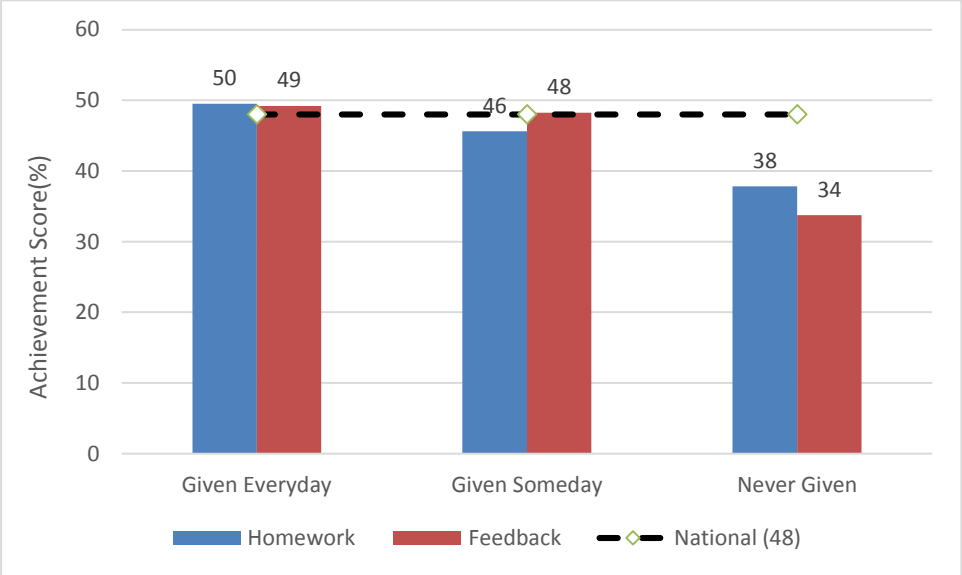
Out of 14629 students who responded to the question, 2.7% did not receive textbook even up to the end of the session. The relation between the availability of textbook and achievement is significant ($p < 0.001$) though the effect size is medium ($d = 0.69$) due to the small size. The difference in achievement is 14%.

Data shows that 2.7% of the students lack the textbook in Nepali. The achievement level of students without textbooks is significantly lower than those who have access to the textbooks.

Homework and Achievement

Homework is considered as one of the ways to enhance learning which can be used as drill, exercise and an evaluation tool as well. When homework is regularly given and feedback is provided to the students, it is likely to boost their achievement level. Statistics related to assigning of homework and providing feedback is presented in figure 4.4.15.

Figure 4.4.15 Relation between assigning homework and achievement in the Nepali language



Based on the dataset, if the teacher assigns homework and checks/provides feedback regularly to the students, achievement is higher (50%) than in the case of those students who are not assigned homework regularly (46%) or not assigned at all (38%). The differences are statistically significant ($p < 0.001$). Those groups without homework are, however, very small, and hence the effect size is small ($f = 0.08$), which explains only 0.7% of the variance in the data ($\eta^2 = 0.007$).

Dataset explains that if the teacher assigns homework and provides feedback regularly, the achievement is higher than without assigning homework.

Positive and Negative Activities at the School and Student Achievement

The activities of the students and teachers determine the learning environment of the school. Bullying, for example, is one of the hindering incidents of the students in the school that may affect learning. In the background questionnaire for students , several school related activities were asked - some positive and some negative. Here, bullying is handled as one of the negative indicators and students' impression from school and teacher activities are taken as the example of positive indicators.

Bullying at school

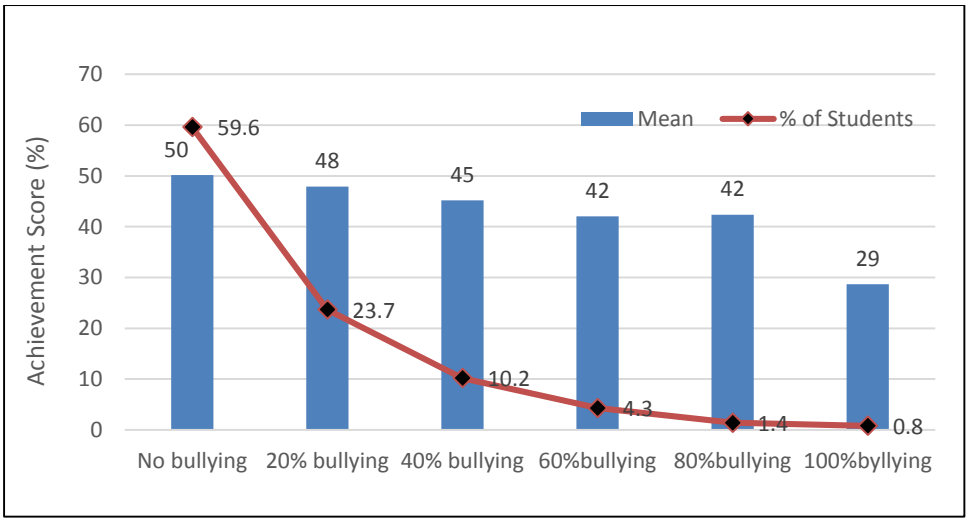
Bullying is one of the problems in the school that worsens the learning environment. International studies like TIMSS and PISA emphasize to identify such indicators. In NASA 2013 student background

questionnaire, five questions indicate the varieties of bullying that tend to happen at schools. All the incidences were stemmed by the phrase "*which of the following activities happened in your school in the last month?*" Students' responses are presented in table 3.4.5 and further visualized in figure 4.4.16. 'No (%)' indicates the percentage of the students' response of no such activity happened in the school and 'Yes (%)' indicates the percentage of the students who experienced the particular type of bullying within the last month.

Table 4.4.5 Bullying and achievement

Type of Bullying	No (%)	Yes (%)
Something of mine was stolen	79.8	20.2
I was made fun of or called names	85.8	14.2
I was hit or hurt by other student(s)	85.7	14.3
I was made to do things I didn't want to do by other students	89.9	11.1
Fellow students kept outside without involving me in activities	90.5	9.5

Figure 4.4.16 Effect of bullying in student achievement in the Nepali language



The sum of all five items is taken as an indicator of 100% bullying. Figure 4.4.16 shows the extent of bullying with the percentage of the students and achievement of the students in each category of bullying. If only one activity of bullying is reported, it is categorized as 20% bullying. If all five activities are reported, it is categorized as 100% bullying. When knowing that 57.8% of the students did not encounter

any bullying during the last month, one can infer that the remaining 42.2% did encounter at least one type of bullying, which is a remarkable number of the students. About 2.2% of the students had experienced some kind of severe bullying (the sum of 80% and 100% bullying). It is found that learning outcomes are remarkably lower with 6.5% of the students who have encountered more than two different types of bullying (42%). Students who did not experience bullying and students who encountered extreme bullying of four or five kinds have 21 percent achievement gap; though there are a few number of the students who reported this kind of bullying (n = 963). The difference is statistically significant (p = 0.001), but the effect size is small (f = 0.14). Though extreme cases of severe bullying are rare (n = 121), bullying seems to be quite common incidents in schools.

The dataset indicates that a large number of the students (42%) have encountered bullying in schools within the last month from the date of this assessment. The phenomenon seems to affect the learning outcomes in almost all the groups of the students who experienced bullying.

Positive activities in school

The activities that can boost the learning and achievement of the students are termed as positive activities. The students were asked about such positive activities in the school in two sets of questions listed in table 4.4.6. The table shows the responses of the students in all four categories, which are in the 4-point rating scale, anchored to fully agree and fully disagree.

Table 4.4.6 Students' responses towards teacher and school related activities in the school

Teachers' and students' activities	Respondents in % (valid percentage)			
	Fully agree	Partially agree	Partially disagree	Fully disagree
q27a School: Students get along well with most teachers	77.9	17.4	3.2	1.5
q27b School: Most teachers are interested in student's well-being	85.9	10.1	2.4	1.7
q27c School: Most of the teachers really listen to what I have to say	60.8	30.1	6.5	2.6
q27d School: If I need extra help. I will receive it from my teacher	79.9	16	2.6	1.5
q27e School: Most of my teachers treat me fairly	56.7	21.7	8.5	13.1
q28a School: I like to come and stay in school	89.6	6.8	1.5	2
q28b School: Students in my school try to do their best	77.7	18	2.8	1.5
q28c School: Teacher in the school care about the students	81.7	14	2.7	1.5
q28d School: Teacher wants the students to do their best	90.5	6.3	1.5	1.8
Average	77.9	17.4	3.2	1.5

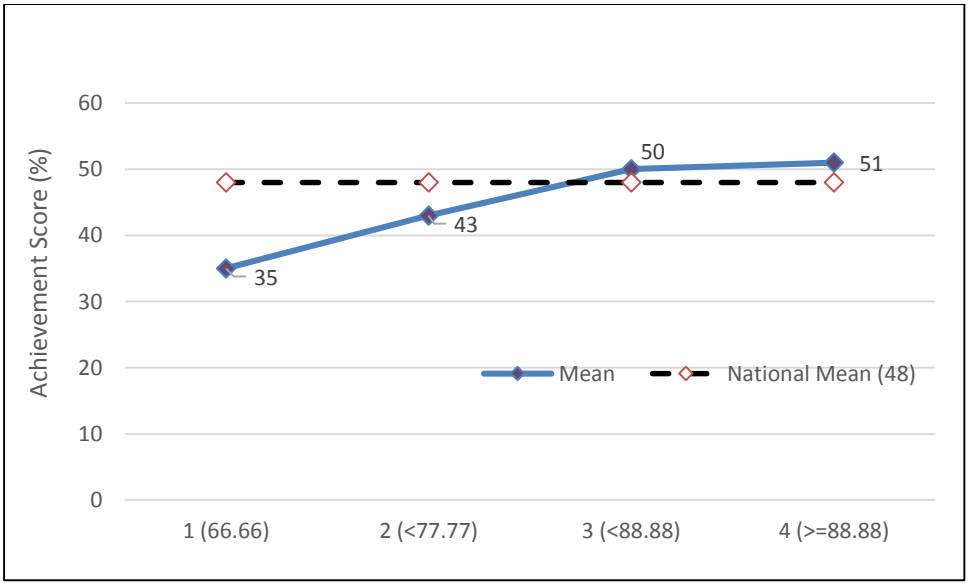
Further analysis was carried out by recoding the variables into two categories: 1 for agree, and 0 for disagree. Furthermore, the sum of nine indicators is converted into the percentage of maximum score to analyse the level of positive activities and its relation to the achievement.

DTA finds four attitude groups in the indicator. These boundaries and descriptive statistics are presented in table 4.4.7 and illustrated in figure 4.4.17. The overall result is that the feeling the positive actions in the school relates positively with the student achievement. The correlation between the sum of nine positive activities and achievement is found positive ($r = 0.25$, $p < 0.001$).

Table 4.4.7 Students' response towards teacher and school related activities in the schools

Percentage of positive action	N	Mean	SD
1 (> 66.66)	1690	35	21.3
2 (< 77.77)	1256	43	21.3
3 (< 88.88)	3484	50	20.0
4 (≥ 88.88)	8783	51	19.8

Figure 4.4.17 Relation between positive actions in school and achievement in the Nepali language



The data shows that there is a positive relation between the positive activities of the school and the achievement. The increase in achievement is directly proportional to increase in the intensity of such

activities. After dividing the indicator into four groups based on DTA, the differences between the groups are statistically significant ($p < 0.001$), however, the effect size is medium ($f = 0.24$). The learning achievement is much higher than the average (51%) only when the students are extremely positive towards school and teachers' behaviour,. However, the difference between the most positive group and the most negative group is notable (16%).

Dataset indicates that when the students think that the actions of the teachers and the schools are ultimately good, the results are higher than the average (51%). At the other extreme of feeling, when such actions are ultimately negative the results are far below the average (35%).

4.5 Summary of Findings

The main findings on **Nepali** language subject at grade eight NASA 2013 are as follows:

Basic results

- The achievement in Nepali language subject is normally distributed as there are two distinct population groups.
- The average achievement in Nepali subject is 48 percent. The achievement is the highest in the content area of vocabulary (55%) and the lowest in reading/writing (46%).
- Students' ability to solve complex problems is quite low. Students are much better in the recalling type of questions (57%), which were geared more on remembering the things than solving problems. Students' performance was found low in application than higher ability.
- The students are performing well in recognizing the correct answer and recalling simple facts from the texts, fundamental thinking, the basic interpretation of paragraph, table, chart and a few steps of logical thinking, basically objective items (58%). They are much weaker in producing fluent texts or letters, or preparing synthesis and abstracts from the text, basically subjective items (45%).

Diversity factors

- There is a wide difference in achievement among the districts in Nepali language subject. The results in Mugu (29%), Siraha (30%), Sunsari (34%), Dang (35%), Bajura (36%), Pantchthar (36%), Sarlahi (37%), and Dhanusa (39%) are relatively very low. Out of 28 districts, 17 districts are below the national average. The gap between low performer district (Mugu 29%) and high performer district (Kathmandu 68%) is 40 percent.

- There is a moderate difference in the students' achievement across the four ecological zones. Students in the Kathmandu Valley outperform (66%) the students from other regions. The achievement is found low in Tarai area (37%).
- There is a wide gap among the development regions in students' achievement in Nepali language subject. The difference between the lowest performing area (Eastern Region, 36%) and the highest performing region (Kathmandu Valley, 66%) is remarkable, that is, 30 percent.
- On an average, the students in the institutional schools outperform (65%) the students in the community schools (42%). However, some students from community schools perform equally as those from institutional schools.
- The students in the urban schools gained 12 percent more than the students in the rural areas. Most of the high performing institutional schools were from the urban areas. Excluding the Kathmandu Valley, the difference between urban and rural schools is 7 percent.
- There is an achievement difference across the language groups. The students from Newar (61%) and Sherpa (53%) communities have performed higher than the students from Nepali language group (52%). On the other hand, the students from Limbu (34%), Tharu (35%), "other" (36%), and Urdhu (39%) communities performed much lower than the national average.
- The Madhesi students' performance is lower than the performance of the other castes/ethnic groups/communities. In particular, their performance is low in the mid-Western and Eastern developmental regions (29%). A notable percentage of Madhesi students obtained less score than the national average in reading and writing.
- There is no difference between girls' and boys' performance in Nepali language. Girls are slightly better performing than boys in reading, whereas, boys are better in vocabulary though the differences are not significant. However, there are notable differences in achievement between boys and girls in Madhesi students. Differences are also found between boys and girls across the ecological zones and developmental regions, but the difference is small in development region in comparison to ecological zones.

Selected explanatory factors

- The parents' educational level strongly predicts the children's future achievement level in Nepali language. Especially, achievement level is the lowest when the parents are illiterate.
- Families either of the high economic or intellectual capacity or both have positive effect on their children's achievement score. If the father or mother or both are from agricultural or related

occupation, the students' achievement in Nepali subject is significantly lower than that of the students whose parents are from other occupational groups.

- When the children have very few home possessions i.e., 0 to 2 out of the 12, the achievement level is significantly lower (< 41%) than the achievement of those having more than two home possessions (> 45%). With a family having ten to eleven possessions, the average score is very high (> 55%) compared with the national average.
- The socio-economic status (SES) plays a strong role in the educational processes in Nepali. The difference between the lowest and highest SES groups is remarkable (34 percent points). Especially challenging is the situation in the families where the father or both parents are illiterate or both parents work in the agricultural field.
- Either any paid work or unpaid household work for four hours per day outside school has remarkably reduced the student achievement. However, a decent amount of household work up to two hours per day does not affect learning negatively for the students in Nepali. Rather it gives positive effect.
- Positive attitude toward the subject correlates positively with achievement.
- Achievement decreases as the age increases. The highest performance is with those students studying within their appropriate age group.
- Support to the learning provided by brother or sister, mother and tuition gives more positive effect on achievement level, compared to the support provided by father and teacher.
- It is a big challenge for students to learn without textbook, as 2.7% of the students did not have textbooks of Nepali language subject even at the end of the academic session. The achievement level of these students is significantly lower (35%) than those who have access to the textbook (49%).
- If the teacher gives homework regularly and provides feedback to the students, the achievement is highest.
- High number of the students (42%) have encountered at least one kind of bullying in schools within the last month of the administration of this test. The phenomenon has affected the learning outcomes in almost all the groups of the students who experienced bullying.
- When the students think that the actions of teachers and schools are ultimately good, the results are better than average (51%). At the other extreme, if feeling is ultimately negative, the results are far below the average (35%).

Chapter 5: Analysis of Student Achievement in Science

Science is one of the compulsory subjects in the school curriculum and accorded five out of 40 credit hours a week. The present assessment is entirely based on the learning outcomes or curricular competencies as set in approved curriculum for grade eight. This chapter presents the overall distribution of achievement. Further, it presents achievements across various content areas of the curriculum, various levels of cognitive domain, and types of items. For establishing comparability, the analysis includes the temporal perspective, comparing with the first cycle of study i.e., NASA 2011.

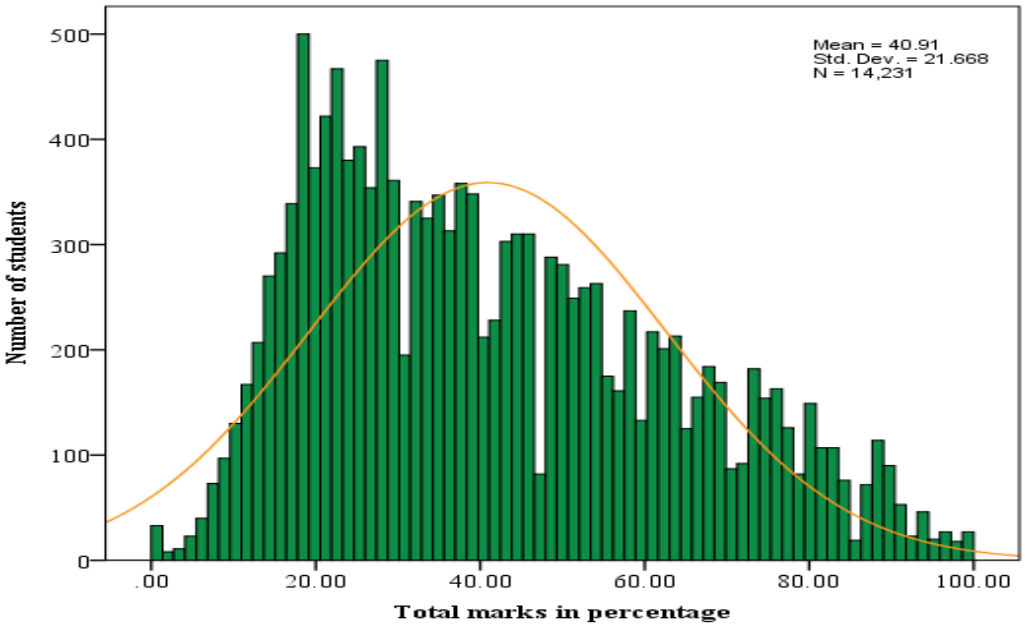
In the second section, there is a discussion to make sense of the level of achievement through comparison with international studies: particularly with TIMSS, an international assessment. The third section of this chapter provides results associated with various diversity factors, including district, ecological zone, development region, school type, school location, language at home, and caste/ethnicity. Section four of this chapter deals with selected explanatory factors about the student achievement in grade eight Science. The factors that explain achievement in Science are parents' education/occupation, home possessions and accessories, SES, work beyond school hour, age of the student, support provided for study, availability of textbook, homework given and checked/feedback provided, negative and positive activities at school. The final section of this chapter summarizes the finding related to the achievement test in Science.

5.1 Basic Achievement Results in Science

Overall Distribution of Achievement Scores

Figure 5.1.1 shows that the achievement scores in Science subject, which is obviously not normally distributed, although the Science sample was big enough to form a normal distribution (altogether 14,231 students). Based on the distribution of achievement scores, the student population can be grouped into three categories: low-performing students, medium-performing students, and high-performing students. Majority of the low-performing students achieve 15 to 25% score, the medium-performing students achieve 35 to 45%, and the high-performing students achieve 60 to 75% score. The figure also shows that the majority of the students fall in the low-performing group.

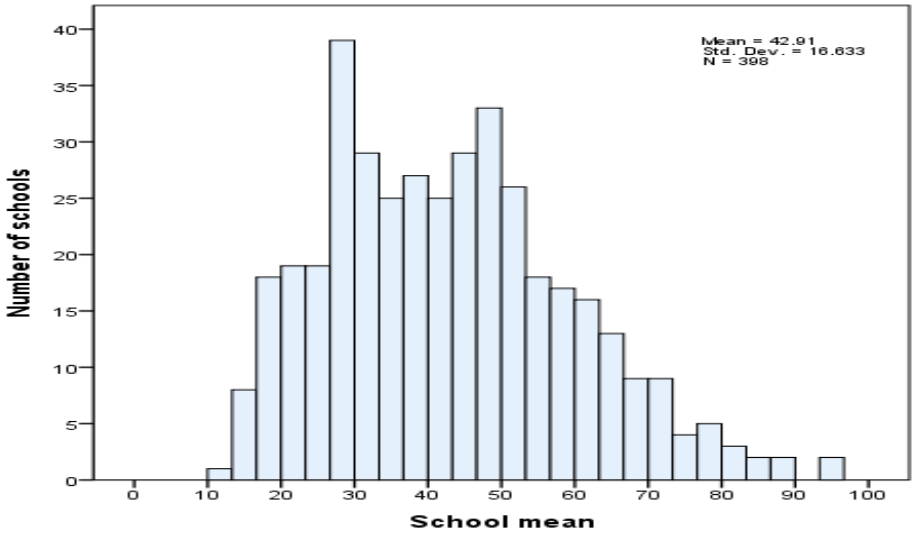
Figure 5.1.1 Distribution of overall achievement scores in Science



The dataset shows that the eighth graders' population is not normally distributed and majority of students in Science are low-performing.

As illustrated in figure 5.1.2 given below, the schools can be divided schools into two categories viz, the high performing schools and the low performing schools. Schools on the left-hand side show the average of 33% of the maximum score and the schools on the right-hand side with the mean of round about 57% or above, which shows the remarkable difference between the populations.

Figure 5.1.2 Distribution of schools' mean achievement score in Science



Student Achievement by Various Content Areas of Science

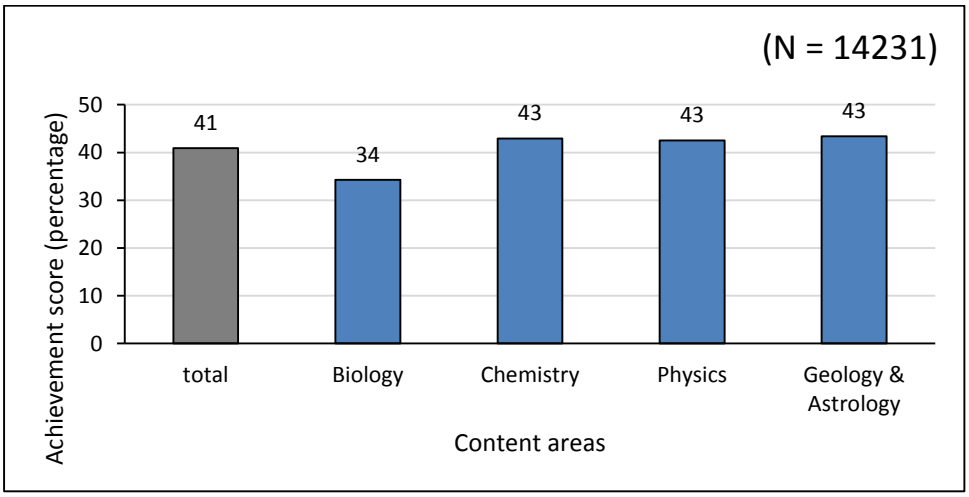
As per the curricular provisions, the Science test includes four content areas: 1) *Biology*, 2) *Chemistry*, 3) *Physics*, and 4) *Geology and Astrology*. The number of items and weighting they carry correspond to the weight allotted in the curriculum. To gain comparability, the achievement scores in all of the areas were converted into percentage. Table 5.1.1 shows the students' achievement in science in total and the achievement level in four content areas in general.

Table 5.1.1 Achievement in various content areas

Content area	N	Mean	Minimum	Maximum	SD
Biology	14231	34	0	100	24.1
Chemistry	14231	43	0	100	23.6
Physics	14231	43	0	100	21.9
Geology & Astrology	14231	43	0	100	24.0
Total	14231	41	0	100	21.7

The table shows the variations of achievement in various content areas of Science. The achievement ranges from 34 percent in biology to 43 percent in chemistry, physics, and geology and astrology, which is 9 percent variation. When we compare the maximum and minimum scores, the situation is the same in all content areas as the maximum score is 100 and the minimum is 0 in each area. Figure 3.1.3 compares the variations of achievements in various content areas of Science.

Figure 5.1.3 Comparison of achievement in various content areas of Science



The overall national average achievement score for science is 41. When the achievement was analysed for the four content areas, students were found performing poorly in Biology (34%), which is far below the national mean (41%). The achievement scores for the remaining three content areas were found greater than the overall achievement score for science: average in Chemistry is 43%, in Physics 43% and in Geology and Astrology 43%.

Dataset indicates that the learning outcomes are poorer in biology than in the other content areas - chemistry, physics, and geology and astrology.

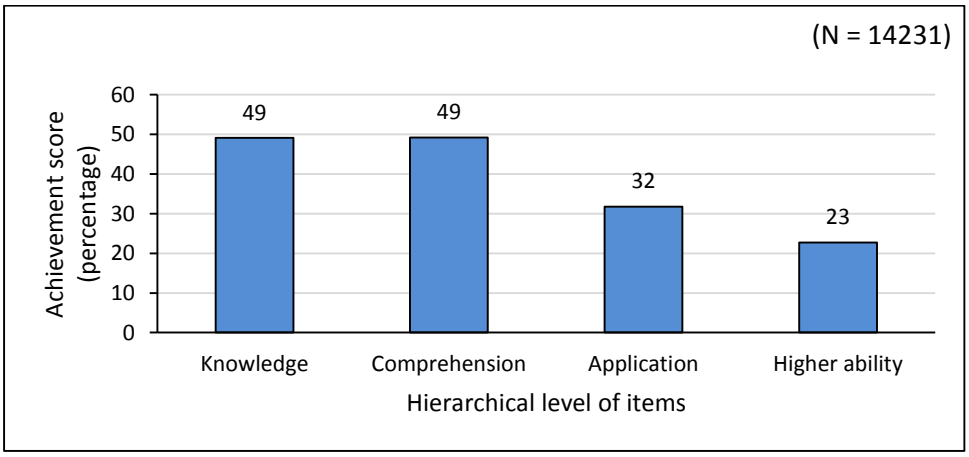
Achievement in Various Levels of Cognitive Domain

The science test comprised of items that can be grouped under the various levels of cognitive domain according to Bloom's taxonomy (Bloom *et al.*, 1956; Metfesser, Michael & Kirsner, 1969), that include knowledge, comprehension, application, and higher ability (reasoning/ problem solving). It is another way of looking at the achievement scores, whether there exists any difference in the achievement by levels of the test items. In fact such an analysis provides valuable insights to the curriculum planners, and instructional designers and implementers. The achievement of students for each of the hierarchical levels is shown in table 5.1.2 and further illustrated in *figure 5.1.5.*

Table 5.1.2 Achievement across various levels of cognitive domain

Level of items	N	Mean	Minimum	Maximum	SD
Knowledge	14231	49	0	100	23.5
Comprehension	14231	49	0	100	20.0
Application	14231	32	0	100	24.8
Higher ability	14231	23	0	100	24.4
Total	14231	41	0	100	21.7

Figure 5.1.4 Comparison of achievement score across various levels of cognitive domain in Science



More than one-third of the students (35%) could not correctly answer any of the items under the 'higher ability' group. In contrast, the students were found much comfortable with the items under knowledge and comprehension level.

The dataset indicates that the students' ability to solve complex problems is low where the mean percentage on higher ability is found only 23%. Students are much better in recall and comprehension type of questions.

Student Achievement by Type of Test Items

There were basically two types of test items, objective and subjective. Objective items covered a wide range of content areas and were very specific because there was only one correct answer or one explicit piece of information required to give the correct answer. There were some subjective items on each test version, which required a longer procedure to get the full marks. Both the objective and subjective types of items represented various levels of cognitive domain including knowledge, comprehension, application and higher ability as well as all the difficulty levels.. Table 5.1.3 comprises the basic statistics of the item type and the achievement levels.

Table 5.1.3 Achievement by type of items

Type of items	N	Mean	SD
Objective	14231	50	20.0
Subjective	14231	32	25.4

It is obvious that the subjectively scored tasks, usually those which are more demanding and need more explanation for the correct answer, are solved much more poorly (32%) than the objective items (50%).

Dataset reveals the fact that students are good in recognizing the correct answers of fundamental knowledge such as choosing the facts and numbers and writing the definitions. They are much weaker in reasoning, problem solving, proving the principle, and constructing the figures. In many cases, the students could not even start doing the open ended questions, hence the low score.

5.2 Comparison of the Result of NASA 2013 with TIMSS

The NASA 2013 was made comparable with the international TIMSS assessment. A limited number of items (9) from the released TIMSS items were used as linking items. Their known difficulty parameters

were fixed in the calibration of the local items. Hence, the international average of θ = 0 was fixed in the Nepalese datasets. If student's ability level in NASA 2013 is zero, it would correspond to the average level of the international students.

Figure 5.2.1 shows the comparison of the students' achievement with the international standard. In the figure, the x-axis shows the content areas of science and y-axis shows the ability shown by the students. The middle horizontal line indicates the international average. As the ability is below the average, the bars are going down; whereas when the ability is above the international average, the bars are going upwards.

Figure 5.2.1 Student achievement in the international TIMSS scale

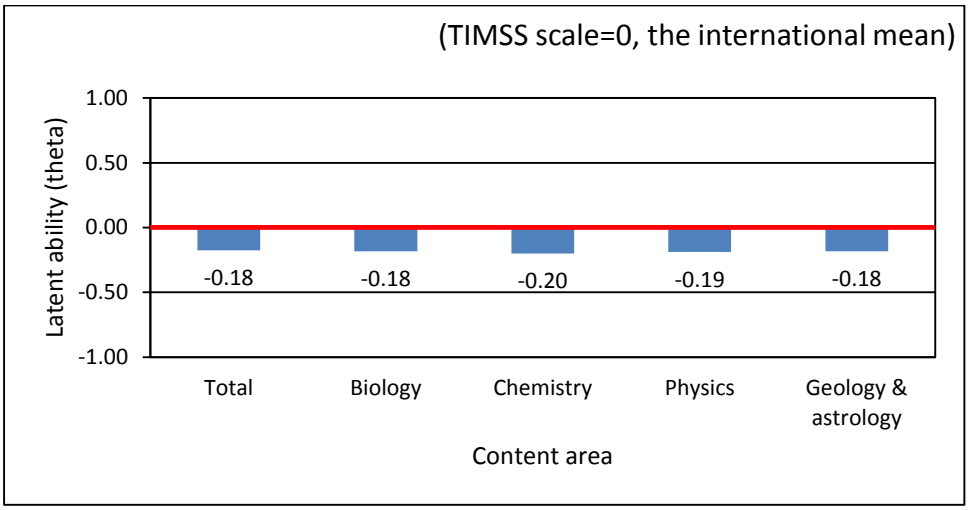


Figure 5.2.1 shows the average ability shown by Nepalese students in science which is below the average. Since the comparison is based on equating the tests with a few selected linking items, the low result only hints that the students in Nepal have not reached the international average, nevertheless how much below is not sure. Actually, in all content areas the results are lower than the international average. On the basis of the raw transformation of the TIMSS scale back to Theta values, Nepalese students' ability level of θ = - 0.18 which is lower than zero (0.00, TIMSS mean).

Dataset indicates that, from the international comparison viewpoint, the average achievement of science for the Nepalese students of grade eight (mean = −0.18) is lower than the international average (mean = 0.00). In all content areas the results are lower than the international mean.

5.3 Achievement Scores by Diversity Factors

Diversity is a relative and contextual term. In the context of NASA, six factors of diversity have been considered, namely geographical/ecological, regional, language, gender, ethnic/caste and economic diversity. NASA 2013 background information questionnaire included the questions related to the six factors of diversity just mentioned. However, this assessment also considered three additional comparisons. They are by district, by school type (community/institutional) and by school location (rural/urban). These comparisons are carried out to assess the equity status of students based on achievement scores.

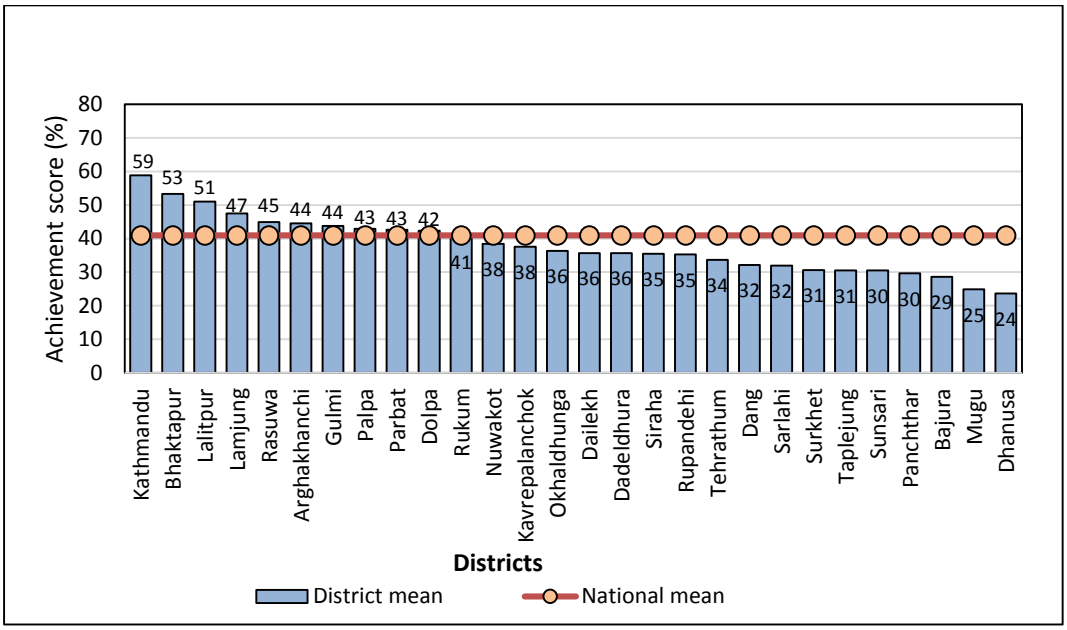
Student Achievement by Districts

The result presented here is based on students' achievement in the 28 sample districts. Student achievement by district is presented in table 5.3.1 and further illustrated in figure 5.3.1.The table and figure present the achievement of districts in descending order according to the students' mean score.

Table 5.3.1 Average achievement score of the sample districts

S.No.	District	Mean	N	SD	S.No.	District	Mean	N	SD
1	Kathmandu	59	2352	20.8	15	Dailekh	36	564	17.3
2	Bhaktapur	53	654	22.5	16	Okhaldhunga	36	267	16.1
3	Lalitpur	51	681	19.1	17	Rupandehi	35	1057	19.1
4	Lamjung	47	318	17.5	18	Siraha	35	499	18.7
5	Rasuwa	45	104	18.8	19	Tehrathum	34	233	19.3
6	Arghakhanchi	44	351	21.2	20	Dang	32	649	18.8
7	Gulmi	44	510	18.8	21	Sarlahi	32	295	18.7
8	Palpa	43	487	18	22	Surkhet	31	719	16.5
9	Parbat	43	263	21.3	23	Taplejung	31	266	18
10	Dolpa	42	108	28.4	24	Panchthar	30	371	17.9
11	Rukum	41	426	21	25	Sunsari	30	863	17.6
12	Kavrepalanchok	38	702	17.6	26	Bajura	29	259	14.4
13	Nuwakot	38	400	17.3	27	Mugu	25	184	13.2
14	Dadeldhura	36	254	15.7	28	Dhanusa	24	395	17.3
						Total	41	14231	21.7

Figure 5.3.1 Comparison of average achievement of sample districts



Of the randomly selected districts in the sample, the student performance was very low in Dhanusa (24%) from Central region; Mugu (25%) from Mid-western region; Bajura (29%) from Far-western region; Panchthar (30%) and Sunsari (30%) from Eastern region. Districts from the Kathmandu Valley are at the top on the achievement level. Of the best performing top most ten schools, nine are from the Kathmandu Valley and one is from Sunsari district, Eastern region. The data-mining module of the SPSS software, Decision Tree Analysis (DTA) points out that there are lowest performing schools in Tarai from Central and schools in Mountain from Mid Far-western region (mean = 27.18).

The difference in achievement scores among the districts is statistically significant ($p < 0.001$). The district variation explained in achievement is $\eta^2 = 0.24$, which explains 24% of the variation in the data. Effect size is $f = 0.55$, indicating remarkably a high difference between the lowest performing district (24%) and highest performing district (59%).

The dataset reveals that there is a wide difference between the districts when it comes to the equal opportunities of children to reach the pre-set goals in science. The results are bound to the 28 districts selected randomly in the sample. In any case, the results in Dhanusa (24%) from Central region; Mugu (25%) from Mid-western region; Bajura (29%) from Far-western region; Panchthar (30%) and Sunsari (30%) from Eastern region are very poor. Districts from the Kathmandu Valley are the top first, second and third districts in achievement score.

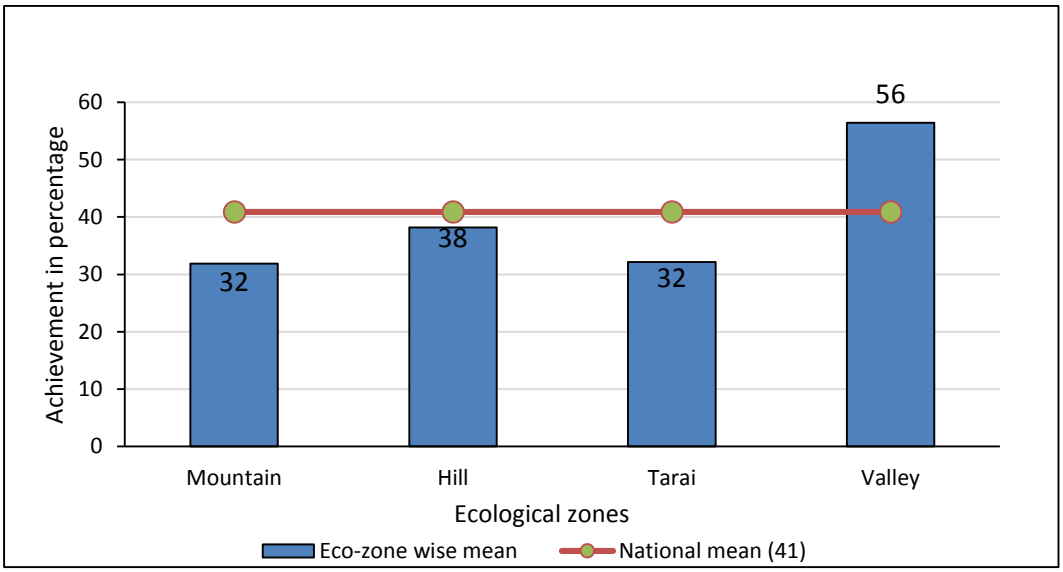
Ecological Zone and Student Achievement

Access to educational facilities can play a vital role in students' achievement. The Mountain, Hill and Tarai are three geographical features in Nepal though the Kathmandu Valley has been considered as a special geographical feature because of being the most densely populated area in the country with more opportunities than other areas. From not only the population point of view, also the mixed ethnicities, weather conditions, economic activities, urbanization as well as the dense human capacity make the Kathmandu Valley a unique geographical area. The variation in the ecological zones in NASA 2013 is presented in table 5.3.2 and further illustrated in figure 5.3.2. Table 5.3.2 also included the number of students, standard deviation, standard error and minimum and maximum scores in each categories

Table 5.3.2 Achievement in various ecological zones

Eco-Zone	N	Mean	SD	SE	Minimum	Maximum
Mountain	921	32	19.1	0.630	1	93
Hill	5865	38	18.9	0.247	0	100
Tarai	3758	32	18.8	0.306	0	95
Valley	3687	56	21.1	0.347	0	100
Total	14231	41	21.7	0.182	0	100

Figure 5.3.2 Comparison of achievement scores by ecological zones in Science



The data shows that, on an average, the students from the Kathmandu Valley (56%) have outperformed the students from all the other ecological zones. The students from the Mountain and Tarai areas performed

the lowest (32%). Students' performance in Rasuwa (45%) was found higher than the average whereas Bajura (25%) was remarkably lower than the average, showing a wide differences between the districts even within the Mountain area.

The achievement in the regions differs significantly ($p < 0.001$). Based on Tukey's *post hoc* test, all the ecological zones deviate from each other in a statistically significant manner at $p = 0.05$ level. The effect size is $f = 0.49$, showing a w ide difference between the highest and the lowest performing ecological zones. In this comparison, the district explains more than 24% of the variation. Variation within the zone is highest among mountain districts (f = 0.37) than other zones, that is, hill (f = 0.28), Tarai (f = 0.19) and the Kathmandu Valley (f = 0.16) which reveals that there are more similarities in the performance of students from the districts in the Kathmandu Valley, Tarai and Hill in comparison to Mountain.

Dataset indicates that there is a remarkable difference between the highest and lowest achieving ecological zones on student performance. Students in the Kathmandu Valley outperformed the other zones whereas it is lowest in the Mountain and Tarai zones. Variation in the districts performance of districts within the mountain zone is widest.

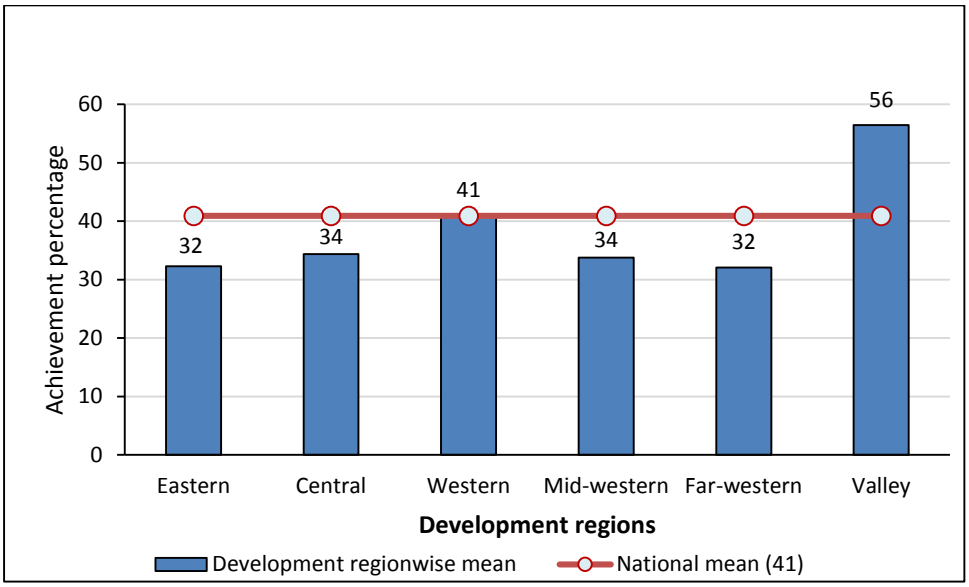
Development Region and Student Achievement

Student achievement varies according to the development regions which are divided into 1) Eastern, 2) Central, 3) Western, 4) Mid-Western, and 5) Far-Western regions. Additionally, the Kathmandu Valley is considered as a separate Development Region though administratively it falls under the Central Developmental Region. The mean achievements in the development regions are given in table 5.3.3 and further illustrated in figure 5.3.3.

Table 5.3.3 Achievement scores in various development regions

Region	N	Mean	SD	SE	Minimum	Maximum
Eastern	2499	32	18.1	0.362	0	97
Central	1896	34	18.7	0.430	0	95
Western	2986	41	19.7	0.360	0	95
Mid-Western	2650	34	19.0	0.368	0	95
Far-Western	513	32	15.4	0.682	4	100
Kathmandu Valley	3687	56	21.1	0.347	0	100
Total	14231	41	21.7	0.182	0	100

Figure 5.3.3 Comparison of student achievement in various development regions in Science



The highest performance can be found in the Kathmandu Valley (56%) followed by the Western region (41%), both are above the average. The performance is low in Eastern and Far-western regions (32%) though the Mid-West (34% and Central (34%) are also the regions where students' performance is lower than the average. In total the difference among the regions is statistically significant ($p < 0.001$). In terms of Tukey's *post hoc* test, there is statistically significant difference between the Kathmandu Valley and all other regions. Similarly, Western Development Region also differs significantly from all other regions, whereas there is no significant difference between Central and Mid-western, and eastern and Far-western regions. Effect size is found high ($f = 0.50$), and development region explains 19% variation in achievement.

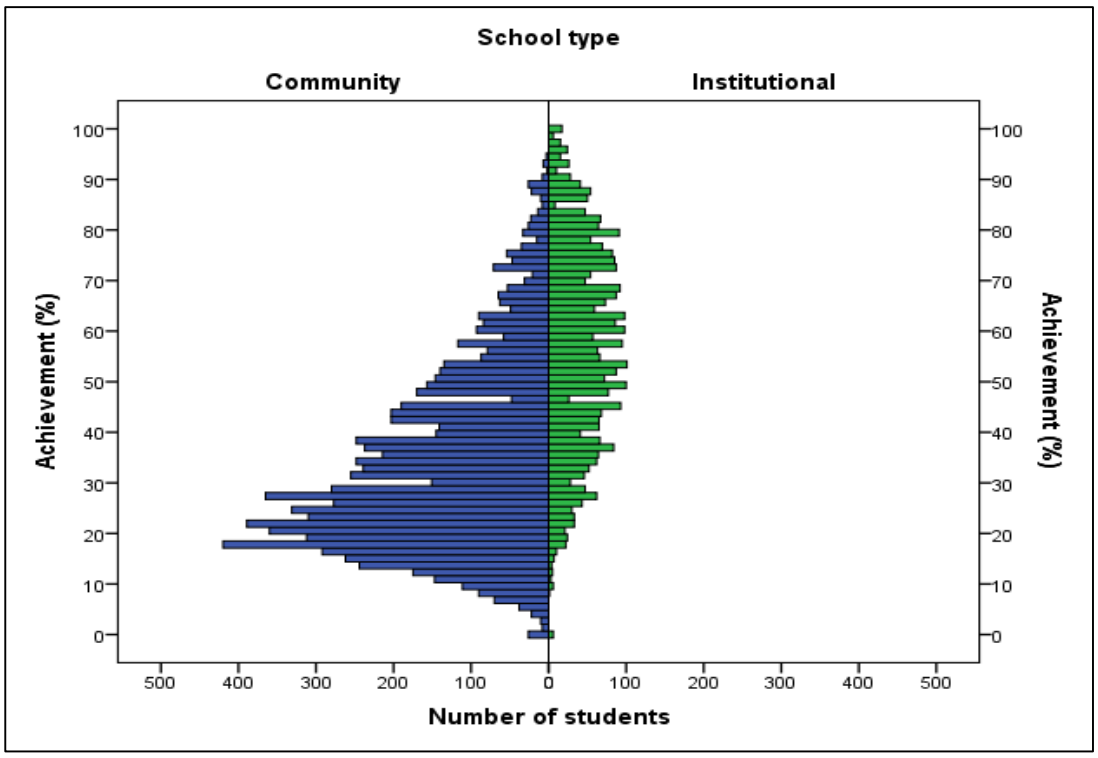
The difference between the lowest performing region (Eastern and Far-western region, 32%) and the highest performing region (Kathmandu Valley, 56%) is remarkable as it is 25 percent.

Student Achievement by School Type

While plotting the sample of community and institutional schools separately we get a chart as given in figure 5.3.4. In the chart, the left hand side distribution shows the community school students, distributed at least in three populations (groups) of low, medium and high performing in the same way as the total student population. There are remarkably high number of the students below 30% of the maximum marks, which is similar to Mathematics score of grade eight NASA of 2011. Similarly, the right hand side distribution shows the population of institutional school students where majority of the students shifted

towards better performing levels though some students are still obtaining very low marks. This indicates that the students from institutional schools are varying from the low performers to the highest performers though most of the students are high performers. On the other hand, in the community schools relatively larger number of students are low performers.

Figure 5.3.4 Distribution of the students' mean scores in community and institutional schools in Science



All the schools are categorized into public and institutional types, the latter are also called private schools. The difference in the achievement of the two types of schools is presented in table 5.3.4.

Table 5.3.4 Student achievement by school type

School type	N	Mean	SD	SE
Community	9103	34	18.0	0.189
Institutional	3469	57	20.0	0.340
Total	12572	41	21.3	0.190

The achievement gap between the community and institutional schools is big. The average achievement score for the private schools is 57% whereas it is 34% in the community schools, with a difference of 23 percent. The difference is statistically significant ($p < 0.001$) and the effect size is high ($d = 1.25$), showing a wide difference between the community and institutional schools. Division of the children to the community and institutional schools explains 24% of the variation in student achievement ($\eta^2 = 0.238$).

Among the 50 high performing schools, only nine are community schools. Further, none of the high performing community schools is from the Kathmandu Valley. On the other hand, most of the private schools in the sample demonstrated high level of achievement. It is also observed that class size in these private schools is smaller.

The dataset indicates that the students in the institutional schools outperformed the students in the community schools.

School Location and Student Achievement

One of the strata of sampling in NASA 2013 was the school location. The schools were categorized into rural and urban schools in terms of their location. The achievement data disaggregated by school location is presented in table 5.3.5.

Table 5.3.5 Student achievement on the basis of location of school

School location	N	Mean	SD	SE
Rural	10357	37	20.2	0.198
Urban	3874	51	22.1	0.356
Total	14231	41	21.7	0.182

The achievement level of the students in the urban schools (51%) is higher than that of rural schools (37%). The difference in average score is statistically significant ($p < 0.001$), though the effect size is medium ($d = 0.35$). School location (rural and urban) explains only 8.8% of the variation in student achievement ($\eta^2 = 0.085$).

Data reveals that students' performance in urban areas is higher than the rural areas.

Language at Home and Student Achievement

In the context of Nepal, student achievement is found depending on the language spoken at home i.e., the mother tongue. In many cases, the mother tongue reflects the ethnic background and hence any difference may be taken as a possible source for inequity in society. In Nepali language assessment the results was obvious. However, it is not known what the effect is in Science subject.

On the basis of the total data, 25% of the eighth grader students speak a language other than Nepali as their first language. These "other" languages are quite fragmented. The largest groups in the student dataset are Tamang (5%),Tharu (3.2%), Newari (2.6%) Magar and Abadhi/Bhojpuri (0.8%). After dividing the languages into ten groups excluding Nepali, there were still 10.4% of the students remained in the group "Other". Because the languages are much fragmented and the Nepali speakers are the majority of the students, for the purpose of this analysis, all the other languages were grouped into "non-Nepali speakers". The results are presented in table 5.3.5 and further illustrated in figure 5.3.6.

Table 5.3.6 Student achievement of various language groups

Language group	N	Mean	SD	SE
Nepali	10208	43	22	0.216
Non-Nepali	3564	35	20	0.329

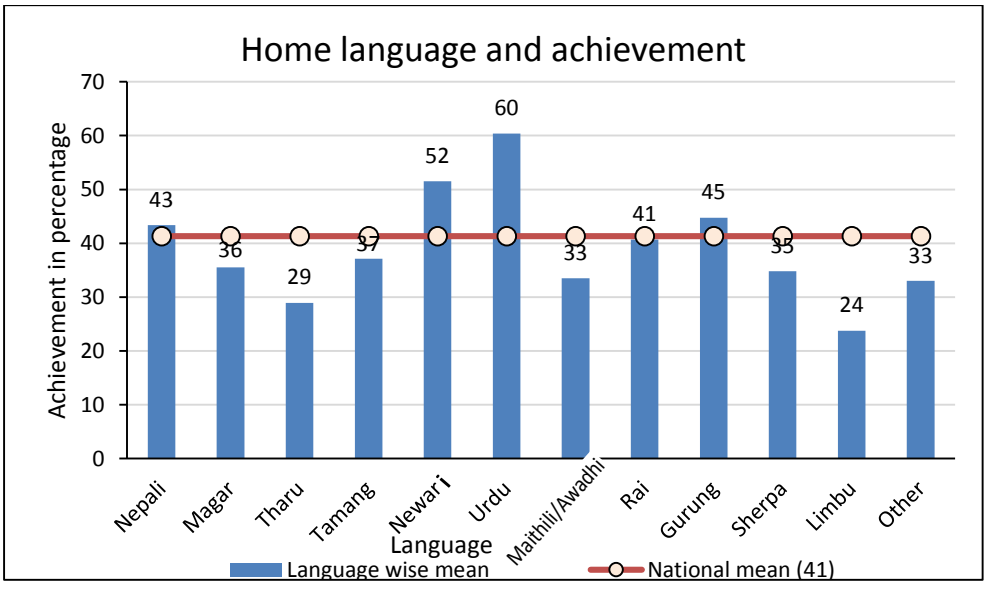
When combining all the minor language groups as "Non-Nepali", there is a significant difference between the Nepali and Non-Nepali language groups ($p < 0.001$). However, when we see each language group separately, it is evident that the small groups such as Urdu (60%), Newari (52%) and Gurung (45%) speakers perform well in science. It is also noteworthy that the equally small language groups, such as Limbu (24%) and Thru (29%) speakers perform very low. The situation demands a deeper analysis to design more appropriate interventions.

Table 5.3.7 Student achievement by linguistic background

Language[1]	N	Mean	Std. Deviation	Std. Error	Minimum	Maximum
Nepali	10208	43	21.9	0.216	0	100
Magar	107	36	17.8	1.723	0	89
Tharu	450	29	14.8	0.697	7	86
Tamang	714	37	18.2	0.680	0	100
Newari	377	52	20.3	1.044	0	100
Urdu	16	60	16.5	4.116	26	81
Abadhi/Bhojpuri	114	33	20.0	1.874	10	88
Rai	39	41	20.6	3.301	7	100
Gurung	59	45	19.3	2.515	16	88
Sherpa	31	35	19.4	3.492	12	79
Limbu	158	24	14.8	1.178	4	75
Other language	1473	33	19.3	0.503	0	97
Total	13781	41	21.6	0.184	0	100

1) Language groups with less than 14 students are omitted.

Figure 5.3.5 Student achievement by language background in Science



The difference between the language groups is statistically significant ($p < 0.001$). However, the effect size is medium ($f = 0.23$) and division into smaller language group explains 5% of the variation in achievement ($\eta^2 = 0.052$). When analysing only the minority languages and hence excluding the Nepali speakers and

the group "Others", the effect size is high (f = 0.37), indicating a remarkable difference between the highest performing group (Urdu, 60%) and the lowest performing group (Limbu, 24%).

Ethnicity/Caste and Student Achievement

Modern education in Nepal has been influenced in several ways by the legacy of the historical caste system which still exists in the mind-set of the Nepalese society. Historically, the Brahmans and Chhetries have been heavily involved in education, but Dalits, for example, have remained outside from educational system. Hence, the government has made several efforts to make education possible and accessible for all children regardless of the ethnicity and caste. The recent National Population Census 2011 shows that the number of Hill Dalits has been increased remarkably at the lower level of schooling but their number at the secondary and higher education is still very small (CBS, 2012). The results concerning the castes and achievement are presented in table 5.3.10.

Table 5.3.8 Student achievement by ethnic/caste background

Caste/ethnicity	N	Mean	Std. Deviation	Std. Error	Minimum	Maximum
Brahman/Chhetri	5076	41	21.2	0.298	0	100
Janjati	5438	45	22.1	0.300	0	100
Dalit	1278	35	18.6	0.519	0	100
Madhesi	1169	36	21.5	0.629	0	99
Minorities	37	34	20.2	3.317	11	77
Other	664	39	20.4	0.792	1	93
Total	13662	41	21.6	0.185	0	100

The NASA 2013 shows that the Dalits are still performing poorly in Science (35%) than the other caste/ethnic groups. The overall difference between the groups is statistically significant ($p < 0.001$), but the effect size is small (f = 0.16). Students' ethnic/caste background explains only 2.4% variation in the result ($\eta^2 = 0.024$). In terms of Tukey's *post hoc* test, Brahman/Chhetri and minorities, Dalit and Madhesi do not differ from each other statistically. Otherwise, all the means of the ethnic/caste groups differ from each other statistically with the significance of at least at $p < 0.05$.

Table 5.3.9 Dalit students' achievement in various ecological zones and development regions

	Eastern	Central	Western	Mid-western	Far-western	Valley
Mountain	32	37		24	27	
Hill	31	33	39	33	36	
Tarai	30	22	33	30		
Valley						50
Total	30	28	38	32	30	50

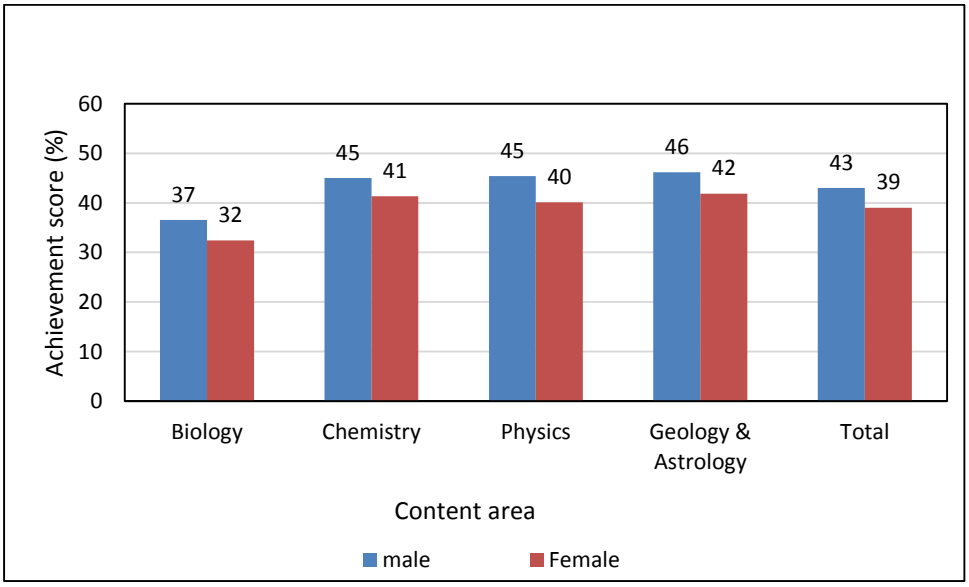
Except in the Kathmandu Valley, Dalit students' achievement in all development and ecological regions is below the national average. However, Dalit students' achievement in Science in the Kathmandu Valley is above the national average.

Dataset indicates that, except for the Kathmandu Valley, Dalit students' performance is lower than the other ethnic/caste groups. Their performance is lowest in the Central Tarai.

Gender and Student Achievement

A significant difference between the genders was found in earlier assessments (see ERO, 2012, ESD, 2008). Gender difference in achievement is also found in Science in this assessment. Figure 5.3.6 presents the comparison between the achievement scores of boys and girls in Science.

Figure 5.3.6 Comparison of achievement of boys and girls in various content areas of Science



A statistically significant difference between boys (43%) and girls (39%) has been found in NASA 2013 ($p < 0.001$). It is noteworthy that in all content areas boys have outperformed girls. However, the effect size

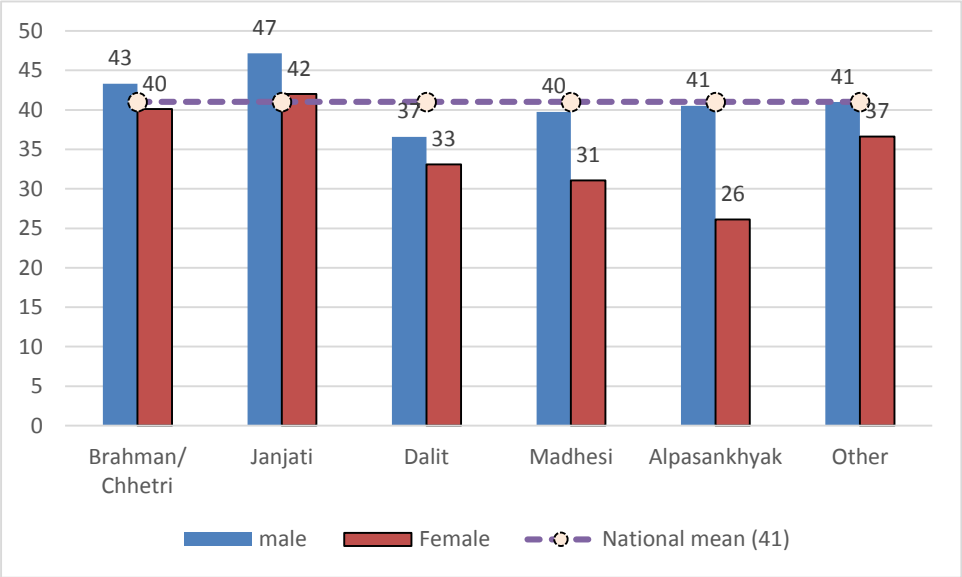
is small (f = 0.11), indicating that the difference is very low. Gender explains only 0.2% (η^2 = 0.024) of the variation in student achievement.

Boys have outperformed girls in all content areas and difference is statistically significant in Science, however, the effect size is small.

Gender and caste/ethnicity and achievement

In NASA 2013, difference in achievement between boys and girls is the highest within the minorities (15%), Madhesis (9%) and Janajati (5%). Tukey's *post hoc* test shows that the differences are statistically significant with at least at 5% risk level. There is statistically significant difference from gender point of view within all castes/ethnic groups, however, wide difference is found within the minorities and Madhesis (see figure 5.3.8).

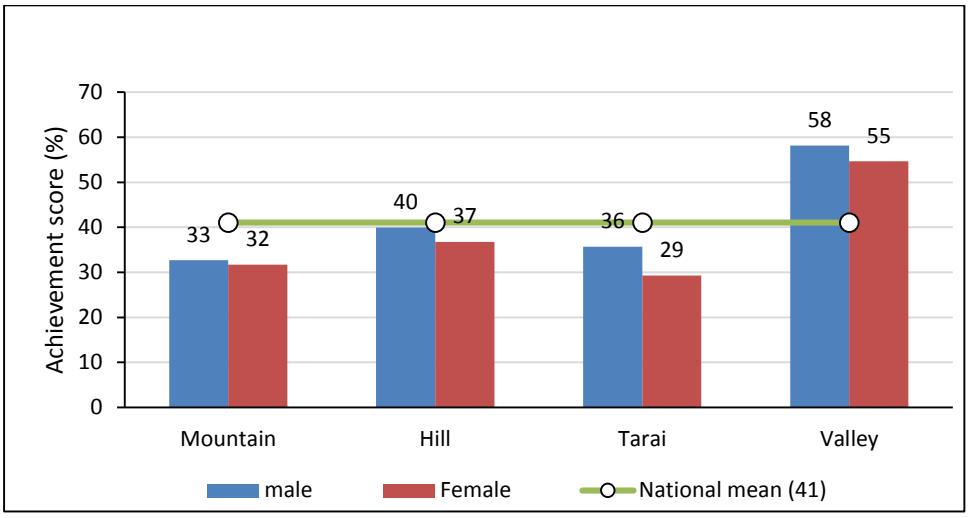
Figure 5.3.7 Comparison of achievement in Science between boys and girls in various ethnic/caste groups



Achievement by gender and ecological zone

The achievement of girls and boys in science differs significantly across three ecological belts: Kathmandu Valley, Tarai and Hill (at p <0.001); however, the difference is not significant in the Mountain region. Mainly, in the Kathmandu Valley, Hill, and Tarai regions boys outperform girls. However, in the Mountain, the achievement scores of boys and girls are the same (see figure 5.3.8).

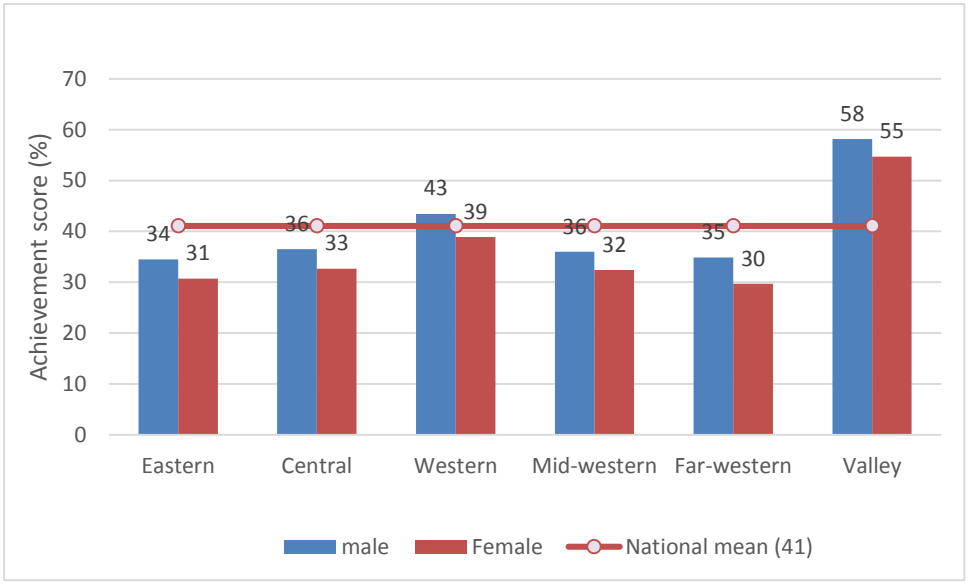
Figure 5.3.8 Comparison of achievement of boys and girls in various ecological zones



Gender, developmental regions and achievement

The difference between the boys and girls is the widest by 5% in the Far-Western region whereas it is smaller in other regions including the Kathmandu Valley, ranging from 3% to 4%. All the differences are statistically significant (at $p < 0.001$) (see figure 5.3.9).

Figure 5.3.9 Comparison of achievement between boys and girls in various developmental regions in Science



Dataset shows that the boys are outperforming girls in all content areas of science. Differences are not very wide though they are significant. Differences are wider in Western and Far-Western regions.

5.4 Selected Explanatory Factors and Achievement

Many of the factors have already been handled in previous sections. They are geographical factors such as districts, ecological zone, developmental region as well as school-related factors such as school type and school location. Some individual related factors such as home language, ethnicity/caste and gender also were handled. In this section, several other factors such as the socio-economic status (SES) of the students' families, paid work beyond school time, students' attitude towards science as a school subject, age of the student, and support provided for studies, which are mainly the family and individual related factors, are taken into consideration. As a sample of deepening school and teacher-related factors, availability of text books, homework given by the teacher, and selected activities in the schools are handled.

Socio-economic Status (SES) and Student Achievement

There were many variables indicating the socioeconomic status. In NASA 2013, they were categorized into parents' education, parents' occupation, home possessions (whether or not the student has his own space to do homework, or a dictionary, for example), home accessories (number of mobile phones, radio and so on in the students' home), and whether the student attends a private school or not. Finally, the SES is estimated on the basis of seven indicators related to the economical, educational, and occupational background of the family. In this section, the education including literacy of the parents is further elaborated to look at the relation in science achievement.

Several SES-related variables were analysed by using a data-mining tool of SPSS-the Decision Tree Analysis (DTA). The method is very effective in finding the cut-offs of the predicting variables, such as mother's education, and classifying the factor into several groups which differ statistically in the most significant way from each other in relation to student achievement.

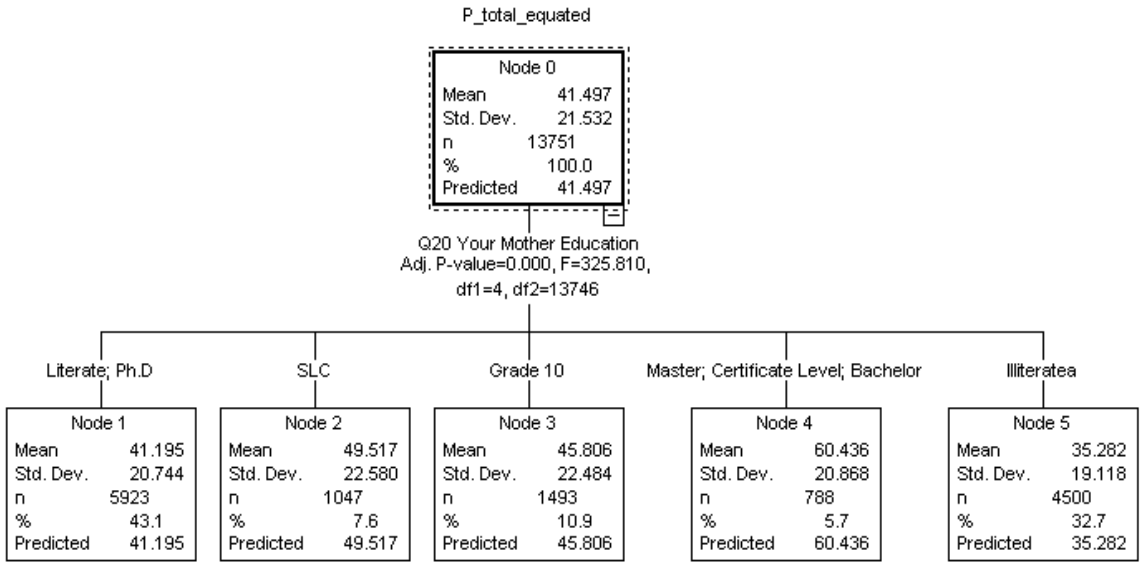
Parents' education

In NASA 2013 background questionnaire, the parents' education is divided into seven categories. These seven categories are: 1) illiterate, 2) literate, 3) grade 10, 4) SLC, 5) Certificate level, 6) Bachelor's level, 7) Master's level and above.

DTA classifies mother's education into five groups with statistically significant differences in students' achievement levels, for example, students of illiterate mother obtained 35% score on and students of having bachelor's and master's degree holder mothers obtained 59% score. In each group, the number of mothers

is high enough to make a credible prediction. The difference between each group is statistically significant (at $p < 0.001$) (see figure 5.4.1).

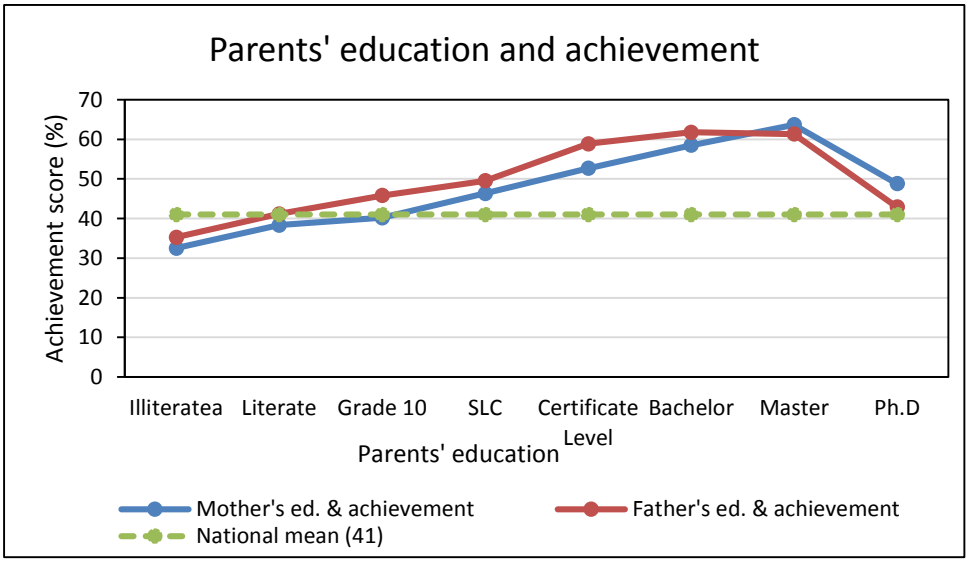
Figure 5.4.1 DTA of mother's education and students' achievement in science



DTA analysis shows that there is a remarkable effect of mother's education in student achievement, m which means: higher the mother's educational qualification, the better level of student achievement.

Similar pattern is found regarding the relationship between father's education and student achievement. A comparative illustration of father's education in relation to student achievement is presented in figure 5.4.2.

Figure 5.4.2 Parent's education and students' achievement in science



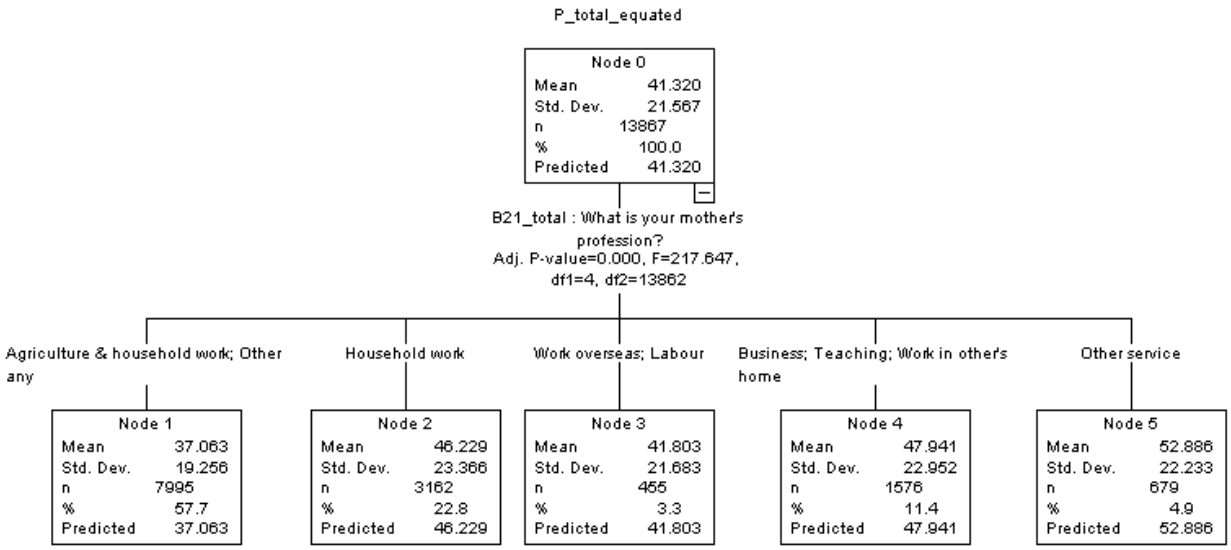
There is a positive relation of parents' education and student achievement. Mother's education can explain somewhat 9% (R^2 = 0.086) and father's education 13% (R^2 = 0.132) of the variation; however, a strong relationship was found between mother's education and student achievement, for example, Pearson's correlation with mother's education r = 0.278 and father's education r = 0.179. DTA shows that when both parents are illiterate, students achieved very low (32%) and highest (60%) when mother is grade 10 passed and father is certificate level or higher level.

Dataset indicates that the educational level of the parents predicts the children's future achievement level in science. The situation is worse when both parents are illiterate, and highest performance is found when mother is at least grade 10 passed and father's qualification is certificate level or higher.

Parent's occupation

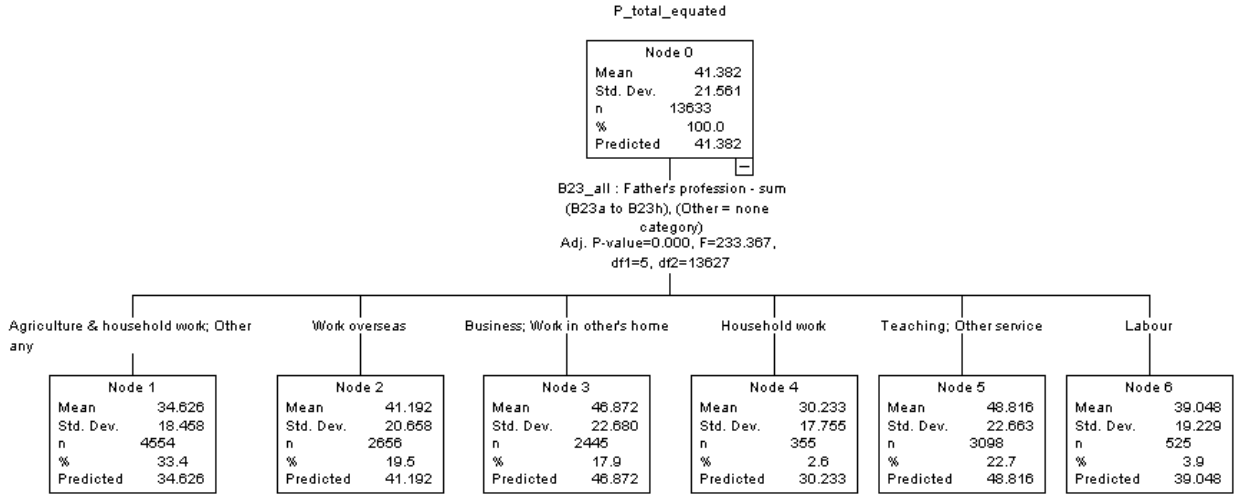
The occupation of parents was grouped into eight categories. These categories were: 1) working abroad 2) agriculture, 3) household work, 4) teaching, 5) other job, 6) business, 7) daily wages 8) work in others' home. In a similar manner as with the parents' education, the DTA was used to find the statistically most deviating groups related to student achievement. Student achievement is the lowest when the mother's occupational background is agriculture, household work and any other jobs (37%). The achievement is higher when the mother is from "working abroad/daily wages" (41%), "household" (46%) or "business, teaching and work in other's home" (48%). It shows that either economical or intellectual ability or both contribute the children to raise their achievement level.

Figure 5.4.3 DTA of mother's occupation and students' achievement in Science



Regarding father's occupation, the main division is whether the father works in "other's home". Student achievement is lowest when the father's occupation is household work (30%) and "it is 35% when father's occupation is "agriculture and household work" and "non-category works" (see figure 5.4.3). If the father was "teacher or in other service ", the students achieved 49%.

Figure 5.4.4 DTA of father's occupation and students' achievement in Science



Combining the mother's and father's occupation, the lowest achievement is found in the families where both parents are from agricultural background or any other non-category works (24%), or when the father works in other's home (30%). Best performers were those whose parents are from business, teaching and other services (56%), having remarkably high achievement compared to national average (41%).

The dataset indicates that if the father or mother or both are from agricultural and household work, working in other's home, the students' achievement in science is significantly lower than the other occupational groups.

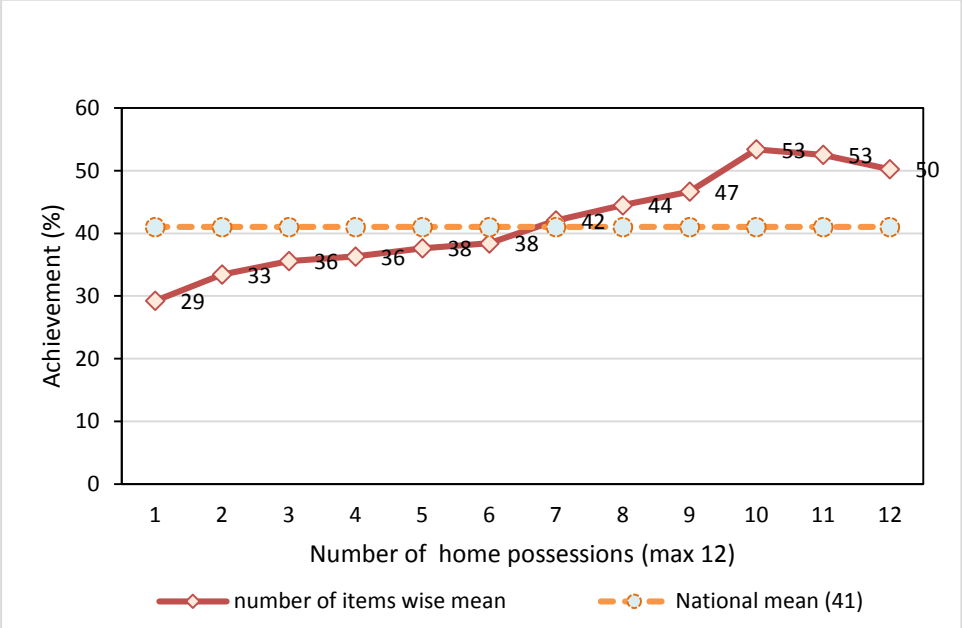
Home possessions and accessories

Facilities and resources available at home have some effects on student achievement. There were two kinds of home possessions defined in the background information questionnaire for the students. One is related to the facilities that support the study at home, for example whether they have a table for study, a separate room for them, a peaceful place for study, a computer for school work, software for the computer assisted learning, internet facilities, their own calculator, access to classical literature, poetry books, or artistic things like pictures, and books that help them for study such as a dictionary. Another category of home possessions

includes different types of home accessories such as the number of mobile phones, televisions, computers, and radios.

There were 12 questions in the student background questionnaire related to home possessions. Each scored 1 if the student had access to this possession. Adding these items up, the maximum score was 12. Figure 5.4.5 shows that home possessions and achievement level are positively correlated. Pearson product moment correlation coefficient between the achievement level and the factor (r = 0.27) is statistically significant (p < 0.001) and the effect size is medium (f = 0.28).

Figure 5.4.5 Relation between the home possessions and achievement in Science



The same pattern, that is the more possession, the better results, can also be seen in the case of home accessories (seen figure 5.4.5.) The question in the background questionnaire was set differently compared with home possessions. Regarding the accessories the question asked was "how *many* of the following accessories do you have in your family?" with the options 0, 1,2, 3 (or more). After dichotomizing the items individually by using meaningful cut-offs as found with ANOVA and DTA (and maximizing the differences in achievement level), all the five indicators were summed. The maximum score was 4 indicating that the students possessed a set number of all the accessories.

Table 5.4.1 Dichotomizing the indicators for home accessories

Accessory	cut-off for 1	cut-off for 0
Mobile phone	2, 3	0,1, missing
Television	1 to 3	0, missing
Computer	1 to 3	0, missing
Radio	1 to3	0, missing

In all of the cases, there is a positive relation between home accessories and student achievement. However, as the number of radio increases, achievement decreases. Hence, radio was excluded from the home accessories plot as shown in figure 5.4.6.

Figure 5.4.6 Relation between the number of home accessories and achievement in Science

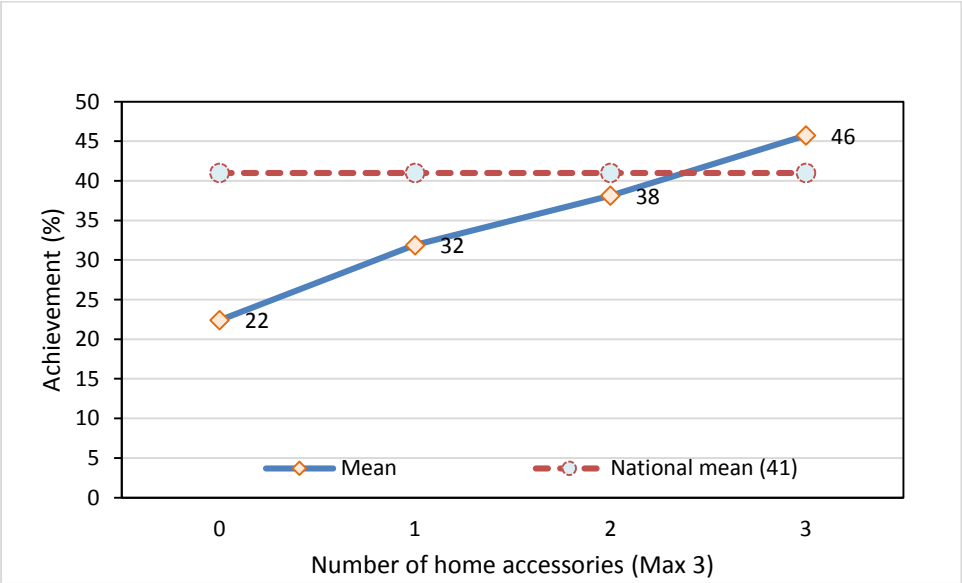


Figure 5.4.6 clarifies how the increased number of home accessories has affected the students' assessment. It is found that students' achievement increases from 22% (none of them are available) to 46% (all of them are available). Availability of all the stated facilities indicates the high SES of the family. Correlation between home accessories and achievement is $r = 0.27$ ($p < 0.001$), which is certainly positive though it is not high. The effect size $f = 0.28$, which indicates a medium size of difference between the groups. The difference between the lowest group with no mobiles phones, television and computer (22%), and the highest group with all these accessories (46%) is remarkable. Note that 26% of those students who have the access to all three accessories are from the Kathmandu Valley.

Data shows that when the children have very few home possessions, that is 0 to 3 out of the 12, the achievement level is statistically lower (below 35%) than if there are more than seven (> 40%). With ten to twelve possessions the average score is very high (> 50%) compared with the national average. The same pattern shows similar results with home accessories. With no or only one accessory indicator met, the results are very poor (22 to 32%) and when there are all three accessories, the results are remarkably higher (46%). This indicates that student achievement is positively associated with the possibilities of access to support materials.

SES and Achievement

The socio-economic status was defined on the basis of seven indicators which were all first dichotomized. The variables, for example, mother's education, father's education, mother's occupation, father's occupation, home possessions, home accessories, and type of school where students were studying, were summed as SES and changed into a percentage of the maximum score (PSES). The deeper description of the transformations is seen in previous section. The PSES represents the percentage of SES the student possesses; 100 means that the student has the highest SES possible measured with these variables and with these transformations, that is, all the seven indicators of SES positive, and 0 refers to the lowest possible SES, that is, all the seven indicators of SES are negative. The analysis of the PSES by using Univariate GLM (that is the regression modelling) shows the strong relation between SES and achievement. Figure 5.4.7 presents the relationship between SES of the students and their achievement.

Figure 5.4.7 Relation between SES and achievement in Science

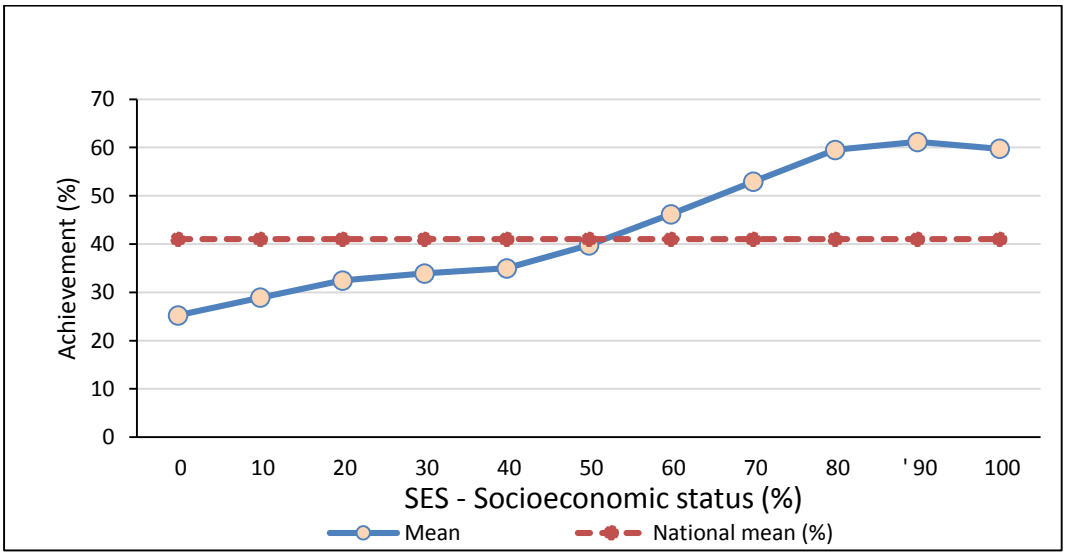
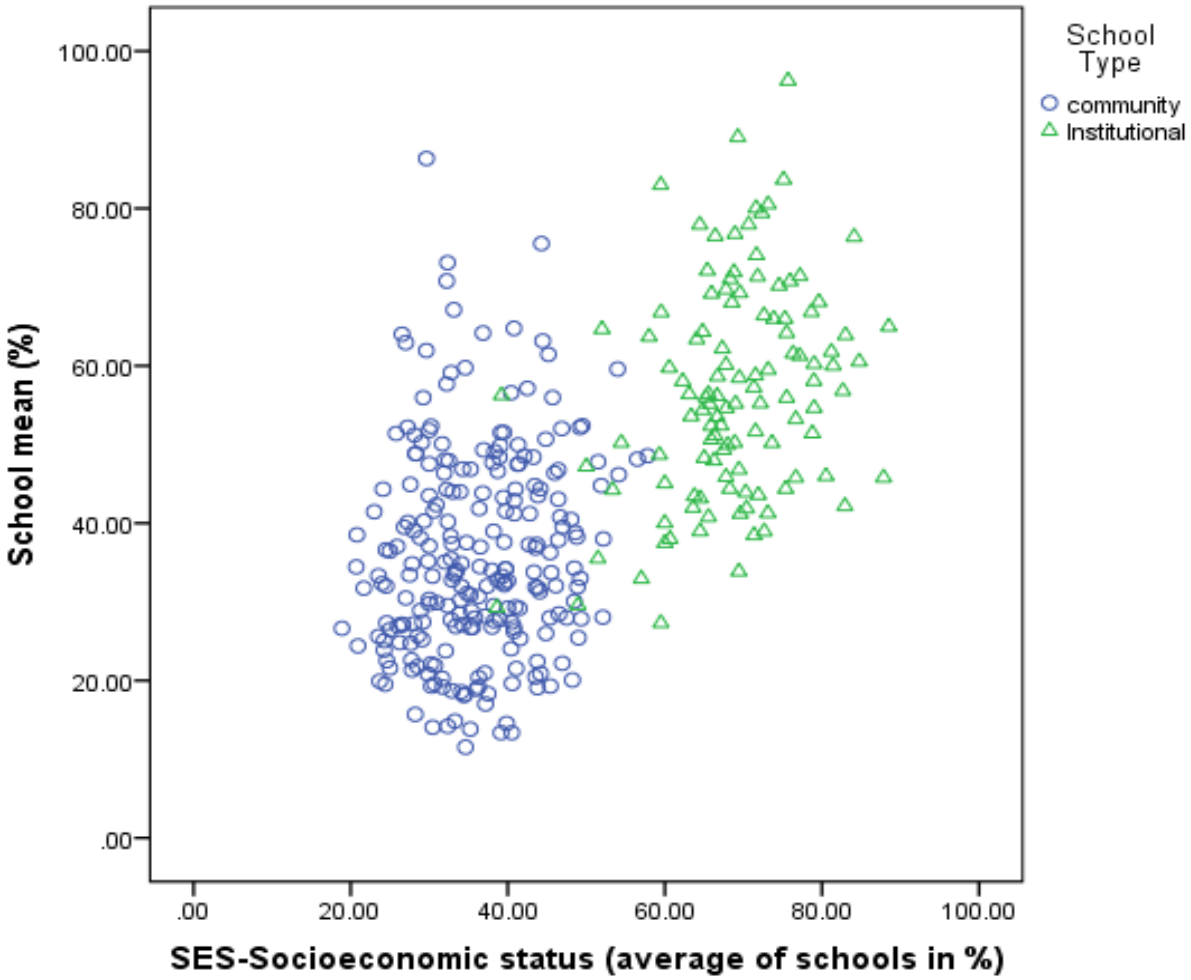


Figure shows a positive relationship between SES and the achievement; the correlation between the variables is $r = 0.44$, which is a significant association ($p < 0.001$). The differences between the SES groups are statistically significant (ANOVA, $p < 0.001$), the effect size is high ($f = 0.52$); that is, the highest groupand lowest group differ from each other remarkably. SES explains 21% of the variation in student achievement ($\eta^2 = 0.209$).

By analysing the matter further with the scatter diagram as in figure 5.4.8, we see that two types of schools, community schools in circle and institutional schools in triangle, fall into two groups where most of the institutional schools are performing relatively high but the community schools vary from very high-performing schools to very low-performing schools.

Figure 5.4.8 Distribution of achievement by socio-economic status and type of schools in Science

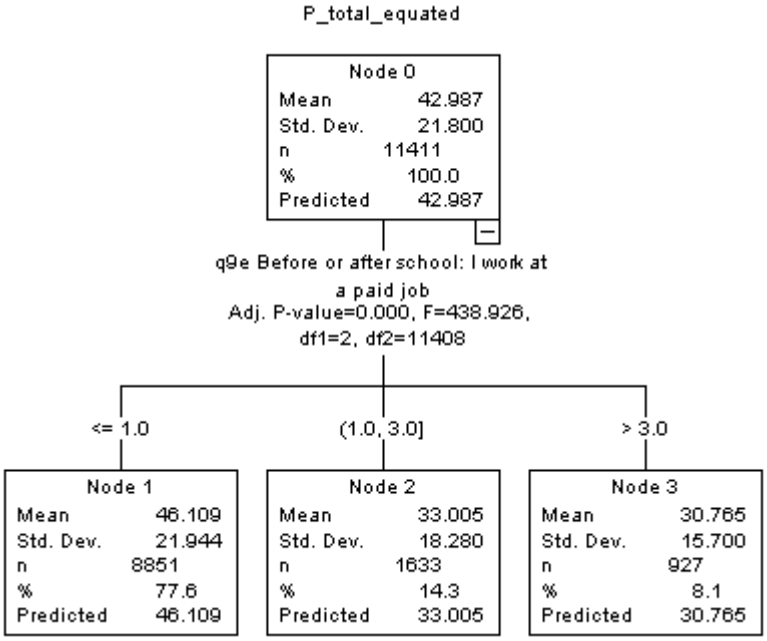


The dataset shows that the students in the institutional schools have performed relatively higher than and the students in the community schools.

Working Beyond the School Time and Achievement

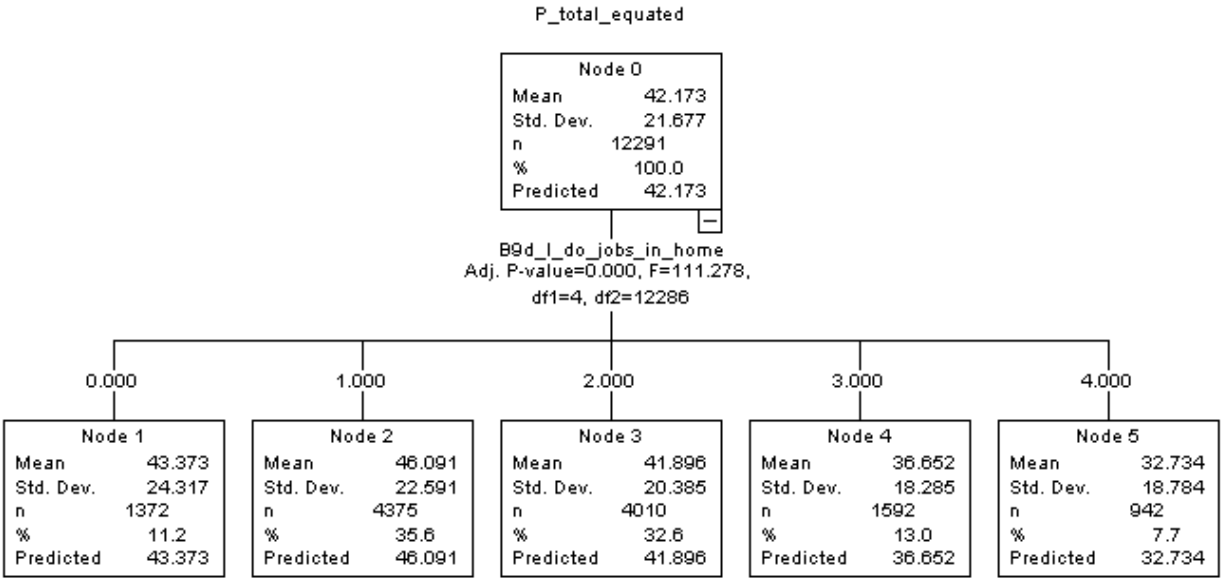
Questions were set in the student background questionnaire related to the students' activities beyond the school time. The questions were related to 'working before/after the school in a paid job' and 'involving in household chores'. The values of the variables are divided into five categories, namely, 1 point for working not at all, 2 points for working less than 1 hour per day, 3 points for working 1 to 2 hours per day, 4 points for working 2 to 4 hours per day, and 5 points for working more than 4 hours per day. DTA shows that when the children are not engaged in paid work *at all*, the results are notably above the national average (46%) (see figure 5.4.8a). When the students involved in paid work even less than two hour, the results were remarkably lower (30%) than the national average. The differences are statistically significant ($p < 0.001$) though the effect size is moderate ($f = 0.28$), that is, the number of children working in paid jobs is not so big. Working beyond school indicates that the family is poor and the extra earnings are needed. When the student needs to work more than 4 hours per day there is no time or energy to handle school homework.

Figure 5.4.8a DTA of paid work and achievement in Science



It is usual practice in families that children support in the household chores at home, which is a part of the socializing process of the children. The DTA shows that when the child spends more than 2 hours per day doing household chores, the results are as poor as (< 37%) if the student needs to work in a paid capacity (see fig. 5.4.8a). However, when the amount of time spent in household work is less than one hour per day, the achievement level is higher than the average (see figure 5.4.9). Differences are significant ($p < 0.001$) though the effect size is small, that is, a few children are participating in household work for more than 2 hours per day.

Figure 5.4.9 DTA of working beyond the school time and achievement in Science



The dataset indicates that working beyond the school time reduces the achievement of students. The phenomenon is most probably connected with the poor economic condition of the family. Especially, when the children need to work for paid job and spend more than 2 hours per day in household chores, the achievement level is remarkably low. If the students involve in household chores up to 1 hour, the achievement is found high.

Student Attitude Towards the Subject of study and Achievement

In the context of Science achievement assessment, attitude tells us what the students think about science and its usefulness in their daily life and future utility. More or less, relationship is found between the attitude of the students and achievement. Though the connection is not always clear, the correlation between science achievement and attitude toward science has widely been studied (see, for example Metsämuuronen 2012a; 2012b; House & Telese, 2008; Shen & Tam, 2008; Kadijevich, 2006; 2008).

In NASA 2013, the same shortened version of Fennema–Sherman Attitude Scales (FSAS, Fennema & Sherman, 1976) as used in several international comparisons, like in TIMSS and PISA studies, was used to identify the relation between attitude towards the subject and their achievement. The original scales included nine dimensions but in these international comparisons only three are used with four items on each and two negative items on each of the first two dimensions. The names of the factors can be "Liking Science", "Self-Efficacy in Science", and "Experiencing utility in Science" (compare naming in, e.g., Kadijevich, 2006; 2008). Factor analysis was used to identify the factors of the responses in FSAS and the negative items were reversed to make the whole test unidirectional. As in several countries of Asia, the expected factor structure cannot be found in Nepal (for a deconstruction of the test scales, see Metsämuuronen, 2012a; 2012b). Hence, only the total score is used to show the connection of attitude and achievement. The relation between attitude of students towards science subject and achievement score is shown in figure 5.4.10.

Figure 5.4.10 Relation between students' attitude towards the subject and achievement in Science

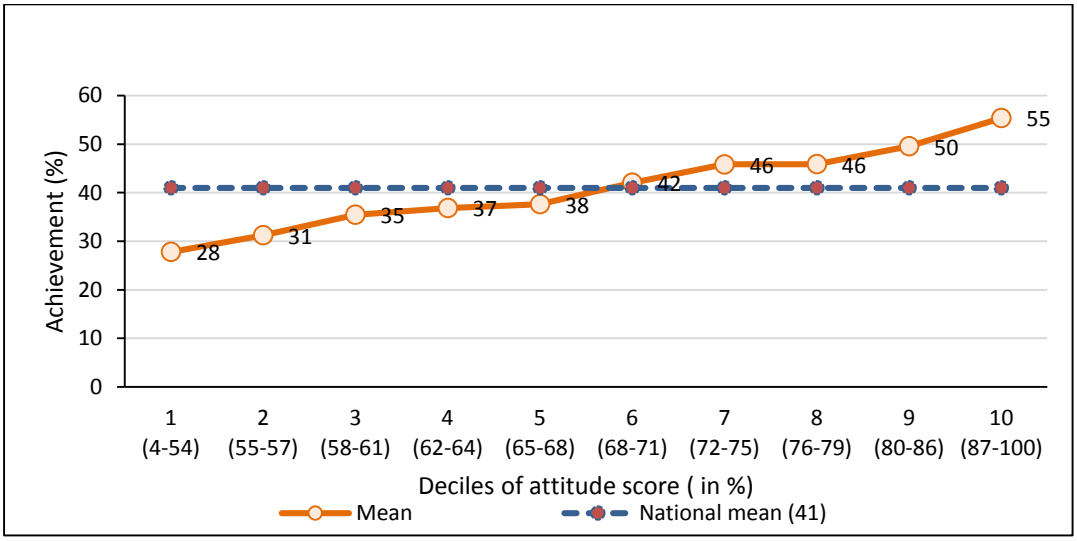


Figure 5.4.10 shows that the difference between the lowest attitude group (28%) and highest attitude group (55%) is remarkable. Positive attitude towards science subject can positively affect the achievement. The correlation between the positive attitude towards science and achievement is $r = 0.35$ ($p < 0.001$); the effect size is high ($f = 0.40$).

Dataset reveals the fact that the more positive the attitude towards science, the higher the achievement. The data also supports the fact that positive attitude influences the positive achievement (and not the other way round).

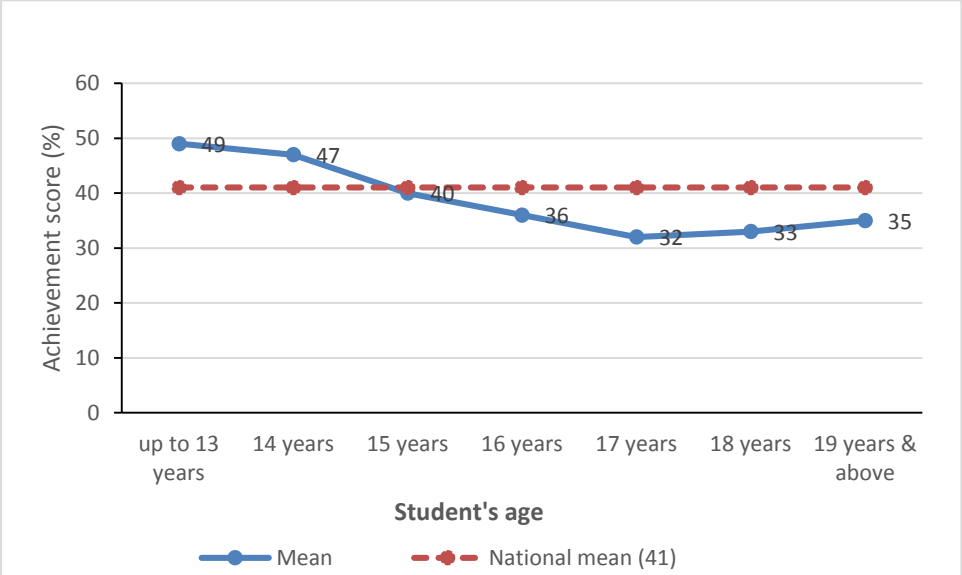
Age of the Student and Achievement

In Nepalese context, although the official age for grade eight students is 13 year, the age of the students attending grade eight varies widely. Some students have mentioned their age below thirteen years and some above 20. All the ages of the students below 13 years were encoded as '13 years', and all students above 19 were encoded as '19 years and above'. The descriptive statistics of the mean for each year are is presented in table 5.4.2 and further visualized in figure 5.4.12.

Table 5.4.2 Descriptive statistics of the students' achievement in different age groups

Age	N	Mean	SD
Age of students	N	Mean	SD
13 years	3259	47	22.4
14 years	4599	45	22.1
15 years	3395	37	19.9
16 years	1456	32	17.0
17 years	530	29	15.4
18 years	207	27	14.2
19 years and above	125	31	17.0

Figure 5.4.12 Relation between the age of the student and achievement in Science



It is evident that the best achievers are those students who are at the proper age for grade eight studies, that is, 13 and 14 years scoring 49% and 47% respectively. The higher age means either the students have started much later than they should have, or they have repeated the classes. The achievement level is remarkably lower when the students are at the age of 16 or above (32–36%). Correlation between the variables is -0.26 ($p < 0.001$), indicating small effect size ($f = 0.29$). The ANOVA hints that the age (the prolonged studies) explains the achievement level by 8%.

Dataset indicates that the highest performance is with those students in their proper age group, that is, at the age of 13 and 14 years. Otherwise the achievement decreases as the age increases.

Support Provided for Study and Student Achievement

The relation between the support in studies and achievement was analysed based on the responses to the question *"who supports you when you do not understand what you have studied or felt difficult?"* In the question, only one option was selected. In many cases, there might be several supporters which cannot be analysed in this study. The descriptive statistics of the supporters are given in table 5.4.3.

Table 5.4.3 Descriptive statistics of helpers of the students

Support from	N	Mean	SD
Teacher	8561	40	20.7
Mother	256	42	23.7
Father	553	42	23.6
Brother/Sister	2745	44	22.4
Consult with more than one	465	44	24.2
No one	183	45	23.0
Tuition	988	48	21.6
Total	13751	41	21.6

From the dataset we see that the support in science is necessary for the students to obtain better than average marks on the test. There is about 2 to 8 point difference between those who get support from teacher and the support from taking tuition. Tuition contributes to advance the achievement in science (48%).

The dataset indicates that the support provided by the student's brother/sister raises the achievement level more than the of support provided by father and mother. The highest achieving groups are those who receive private tuition. It is likely that this group also spends more time on homework which may explain the high score.

Availability of Textbook and Student Achievement

The data shows that there were some students who did not have Science textbook up to the end of the academic session. Table 5.4.4 shows the descriptive statistics of the availability of the textbook and the achievement (mean).

Table 5.4.4 Availability of textbook of Science and the achievement

Availability of Science textbook	N	Mean	SD
Yes	13140	42	21.5
No	372	34	21.2
Total	13512	42	21.6

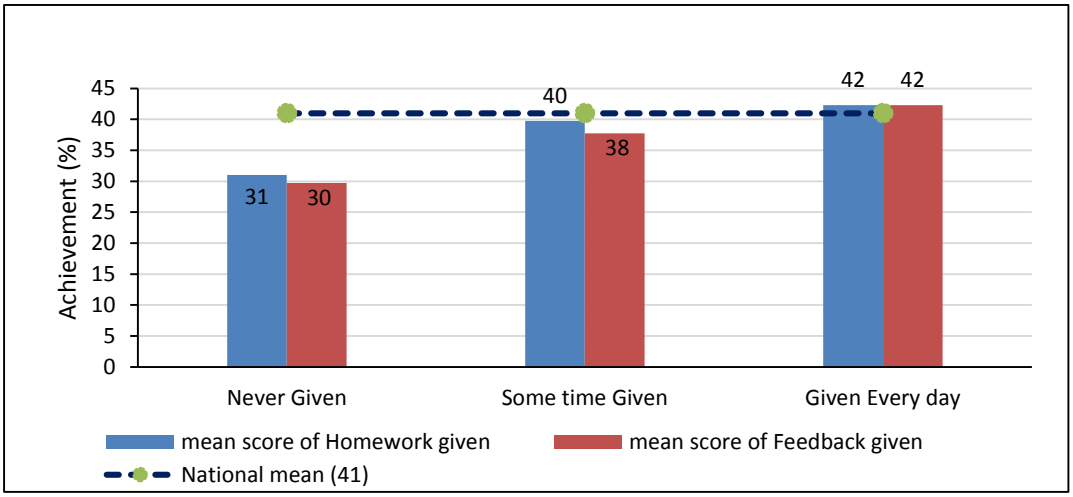
Out of 13512 students who responded to the question, 2.8% of them did not receive a science textbook even up to the end of the academic session. The relation between the availability of textbook and achievement is significant ($p < 0.001$) though the effect size is small due to the small size. The difference in achievement is quite big, that is, 8%.

Data shows that 2.8% of the students did not receive the textbook in science even at the end of the academic session. The achievement level of these students is significantly lower than those who have access to the textbook.

Homework Given/Checked and Achievement

Homework is considered as one of the ways to enhance learning, which can be used as drill, exercise or an evaluation tool. When homework is regularly checked, it is likely to boost achievement level. Statistics related to homework given and checked is condensed and visualized in figure 5.4.13.

Figure 5.4.13 Effects of homework in students' achievement in Science



If the teachers assign homework and provides feedback regularly, students achievement is found higher (42%) than when homework is not assigned (30%). The difference is statistically significant ($p < 0.001$). Those groups without homework or no feedbacks provided are, however, small and hence the effect size is also small ($\eta^2 = 0.001$). Dataset does not show any positive effect if homework is given but not checked.

Dataset explains that if the teacher assigns the homework and provides feedback regularly, the achievement is higher than when homework is or not checked.

Positive and Negative Activities in the School and Student Achievement

The activities of the students and teachers determine the learning environment of the school. Bullying, for example, is one of the hindering incidents for the students in school that may affect learning. In the student background information questionnaire, several student-related and school-related activities were asked, some of which are positive and some are negative. Here, bullying is handled as one of the negative, and students' impressions of school and teacher's activities are taken as of positive indicators.

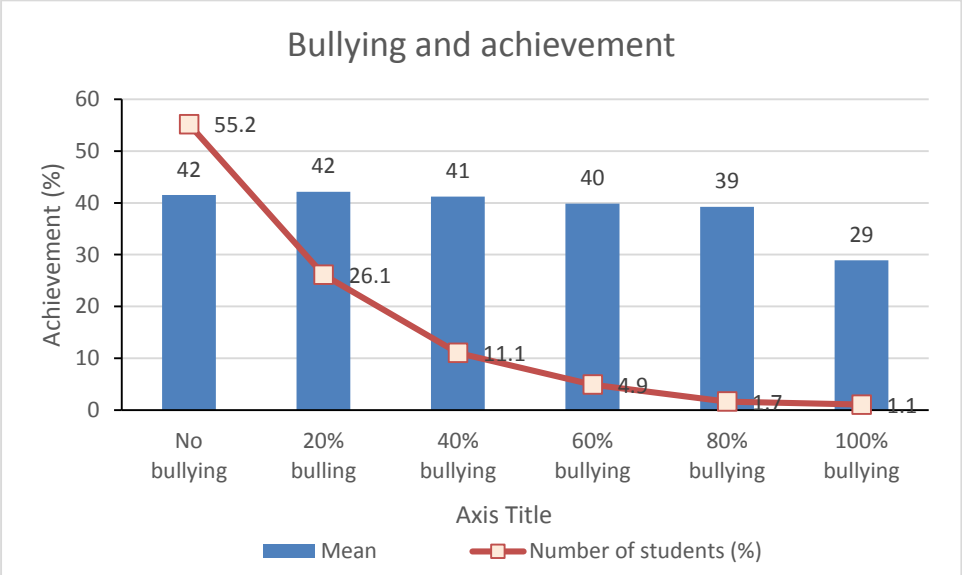
Bullying at school

Bullying is one of the problems in the school that worsens the learning environment for the students. International Studies like TIMSS and PISA emphasize to identify such indicators. In NASA 2013 student background questionnaire, five questions indicate the varieties of bullying that tend to happen in schools. All the incidences were stemmed by the phrase "*which of the following activities happened in your school in the last month?*" The students' response is presented in table 5.4.5 and further visualized in figure 5.4.14 'No' (%) indicates the percentage of the students' response which describes that no such activity happened in the school and 'Yes' (%) indicates the percentage of the students who experienced the particular type of bullying happened within the last month. Altogether 44.8% students have mentioned that they had experienced at least on incident during the last month.

Table 5.4.5 Bullying and achievement in Science

Type of Bullying	No (%)	Yes (%)
Something of mine was stolen	68	24
I was made fun of or called names	78	14
I was hit or hurt by other student(s)	81	10
I was made to do things I didn't want to do by other students	75	15
Fellow students kept outside without involving me in activities	82	9

Figure 5.4.14 Effect of bullying in students' achievement in Science



The sum of all five items is done as an indicator of bullying. Figure 5.4.14 shows the extent bullying with the percentage of the students and achievement of the students in each category of bullying. If only one activity of bullying is reported, it is categorized as 20% of bullying, and if all five activities are reported it is categorized as 100% bullying. Knowing that 55.2% of the students did not encounter any kind of bullying during the last month, one can infer that the remaining 44.8% have encountered at least one type of bullying. This is a remarkable number of students. About 1.1% of the students are experiencing some kind of severe bullying. It is found that learning outcomes are remarkably lower with that 1.1% of the students who have encountered more than two kinds of bullying including all five types of harassments (29%). There is 13% achievement gap between the students who did not experience bullying and students who experienced extreme bullying of all five kinds. Though very small number of students reported all kinds of bullying (n = 144), However, the difference is statistically significant (p = 0.001) and the effect size is small (f = 0.06). Though the extreme case of severe bullying is rare, bullying of some kind seems to be quite common in schools. This negative phenomenon causes harm to young children and has to be rooted out from schools.

The dataset indicates that a high number of the students (45%) encountered bullying in schools within the last month of the this assessment. The phenomenon does not have a great effect except in the group of extremely bullied students. Further, efforts are required to root out the phenomenon from schools.

Positive activities at school

The activities that can boost the learning and achievement of students are categorized as positive activities. Such positive activities happened in the schools were asked from the students in two sets of questions listed in table 5.4.6. The table shows the responses of the students in all four categories, which are in the 4-point rating scale, anchored to fully agree and fully disagree.

Table 5.4.6 Students' response towards teacher and school-related activities in the schools

Teacher and Students activities	Respondents in %			
	Fully agree	Partially agree	Partially disagree	Fully disagree
Students get along well with most teachers	73.2	18.4	3.2	1.4
Most teachers are interested in student's well-being	82.2	9.6	2	1.5
Most of the teachers really listen to what I have to say	56.3	29.9	5.7	2.2
If I need extra help. I will receive it from my teacher	74.7	15.6	2.4	1.5
Most of my teachers treat me fairly	51.8	20.8	7.3	11.5
I like to come and stay in school	85.5	6.8	1.5	1.9
Students in my school try to do their best	72.8	18.6	2.5	1.5
Teacher in the school care about the students	77.2	13.4	2.7	1.5
Teacher wants the students to do their best	86.4	5.6	1.5	1.7
Average	66.1	14.1	3.2	2.9

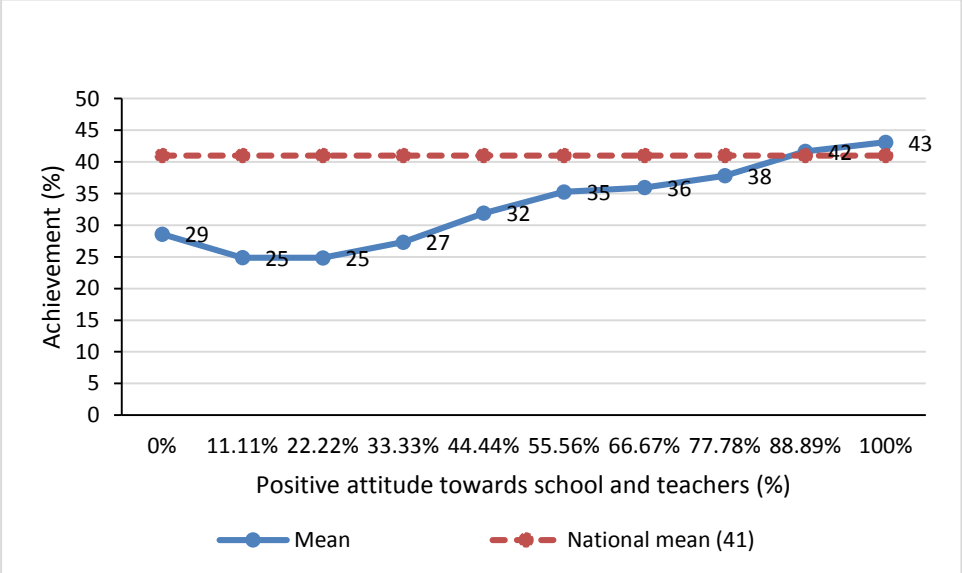
Further analysis was carried out by recoding the variables into two categories, 0 for disagree and 1 for agree. Furthermore, the sum of nine indicators is converted into the percentage of maximum score to analyze the level of positive activities and its relation to the achievement. The overall result is that the feeling of the positive actions in the school relates positively with the student achievement. The correlation between the sum of nine positive activities and achievement is found positive ($r = 0.11$, $p < 0.001$).

DTA finds 6 attitude groups in the indicator. These boundaries and descriptive statistics are seen in table 5.4.7 and further illustrated in figure 5.4.15.

Table 5.4.7 Positive activities at school and student achievement

% positive attitude	N	Achievement	SD
0 – 73.3	1384	38	19.8
73.3 - 81.5	1258	39	20.2
81.5 - 85.2	1139	41	20.8
85.2 - 87.5	1632	42	20.4
87.5 - 96.3	4270	43	20.9
96.3 - 100	6251	46	22.0
Total	15934	43	21.3

Figure 5.4.15 Relation between positive actions in school and achievement in Science



The data shows that there is a positive relation between the student's feeling about the school activities and achievement. Increase in achievement is directly proportional to the increase in the intensity of such activities. After dividing the indicator into six groups on the basis of DTA, the differences between the groups are statistically significant ($p < 0.001$); however, the effect size is small ($f = 0.17$). Only when the

students are extremely positive towards school and teachers' behaviour, the learning achievement is much higher than the average (43%). However, the difference between the most positive group and the most negative group is notable, that is 18%.

Dataset reveals that when students have a feeling that the actions of the teachers and the school are ultimately good, the results are better than the national average (43%). At the other extreme, when they have a feeling that such actions are negative, the results are far below the national average (25-29%).

5.5 Summary of Findings

The main findings on **Science** subject at grade eight in NASA 2013 are as follows:

Basic results

- The average achievement score in Science is 41%, and the achievement is normally distributed as there were three populations: low-performing, medium-performing and high-performing students.
- The learning outcomes are lower in Biology than in all other content areas such as chemistry, physics, and geology and astrology.
- The students' ability to solve complex problems is low as only 23% scores were obtained in higher ability items. Students are much better in the recalling and comprehension type of questions (49% each). The students are good in recognizing the correct answer and in very fundamental knowledge, such as choosing the facts and numbers, and writing the definitions. They are much weaker in reasoning, problem solving, plotting, proving the theory or formula, and constructing the shapes and figures. In many cases, the students did not even start to do the open ended questions and hence, they got poor score.
- The average achievement of Science in Nepal (mean = – 0.18) is lower than the international average –TIMSS (mean = 0.00). In all content areas the results are lower than the international mean.

Equity indicators

- There is a wide difference in students' achivement scores in science among the districts. The results in Dhanusa (24%) from Central region; Mugu (25%) from Mid-western region; Bajura (29%) from Far-western region; Panchthar (30%) and Sunsari (30%) from Eastern region are so poor that raising

the standard in these districts would raise the standard in the whole country. The Valley districts Kathmandu (59%), Lalitpur (53%) and Bhaktapur (51%) are at the top in achievement score.

- There is a remarkably wide difference between the highest and lowest achiever ecological zones in Science. Students in the Kathmandu Valley have outperformed the other zones. Variation in the achievement scores among the districts within the mountain zone is the widest.
- There is wide difference in students' achievement scores among the various development regions in Science. The difference between the lowest performing regions (Eastern and Far-western regions) (32%) and the highest performing region the Kathmandu Valley (56%) is remarkable.
- The students in the institutional schools outperform the students in the community schools. The students in the institutional schools are more concentrated on high performing group whereas the students in the community schools form two groups of schools: high-performing schools and low-performing schools. The variation between the community schools is also remarkable.
- The achievement scores are better in urban areas than in the rural areas, and the difference is medium.
- Dalit students' performance is lower than the results of the other castes/ethnic groups. Their performance is poor even in the Central Development Region if we exclude the Kathmandu Valley. However, Dalits are achieving high in the KathmanduValley.
- Boys are outperforming girls in all content areas of Science. Differences are not very wide though they are significant. Gender explains only 2% of the variation in achievement. However, the differences are widest in the Western and Far-Western regions.

Selected explanatory factors

- The educational level of the parents predicts the children's future achievement level in Science. Especially, the achievement level is low if both the parents are illiterate, and achievement is the highest when mother is at least grade 10 passed and father is having tion the qualification of certificate level or higher.
- If the parents are coming from an agricultural or related occupation and father is working in other's home, the students' achievement in Science is significantly lower than the other occupational groups.
- When the children have very few home possessions, (i.e., 0 to 3 out of the 12), the achievement level is lower (below 35%) than if there are more than seven home possessions (above 40%). Children having 10 to 12 possessions have a very high (> 50%) scorecompared with the national

average. The same pattern shows with home accessories, as the results are very poor when as no or only one accessory indicator is met, (22 to 32%) and when there are three accessories, the results are remarkably higher (46%). This indicates that student achievement is positively associated with the possibilities of access to support materials.

- Working beyond the school hours reduces the school achievement of the students. The phenomenon is most probably connected with the poor economic situation in the family. Especially, when the children need to work more than 2 hours per day either paid or unpaid, the achievement level is remarkably lower. Achievement is better if they are working 1 hour or less per day.
- The more positive the attitude towards the Science subject, the higher is the achievement. The data also supports the fact that positive attitude influence positive achievement.
- The highest performance is found with those students who are studying in their normal age group, that is, at the age of 13 and 14 years. Otherwise, the achievement decreases as the age increases.
- Study support provided by the mother raises the achievement level more than any other family members' support. The highest achieving groups are those who receive private tuition. It is possible that this group also had spent more time on their homework which may explain higher score.
- There were 2.8% students without textbooks in Science; and the achievement level of those students is significantly lower than the achievement of those who have access to the textbooks.
- If the teacher gives homework regularly and checks it, the achievement of students is higher than the achievement of those students who were not assigned homeworks and not given feedback. Dataset does not show any positive effect if homework is given but not checked, or not provided any feedback based on the homework.
- A large number of the students (45%) encountered bullying in schools within the last month of this test. However, the phenomenon does not have a great effect except in the group of extremely bullied students.
- When students have aeeling that the actions of the teachers and the schools are ultimately good, the results are better than average (43%). At the other extreme of feeling that such actions are negative, the results are far below the average (25 to 29%).

Chapter 6: Major Findings, Implication and Conclusion

It is acknowledged worldwide that schooling fosters knowledge and skills to youngsters and prepare them for better adult life on which prosperity of a nation largely depends. With this motif, nations around the world express commitment for ensuring universal schooling with good quality as a central part of development strategy in order to enhance skills and employability of youth, raising national productivity, and reducing poverty. So, it is an important concern for all parents, teachers, governments and general public to know how well the school education systems equip youths with knowledge and skills they need to make their lives better, to play a role in building more peaceful and equitable societies (Matsura, 2004), and to be able face future challenges. With this concern, measuring and monitoring students learning achievement by means of national assessment developed worldwide during the last decade of 20th century. For the same purpose at international level of assessment like PISA, PIRLS, TIMSS in industrialized countries; at regional level assessment like Southern and Eastern African Consortium for Monitoring educational Quality (SACMQ) in African countries, Segundo studio Regional Comparativeo Y Explicativo (SERCE) in Latin American countries(UNESCO, 2005, 2009) also were institutionalized during the same period. Nepal initiated it as National Assessment of Student Achievement (NASA) since the late 1990s.

This final chapter of NASA 2013 report begins with restating the main objectives of the assessment and briefly includes context of the study, the sampling method and process as well as test development, conduction and analysis process. In the next section, it states overall major finding of this assessment study. Before concluding, it draws the main implications of finding indicating required systemic reform based on the key results of this assessment.

6.1 Objectives of the Assessment

As other student assessments institutionalized at international, regional and national level, National Assessments of Student Achievement (NASA) carried out recently in 2013 for grade 8 in Nepal have some specific objectives that can be mentioned as follow:

- To determine the current national level of achievement of grade 8 students in mathematics, Nepali language and Science subjects;
- To determine variations in student achievement in the subjects among different ecological zones, developmental regions, districts, school location (rural/urban), school type (community/institutional schools), ethnicity and language groups, gender, and socio-economic conditions;

- To examine the extent to which the home background and other pupil related factors influence the learning achievement in the subjects;
- To compare student learning achievement in the current study with that of the previous studies of Nepal and international ones like TIMSS and PISA.
- To generate evidence based data for monitoring the trends in students achievement for these subjects over the period of time.

6.2 Summary of Process and Method of Assessment

Nepal is no exception in adopting national assessment, which introduced it since 1997 for measuring students' achievement in order for monitoring it and providing policy feed back to the system. With this objective, various assessments of student achievement were carried out for grade 3 in 1997 and 2001; grade 5 in 1999, 2003 and 2008, grades 6 and 8 in 1999 which were small scale in nature. After the establishment of the ERO, large-scale assessment of student achievement have been initiated for grade 3, 5 and 8 as envisioned by the SSRP 2009-2015. Accordingly, two round of NASA for 8 in 2011 and 2013 and one round for grade 3 and 5 in 2012 have already been accomplished, of which NASA 2013 is the latest one.

This study was conducted in 1199 randomly selected schools from 28 sample districts covering all ecological zones and development regions. Total number of student sampled for the assessment was 48,000, almost equally distributed in each of the three subjects. Each student participated in the assessment responded a set of background questionnaires together with one set of test item among three subjects Mathematics, Nepali language and Science. In addition, 1199 teachers and 1199 head teachers from sample schools were also participated in this assessment by responding related questionnaires.

Standardized test items were used to assess the students' achievement. Individual items were developed through the workshops of trained subject teachers and subject experts. The subject committee of each subject revised the items and selected at least six different items to measure the same performance in each subject. Selected items were pretested and results were analysed by calculating difficulty level of each items. Based on the pre-testing results of the items, the difficulty levels of the tests were set around 50 – 60%. Some linking items from NASA 2011 and international tests like TIMSS and PISA were selected and used. Items were analysed and equated using IRT modelling.

6.3 Major Findings

Present study revealed that average score of students' learning achievement varies across the subjects ranging from 35% in Mathematics, 41% in Science to 48% in Nepali. It however also shows that the achievement of students in various subjects has not been improving over the years. This level of achievement of Nepali students is considered low in comparison to the international standard. Having analysed the database for each subject as described in earlier sections, the following conclusions are drawn as the main results.

Wide differences in achievement persist among students, schools, districts across all locations, area and regions.

A wide difference is found in the achievement levels across students and schools in Nepal. Some students in the study were not found able to answer even single test item (0%) while the best students achieved more than 95%. The average achievement of students in the lowest performing schools was below 15% while the students from the highest performing schools gained, on average, over 90%. When knowing that the average score in a school is below 15%, it means that in those schools many students fall below that level. From an equal opportunities viewpoint, it is not a good sign that 'low' and 'high' performing students are concentrated in certain schools particularly high number of low performing students are concentrated in community schools.

Not only is the variation great amongst students and schools, but it is also significant across the 28 districts covered by the study. In the lowest performing districts the average performance was below 30% (vary by subjects), while in the highest performing districts it ranges from 50% to 70%. The students in the Kathmandu Valley schools exceed the average students in numerous indicators in across the regions other than Kathmandu Valley. The achievement level of the students in Science, for example, from the Eastern (32%), Far-Western (32%), and Mid-Western (34%) development regions is far behind that of the students in the Kathmandu Valley (56%). Once again, from an equal opportunities standpoint, this is not a positive sign.

Further analysis of the highest and lowest performing community schools reveals that students tend to perform higher who afford more time on homework, have positive attitude towards the subjects, receive required support from siblings or private tuition from teachers, do not need to work for earning while studying, and reach the grades at their correct age and so on. On the other extreme, lowest achievers are those who belong to illiterate parents, especially the mother and involved in agriculture work, need to work

for earning, receive neither private tuition nor support from family members, and the children not receiving textbook and so on.

Gender gap in achievement is still visible in some subjects.

The data base for each subject analysed in the earlier sections also reveals that achievement gap between boys and girls is still persist in Mathematics and Science. In Nepali, both the boys and girls are found to have been performing at the same level (48%). However, girls still lag behind the boys by 5 percent in Math (33% for girls against 38% for boys) and by 4 percent in science (39% for girls against 43% for boys) achieving even lower than the national average in both the subjects which is 35% and 41% for Mathematics and science respectively.

Remarkable gap in achievement is noticed between the students from institutional and community schools.

Students' average achievement in institutional schools is found remarkably higher (for example, 31% gap in Mathematics) than in the community schools, which is more than twice as high as the private schools (57%) when compared with the students at the community schools (26%). One of the reasons behind this may be the higher socio-economic status of the family sending children to institutional schools as higher socioeconomic status and educational achievement are positively correlated.

It is noteworthy that there are also few number of community schools where the average results are at the same level as in the private schools even though the SES is remarkably lower. In these schools, either the processes are more effective than in the private schools or the students are of the same ability as those in the private schools and are not adversely affected by the processes within the school or their socioeconomic status. It could be concluded that higher achievement of community schools is not due to the system but because of the individual effort of school, teacher or students themselves.

Students are found good at tasks requiring lower level of cognitive ability like knowledge and comprehension but are poor in tasks requiring higher level of cognitive ability like analysis, evaluation and applying the gained knowledge in new situation.

Students are found comparatively poorer in the ability to solve problems, to analyse, deduce logic, generalize, justify an argument or viewpoint, and in the ability to transfer learning from one context to another. A remarkably high number of students were able to solve only 15% or less of the practical problems (15% of the students in Mathematics, 11% in Nepali, and 5% in Science).

In Mathematics, students are able to do basic calculations, but are weak in reasoning, problem solving, proving theory or formula, and in constructing shapes and figures. In many cases, the students did not even attempt to complete the open-ended questions of higher cognitive level.

In Nepali subject, students performed well in the tasks requiring recognizing the correct answer, recalling simple facts from the texts, fundamental thinking, and the basic interpretation of paragraphs. However, they are much weaker in producing fluent texts or essays, and in preparing synthesis and abstracts from a text. The students tended to attempt open-ended tasks but the skills were not high enough for the highest marks. In case of applying the gained knowledge in new situation, they are found very poor than the higher level thinking ability.

Dataset in Science also reveals the fact that the students are good in recognizing the correct answer and in very fundamental knowledge, such as choosing the facts and numbers, and writing the definitions. They are much weaker in reasoning, problem solving, proving the principle, and constructing the figures. In many cases, the students did not even start to do the open-ended questions and hence, the low score. From the international comparison viewpoint, in general and by each content area in particular, of Nepalese students is found lower (mean = – 0.18) than the international average (mean = 0.00).

Caste/ethnicity and home language background have played significant role in determining the students' achievement.

The achievement level of the Dalit students is found lower in all the subjects. Particularly, the low-achievement is found among Dalit girls in Mathematics (23%), and Science (33%), as compared with the Dalit boys (25% in mathematics and 37% in Science). However, it is worth noting that achievement level for Dalit in the Kathmandu Valley is much higher than the average of the country.

Similarly, the achievement of Madhesi students is also found lower than the average in Mathematics (29%), Nepali (37%) and Science (36%). Among them, the girls lag further behind against the boys (24% against 34% in Maths, 31% against 40% in Science and 34% against 39% in Nepali) across all ecological zones.

Rural urban differences in student achievement are remarkably noticed across all subjects.

Urban students outperform the rural students in all subjects. For instance, urban students achieved 57% in Nepali whereas it is only 45% for rural. Similarly, achievement gap between urban and rural is more wider in Mathematics by 24% in which students from urban located schools achieved 52%. Similar level of differences can also be found in Science in which students from urban school exceed far ahead (51%) than the rural students (37%).

Learning achievement over the years has remained stagnant in all subjects rather improving.

While compared the results for 2013 with that 2011, the achievement is not found to have improved over these years. For instances, in 2011 it was 49% for Nepali, whereas it is found only 48% going further down by around 1% in 2013. In Mathematics, it is found more worsening as the achievement score has gone further below by almost 7% that came down from 43% in 2011 to 35% in 2013. The reasons behind such downfall in the learning achievement demand further enquiry to get the root of the fact.

The unavailability of textbooks to students even up to end of academic session is found prevalent which has hampered the students' progress.

Despite the concerted efforts to make the timely available of textbooks in the hands of students, the datasets reveal a distressing fact that a significant number of students (5.1% in Mathematics, 2.7% in Nepali and 2.8% in Science) are found studying the subjects without receiving text books of respective subjects even up to the end of academic session. This problem has resulted in lowering the achievement from minimum 8% to maximum 15% score for the respective subjects. For instance, the students reporting to have received textbook on time are found to have scored 36% in Mathematics, 49% in Nepali and 42% in Science whereas those not receiving textbook even at by the end of academic session are found to have scored 24%, 35% and 34% respectively in the same.

School bullying has been impeding learning potential for a significant number of students.

Though bullying at school goes unnoticed to many of the parents and school teachers and hence they remain reluctant to such incidences students experienced in school, it keeps on worsening learning environment negatively impacting the learning potential of students and thus is taken as negative activities in school. The obtained datasets reveal that bullying is found rampant in Nepalese school, which has negatively been affecting students learning potentials for minimum 9% in Mathematics to the maximum 19% students in Science. Those students experiencing no bullying are found to have scored 35% in Mathematics, 5o% in Nepali and 42% in science, whereas students encountering extreme types of bullying are found to have scored only 26% in Mathematics, 29% in Nepali and Science. Excluding the number of students experiencing minor and moderate types of bullying, an alarming number of students (2.2 % In Mathematics and Nepali and 3.4% in Science) are found experiencing severe bullying.

Majority of schools and teachers are found reluctant to assign and check homework that has been preventing students from boosting their achievement.

Properly assigned and checked homework which is an integral part of teaching learning process boost the learning achievement as it provides an ample opportunity for self-learning and engages students in problem solving. However, the data sets of student assessment evidently shows that nearly 37% in Mathematics, 35% in Nepali and 32% students are found never assigned or checked their home works. The students getting homework with its checking are found to have score 36% in Mathematics, 50% in Nepali and 42% in Science, whereas students with no homework and getting no checked score just 27%, 34% and 30% respectively in the subjects. It further shows that regular assignment of homework along with its checking has contributed to raise the achievement up to 9% points in Mathematics, 16 % points in Nepali and 12% points in Science.

6.4 Implication of this Assessment

The low level of learning along compounded with a wide gap in achievement among and between rural-urban, community-institutional schools, caste/ethnicity, social groups across all developmental regions, ecological districts are neither a good sign for equality and nor the positive indication for the system. Similarly, achievement variations of students among various subjects, content areas within the subjects, and poor achievement in the domain of application and higher ability show some lacking in educational delivery system and process. Given the context, the following would be the main implications for the system to improve the achievement level of Nepali students. The implications mentioned below indicates the areas of interventions with some alternatives, but it does not specify the roles of each agency and stakeholder. In this case, based on these implications each agency and stakeholder are expected to identify the tasks required to perform in order to improve the results.

Reducing inequality in achievement

As the result confirms a wide inequality in achievement level, which continues to persisting, between students from rural and urban locations, among various language speaking and ethnic/cates groups, and across districts, regions and ecological belts; it has been an imperative for the policy makers, curriculum planers, teacher educators as well as education managers to look for the ways for enhancing the capacity of current delivery system to produce equal level of learning opportunities for all children irrespective of caste/ethnicity, social and language groups, family in which one is grown up and the school types on attends. Although a lot of efforts have been put into the system to reduce the inequality, the persisting gap

is still demanding further measures that actually reduce the inequality in practice. In addition, reason behind the low performance of students from community schools and the students of particular communities should be explored with micro level studies going deep down into the root causes. Simultaneously, discussions need to initiate with the teachers of low performing schools, parents of the low achieving students on reducing the inequality and improving low level of achievement that is persisting.

The causes of persisting inequality in achievement between the students from community and institutional schools would be another area to further explore whether it is due to the teacher effect or more rigorous teaching learning practice at institutional school, or socio-economic background of the family that provides students attending institutional schools additional supports. If the rigorous teaching learning practice and caring environment available at private school is found only the contributing factor to raise achievement, then educational managers and policy makers need to find out the ways to encourage community school teachers for providing caring environment and rigorous teaching learning practices. One of the measures to raise and reducing achievement gap would be introducing the performance based incentives to the schools and teachers for raising achievement and reducing achievement gap of the low performing students through benchmarking and setting the target to achieve in the given timeframe against the target given.

Improving reading ability

As confirmed by the dataset of 2011 and 2012, reading proficiency of Nepali students is seen poor, which has not improved over the years. The low level of reading proficiency among the students has kept them in a weak state of comprehending implied meaning, solving complex problem, abstracting of deeper ideas, producing open-ended text not only in Nepali subjects but also in other subjects like Mathematics and Science leaving them less able to perform the tasks demanding higher cognitive ability. Such state of reading proficiency necessitates developing new instructional design for classroom practice with more reading activities, exercises on comprehension, introducing varieties of text with different genres by spelling out specifically the standards on fluency, accuracy, performance and problems solving in curriculum and textbooks across the subjects. Another measure for fostering reading ability would be determining certain amount of texts to be read by students for each grade in addition to the text from the textbooks. The curriculum planners, textbook writers, teacher educators as well as the classroom teachers need to sensitized, oriented and trained in such instructional design. Allocating additional time for reading activity in school hours through curricula provision would also be one option for which the CDC needs to initiate discussions and dialogues with curriculum planners, textbook writers as well as with teachers.

Similarly, ongoing teacher education courses and packages also need to be redesigned incorporating required skills and competencies for teacher to organize reading skill promoting activities in classroom.

In addition, existing student assessment and examination practice needs to be revisited incorporating the reading tasks in assessment activities across all subjects. Generally, existing student assessment and examination practice tend to either ignore or give less emphasis to assess skills on listening, speaking, oral presentation, reading fluency and accuracy. So, teacher are also likely to be reluctant on such skills. Student assessment framework and guidelines also need to be revised incorporating tasks and activities to assess the skills.

Fostering higher cognitive ability

Although cognitive skills are seldom taught explicitly in schools, various researches indicate that schooling through teaching knowledge and skills in language, mathematics and science need to promote cognitive ability. In this regard, fundamental goal of education is to equip students to think critically, solve complex problems and succeed in the society and economy of 21st century (Fin et al, ND). The subject wise datasets reveals that Nepalese students are good at lower level of cognitive skills such as knowledge and recognizing, but found poor across all subjects in higher level ability such as application, analysis, synthesis and evaluation required for solving the novel problems, information processing and applying the knowledge and skill learned one context into the new one. Provided situation leads to imply that either our teaching learning activity deviates from organizing activity to engage in students in the tasks requiring higher ability or reading materials including textbooks lack adequate exercises or practice that foster given ability.

Our education system needs to be well aware of the reasons behind why schooling is not promoting desired higher cognitive skills among eighth graders even after completing eight years school education. Curriculum planners and textbook writers need to pay enough attention to address the issue of low performance level in higher ability tasks and initiate discussion on how to design new curricula and textbooks so as to keep engaged students in tasks demanding higher mental ability. Similarly, teacher educators and training modules designers also need to look for the possible measures to incorporate skills and competencies required for teachers in organizing classroom activity to foster higher ability among students. Teacher and test item writers too are to be reoriented on developing assessment tools to assess these skills. Future researches and studies need to concentrate their focus on this issue to find out the lacking part whether it is because of curricula or classroom practices or teacher preparation.

Improving student assessment system

Modern learning theory conceives student assessment for learning rather than assessment of learning, which implies that assessment should focus on both assessing the processes of learning i.e. inquiring, independent learning, use of generic skills, reflections and the products of learning e.g. knowledge/concepts, problem solving capabilities adopting different varieties of methods such as oral test for oral communication, discussion for collaboration, presentation/performance for creativity, tests and examination for knowledge etc. However, the datasets show that a notable number of students are found to have been not used to answering open ended, problem solving and application types of subjective items. Given circumstances lead one to conclude that our assessment practice and test items have not engaged students in solving novel problems, producing creative works, dealing with open ended tasks which demands an immediate reform to make it able to assess both the process and products. For this to happen, whole assessment mechanism and practice from classrooms at the lower unit to public level examinations at the national level need to overhaul reform.

To bring reform in the classroom assessment practice, teachers require desired capacity to develop standardized test item against learning objectives set in the curricula. Schools should be reoriented on the tools and process of assessing students learning process and products. Furthermore, existing ceiling of 32 percent marks to pass the grade and level wise examination should also be raised at least to 50 percent to raise the expectation and study habits among the students. At the same time, public examination at district and national level need to ensure that students are assessed by means of standardized test items, which corresponds the curriculum objectives set for to go through.

Rooting out the incidences of school bullying

The phenomenon school bullying, as recognized in other parts of the world, is not so much familiar to parents and teacher so has gone unnoticed at schools; but it is found rampant in Nepalese schools in some forms negatively affecting their learning potentials for those children who experience. As results of this study confirm that it is found to have been associated with low level of learning for a notable percentage of school children (from minimum 0.8% experiencing with all forms to maximum 23% experiencing with 20% of bullying). Given situation alarms concerned teachers, education managers and even to the parents to be aware of the phenomenon and to expedite the possible preventive measures that help schools at least minimizing such incidences.

One of the measures for this would be making teachers aware of the issue and its possible consequences to the school children. Students themselves are to be sensitized with the phenomenon and its effects on their colleagues through mobilizing child clubs that are functioning at schools. Child right activists also would be supportive to lead the process. Similarly, child friendly school framework being implemented under the initiative of the DOE would also need to incorporate possible indicators regarding the measures that discourage school bullying. Curricula and packages on teacher training also need to consider these issues seriously in their future revision or repackaging.

Raising parents' educational level

Low level of parents education is found to have been directly associated to the underachievement of their children at school which further has been the main sources of perpetuating social inequality and disparities. Despite the lots of effort put and investments made for raising literacy level of parents, youth and adult literacy rate of nation (84.72 in total with only 80.16 for women among youth population and 59.6 in total with only 48.8 for women among adult population respectively (CBS, 2013)) has not been improved to the desirable extent which also has resulted in the low level of learning of those children whose parents are illiterate.

The results of the study confirm that parents' education level in general and the mother in particular is found major determinant for their children achievement level at the school. So, raising parents literary level especially targeting to mothers is seen an urgent need not only for achieving higher level of students achievement at school but also for achieving social equality and a just society. One of the probable measure for raising parents literacy level would be devising a policy and program that encourage schools to make those illiterate parents literate residing within the catchment area and whose children are attending the particular school. For this strictly enforcement of 'Student Mobilization Guidelines for Literacy 2070' developed the CDC and endorsed by the MOE would an effective instrument. Another measure would be devising and introducing, as implemented in some parts of the world (e.g. Mauritania, in some Latin American countries), the family literacy program through schools. Such provision would help raise literacy level of the parents and improve achievement level of their children on the one and develop strong school community relations that ultimately increases parental involvement in school development endeavours.

Raising learning level equally to all content areas across the subjects

Besides problems of low achievement level, Nepalese students are not found to have developed similar levels of knowledge, skills competencies equally over all content areas within and across subjects. Little learning in one area of subject impedes to acquire expected level of proficiencies in other areas leading ultimately towards underachievement in all content areas within and across all subjects. As shown by the results, reading and writing in Nepali; biology in science; algebra, geometry and sets in Mathematics are poorly learned areas in comparison to others.

Given the context of such unequal level of learning, curriculum planners of the respective subjects first need to give serious consideration over the little learned contents in order to find the answer to the questions as to why students are not able to learn them and how curriculum planning and designing in given contents would be restructured in order to enable learning. Similarly, root sources of low level of learning in the given areas and innovative ways to facilitate students learning in those contents also need to be explored with further researches and studies in the days to come. Existing teacher training courses and packages also require revisiting in order to sensitize and prepare teachers for further facilitating students learning in the areas identified.

Ensuring timely availability of textbooks to all students

Despite the government efforts for the delivery of textbook on time, the data reveals that a notable number of students are compelled to complete the grade without textbooks. In this regard, educational managers from district to central level need to look for the further ways to strengthen the existing delivery mechanism and make school accountable for ensuring on time availability of textbooks to all students. One of the measures towards this would be enforcing the provision to earmark certain amount fund by each school every year or devising a mandatory provision to establish a book corner with certain set of textbooks in each. Similarly, designing multi-year usable textbooks could also be one of the measures for textbooks distribution. In this context, the existing mechanism of textbooks development, printing and distribution should be reviewed based on earlier studies or with some additional studies to ensure timely availability of textbooks with good quality. One of the options for ensure timely availability of textbooks with good quality is to implement multiple textbooks system. One of the crucial factors regarding the use of textbooks is the enhance teachers' capacity to use multiple resources including printed, electronic and online material in order to deliver effectively the curricular competencies.

Implications for the further study

This study, besides calculating mean score of each subject, has compared achievement in various content areas as well as various level of cognitive domain within the subject. Similarly, study has identified variations in achievement based on various diversity factors, including district, region, ecological zone, ethnicity/caste, home language, socio-economic status, age, gender. Variations are found in almost all variables, though variations are not the same as in some factor variation is small, but others have wide. This study, however, indicated that there is a need of some changes in pedagogical practices at school as well as improvement in overall system of school education delivery, but it does not identify the specific causes beyond the poor result and wide variations in results. In this context, in order to identify specific causes to improve the situation, issue or case specific analysis or studies (qualitative/quantitative) are important. Instead of designing large-scale study, such studies mostly medium or small scale, focusing on specific issues of particular district, resource centre, school, community, or in some cases practisers' action research. Regarding large scale and survey-based national assessment of student achievement instead of norm-based assessment criterion-based assessment could be useful to spell out competency level of students. Third, it is imperative to carry out assessment and evaluation of various policies, programs and projects as well as assessment of schools and technical and management related agencies under the ministry of education against their roles, responsibilities and mandates, and redefining their additional or new roles, responsibilities and mandates.

Having considered the low achievement compounded with wider inequality between rural and urban schools, community and institutional schools, Dalits and non-dalits, Nepali and non-Nepali speaking community which has been persisting over the long period, the concerned agencies require to initiate formulating time bound action plan to improve the performance of delivery system in education. The action plan needs to concentrate more specifically on devising appropriate strategies to increase teachers' accountability towards the results. Most importantly, there is also a need to initiate dialogues and discussions with academia, curriculum planners and policy makers on how to raise the performance of the Nepalese students to the international level in order to make them competitive in the globalized world.

6.5 Conclusion

The main objective of this assessment was to find out whether the students of grade eight achieve the goals set by the national curricula in Mathematics, Nepali language and Science by determining their national level of achievement. More specific, this assessment was carried out to determine variations in student

achievement among various diversity factors, such as ecological zones, developmental regions, districts, rural/urban location of schools, community/institutional schools, ethnicity and language groups, gender, socio-economic status and home background. Similarly, other specific objectives of this assessment were to compare students' learning achievement in the current study with the previous studies and create reliable baseline data for the future study. In order to achieve these objectives, the National Assessment of Student Achievement (NASA) 2013 was carried out for grade eight students in three subjects, namely Mathematics, Nepali language and Science. For Mathematics and Nepali language subjects, it was the second cycle in which the first cycle was carried out in 2011, and for Science subject it was the first cycle.

The study was conducted taking 1199 randomly selected schools in 28 sample districts covering all ecological zones and development regions. Total number of student sampled for the assessment was 48,000, almost equally distributed in each of the three subjects. Each student participated in the assessment responded a set of background questionnaires together with one set of test item among three subjects Mathematics, Nepali language and Science. In addition, 1199 teachers and 1199 head teachers from the sample schools were also participated in this assessment by responding related questionnaires.

Standardized test items were used to assess the students' achievement. Individual items were developed through the workshops of trained subject teachers and subject experts. The subject committee of each subject revised the items and selected at least six different items to measure the same performance in each subject. Selected items were pretested and results were analysed by calculating difficulty level of each items. As a result of pre-testing of the items, the difficulty levels of the tests were set around 50–60%. Some linking items from NASA 2011 and international tests like, TIMSS and PISA were selected and used. Items analyse and equating of scores are done using IRT modelling.

Assessment results are analysed in chapters 3, 4 and 5 of this report. In those chapters, various indicators are analysed and interpreted. Subject wise achievements are analysed mainly in relation to the type of school, content area, various levels of cognitive domain, item type, school location, language, ecological zone, development region, gender, ethnicity/caste, language, attitudes of students towards subject, parents' educational status and occupation, home possessions and home accessories. The subject wise achievements are compared with different international assessments as well as the previous assessment carried out in Nepal.

High consideration has been given in the process of assessment, particularly, selecting and using the method for sampling, deciding the sample sizes, preparing items, pretesting, final test administration, data analysis

and report writing, to make the assessment valid, reliable and comparable. This process of standardization in this assessment fully utilised the requirements of international standards for the large-scale assessment of student achievement.

The findings of NASA 2013 indicates that there are some lacking in the current educational system of Nepal. The wide gap in the achievement level of students, schools, districts and regions indicates that there is an unequal distribution of educational opportunities to students. Different actors can play vital role to maximize the equity in their respective organizations and their relevant field. However, some of the required changes are quite fundamental.

The information mentioned in this report provides numerous insights into the variables connected to the results. These information can also be used as a basis for better planning as well as policy alteration. During this assessment a rich dataset with individual level relevant information obtained, which provide the opportunity to search the barriers for better performance. This report mentioned the comparable results to the previous NASA 2011 and the international standards as it followed a standardised process for the assessment of students' achievement. Every stakeholder in education sector could get feedback from the results and the concentration on quality improvement would increase and they can play supportive role in monitoring teaching learning activities. In this way, expected improvement could be made in the education system.

References

Adams, R. & Wu, M. (2002). PISA 2000 Technical Report. OECD.

APA (2007). *Report of the APA Task Force on Socioeconomic status*. Washington, DC: American Psychological Association.

Bartlett, J. E. II, Kortlik, J. W. & Higgins, C. C. (2001). Organizational Research: Determining Appropriate Sample Size in Survey Research. *Information Technology, Learning, and Performance Journal*, **19**(1), 43–50.

Béguin, A. (2000). *Robustness of Equating High-Stake Tests*. Enschede: Febodruk B.V.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: the classification of educational goals*; *Handbook I: Cognitive Domain*. New York, Longmans, Green.

BPEP (1997). The Effect of New Curriculum on the Achievement of Grade 4 Student. Kathmandu: HMG/Ministry of Education, BPEP.

BPEP (1998). The Effect of New Curriculum on the Achievement of Grade 5 Student. Kathmandu: HMG/Ministry of Education, BPEP.

Bradley, R. H. & Corwyn, R. F. (2002). Socioeconomic status and Child Development. *Annual Review of Psychology*, **53**, 371–399.

Brooks-Gunn, J. & Duncan, G. J. (1997). The effects of poverty on children. *Future Child*, **7**(2), 55–71.

Bryk, A. S. & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin*, **101**(1), 147–158.

CBS (2012). *Final result of population and housing census 2011*. Kathmandu: Central Bureau of Statistics, The Government of Nepal.

CERID (1999). Assessment of Learning Achievement of Lower Secondary Children: Grades 6 and 8. Kathmandu: CERID in Association with MOES/Nepal.

CERSOD (2001). *District level assessment, 2001*. Kathmandu: Centre for Educational Research and Social Development, GON/MOE/DOE.

Cochran, W. G (1977). *Sampling Techniques* (3rd ed.), New York: Wiley & Sons.

Cohen, J. (1988). Statistical power analysis for the behavioral Sciences. (2nd Ed.) Hillsdale, NJ: Erlbaum.

DeVellis, R. F. (1991). Scale Development: theory and applications (Applied Social Research Methods Series, Vol. 26). Newbury Park: Sage Publications.

EDSC (1997). *National Assessment of Grade 3 Students*. Kathmandu: Educational and Developmental Service Centre.

EDSC (1999). *National Assessment of Grade 5 Students*. Kathmandu: Educational and Developmental Service Centre.

EDSC (2001). *National Assessment of Grade 3 Students*. Kathmandu: Educational and Developmental Service Centre.

EDSC (2003). *National Assessment of Grade 5 Students*. Kathmandu: Educational and Developmental Service Centre.

EDSC (2008). *National Assessment of Grade 8 Students*. Kathmandu: Educational and Developmental Service Centre

Fennema, E. & Sherman, J. A. (1976). Fennema-Sherman Mathematics attitude scales: Instruments designed to measure attitudes toward the learning of Mathematics by females and males. *Journal for Research in Mathematics Education*, **7**(5), 324—326.

FNBE (2004). *National Core Curriculum for Basic Education 2004*. The Finnish National Board of Education. Helsinki: Yliopistopaino. Retrieved from

http://www.oph.fi/english/sources_of_information/core_curricula_and_qualification_requirements/basic_education.

Fullbright (2008). A *final report on national assessment of grade-V students*, Fulbright Consultancy & Cooperative Hands in restoration (CHIRAG). Kathmandu: Government of Nepal, Ministry of Education and Sports, Department of Education.

Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika*, 73, 43–56.

Hair, J. F. Jr. , Anderson, R. E., Tatham, R. L., & , W. C. (1998). *Multivariate Data Analysis*, (5th Ed.). Upper Saddle River, NJ: Prentice Hall.

House, J. D. & Telese, J. A. (2008). Relationships between Student and Instructional Factors and Algebra Achievement of the students in the United States and Japan: An Analysis of TIMSS 2003 Data. *Educational Research and Evaluation,* **14**(1), 101—112.

Kadijevich, Dj. (2006). Developing trustworthy TIMSS background measures: A case study on Mathematics attitude. *The Teaching of Mathematics,* **9**(2), 41—51. Retrieved from http://elib.mi.sanu.ac.yu/journals/tm/17/tm924.pdf.

Kadijevich, Dj. (2006). Developing trustworthy TIMSS background measures: A case study on Mathematics attitude. *The Teaching of Mathematics*, **9**(2), 41—51. Retrieved from http://elib.mi.sanu.ac.yu/journals/tm/17/tm924.pdf.

Kadijevich, Dj. (2008). *TIMSS 2003: Relating Dimensions of Mathematics Attitude to Mathematics Achievement*. Retrieved from http://www.ipisr.org.rs/Dokumenta/zbornici%20PDF/Zbornik%202008%2040-2/Zbornik-08-2-006.pdf.

Kadijevich, Dj. (2008). *TIMSS 2003: Relating Dimensions of Mathematics Attitude to Mathematics Achievement*. Retrieved from http://www.ipisr.org.rs/Dokumenta/zbornici%20PDF/Zbornik%202008%2040-2/Zbornik-08-2-006.pdf.

Knapp, T. R. & Brown, J. K. (1995). Ten Measurement Commandments That Often Should Be Broken. *Research in Nursing & Health*, **18**, 465–469.

Krieger, N., Williams, D. R & Moss, H. W. (1997). Measuring social class in US public health research: concepts, methodologies, and guidelines. *Annu. Rev. Public Health* **18**, 341–78.

Liberatos, P., Link, B. G., Kelsey, J. I. (1988). The measurement of social class in epidemiology. *Epidemiol. Rev*. **10**, 87–121.

Lord, F.M. (1952). The relationship of the reliability of multiple-choice test to the distribution of item difficulties. *Psychometrika,* **18**, 181–194.

McLoyd, V. C. (1997). The impact of poverty and low socioeconomic status on the socioemotional functioning of African-American children and adolescents: mediating effects. In R. D. Taylor & M. **Wang** (Eds), S*ocial and Emotional Adjustment and Family Relations in Ethnic Minority Families*. Mahwah, NJ: Erlbaum. (pp. 7–34).

Mehrens, W. A. & Lehmann, I. J. (1991). *Measurement and Evaluation in Education and Psychology*. 4th edition. Harcourt Brace College Publishers, Fort Worth.

Metfessel, N., Michael, W. B., & Kirsner, D. A. (1969). Instrumentation of Bloom's and Krathwohl's taxonomies for the Writing of behavioral Objectives. *Psychology in the Schools*. **6**, 227–231

Metsämuuronen J (2012a). **Challenges of the Fennema-Sherman Test in the International Comparisons. *International Journal of Psychological Studies,* 4**(3) September, 1–22. http://www.ccsenet.org/journal/index.php/ijps/article/view/16904/12480.

Metsämuuronen J (2012b). Comparison of Mental Structures of Eighth-Graders in Different Countries on the basis of Fennema-Sherman test**. Available at *International Journal of Psychological Studies,* 4**(4) December, 1–17. **Available at** http://www.ccsenet.org/journal/index.php/ijps/article/view/17305/14331.

Metsämuuronen, J. (2009a). *Methods assisting the assessment*. Oppimistulosten arviointi 1/2009. Opetushallitus. Helsinki: Yliopistopaino. [In Finnish].

Metsämuuronen, J. (2013a). A new method to setting standard for the wide range of language proficiency levels. *International Education Research* 1(1). Available at http://www.todayScience.org/IER/v1/Jari%20Metsamuuronen.pdf.

NASA (2011). *Where are we now? Student achievement in Mathematics, Nepli and Social Studies in 2011.* Sanothimi: Education Review Office.

NASA (2012). *Where are the grade 3 and 5 now? Student achievement in Mathematics, Nepli and Social Studies in 2011.* Sanothimi: Education Review Office.

Nunnally, J. C. (1978). *Psychometric theory*. (2nd ed.) New York: McGraw-Hill.

Oaks M. (2011). *Measuring Socioeconomic status*. In e-Source. Behavioral & Social Science Research. Retrieved from http://www.esourceresearch.org/Portals/0/Uploads/Documents/Public/Oakes_FullChapter.pdf.

OECD. (2009). Take the Test Sample Questions from OECD's PISA Assessments.

Papanastasiou, C. (2000). Effects of Attitudes and Beliefs on Mathematics Achievement. *Studies in Educational Evaluation,* **26**(1), 27—42.

Papanastasiou, E. (2002). Factors That Differentiate Mathematics Students in Cyprus, Hong Kong, and the USA. Educational Research and Evaluation: *An International Journal on Theory and Practice*, **8**(1), 129—146.

PEDP (1998). *Primary Education Development Project: An assessment*. 1998. Kathmandu.

PISA (2003). Education at a Glance. OECD Indicators 2003. Paris: OECD;

PISA (2003b). *PISA School Questionnaire*. Retrieved from http://pisa2003.acer.edu.au/downloads.php.

PISA (2006a). *PISA Student Questionnaire*. Retrieved from http://pisa2006.acer.edu.au/downloads.php.

PISA (2006b). *PISA School Questionnaire*. Retrieved from http://pisa2006.acer.edu.au/downloads.php.

PISA (2006c). *PISA Released Items – Reading*. OECD

PISA (2009). Take the Test. Sample Questions from OECD's PISA Assessments. OECD.

Shen, C. & Tam, H. P. (2008). The Paradoxical Relationship between Student Achievement and Self-Perception: A Cross-National Analysis Based on Three Waves of TIMSS Data. *Educational Research and Evaluation*, **14**(1), 87—100.

Shen, C. (2002). Revisiting the relationship between students' achievement and their self-perceptions: A cross-national analysis based on the TIMSS 1999 data. *Assessment in Education,* **9**(2), 161—184.

Stevenson, H. W. (1998). A Study of Cultures: Germany, Japan and the United States - An Overview of the TIMSS Case Study Project. *Phi Delta Kappan*, **79**(7), 524—529.

TIMSS (2003). *TIMSS Contextual Background Questionnaires*. Retrieved from http://timss.bc.edu/timss2003i/context.html.

TIMSS (2006). *Questionnaires*. Retrieved from http://nces.ed.gov/timss/questionnaire.asp

TIMSS (2007). TIMSS *2003 Science Items. Released Set, Eight grade*. IEA, TIMSS & Pirls International Study Center. Lynch School of Education, Boston College.

TIMSS (2009a). *TIMSS 2007 User Guide for the international Database. Released Items, Mathematics – Eight grade*. IEA, TIMSS & Pirls International Study Center. Lynch School of Education, Boston College.

TIMSS (2009b). *TIMSS 2007 User Guide for the international Database. Released Items, Science – Eight grade*. IEA, TIMSS & Pirls International Study Center. Lynch School of Education, Boston College.

TIMSS (2011). Mullis, I.V.S., Martin, M.O., Foy, P., & Arora, A. (2012). *TIMSS 2011 International Results in Mathematics*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Verhelst, N.D., Glas, C.A.W., Verstralen, H.H.F.M. (1995). *One-Parameter Logistic Model OPLM*. Cito, Arnhem.

White, K. R. (1982). The relation between socioeconomic status and academic achievement. *Psychol. Bull*., **91**, 461–81.

Wilkins, J. L. M. (2004). Mathematics and Science Self-Efficacy: An international investigation. *The Journal of Experimental Education*, **72**(4), 331—346.