National Assessment of Student Achievement (NASA) 2012

(Grade 3 and 5)

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

Education Review Office

National Assessment of Student Achievement 2012

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

Webpage: www.ero.gov.np, email: eronasa@gmail.com
Tel: +977-1-6639556, +977-1-6639557, Fax: +977-1-6639556

National Assessment of Student Achievement 2012

(Grade 3 and 5)

Government of Nepal

Ministry of Education

Education Review Office

Sanothimi, Bhaktapur

Nepal

National Assessment of Student Achievement 2012 (Grade 3 and 5)

Publisher: Education Review Office, Sanothimi, Bhaktapur, Nepal.

This publication can be cited as:

Education Review Office. (2015). *National Assessment of Student Achievement 2012 (Grade 3 and 5).* Sanothimi: Author.

This report is also available at : www.ero.gov.np

Published in: 2015 (2072 BS)

Copies : 500

Printed At : Morang Offset Press
Kathmandu

Phone : 01-4372431

Mobile : 9841344706

Acknowledgements

This is the report of National Assessment of Student Achievement for grade 3 and grade 5 conducted by Education Review Office (ERO) in 2012. Though the first draft of this report was completed in the year 2013 right after the test administration and data analysis, we now have been able to publish it in this form after further verifying and editing the contents as well as language,and reshaping the initial draft of the report with rewriting some parts of the reports, particularly conclusion, implications and executive summary.

This report is, in many ways, a collaborative effort of several institutions and persons working in different sectors who were involved in accomplishing the tasks at various phases. Many school teachers and subject specialists were involved in test development and finalization without which test administration would not have been possible, therefore, they deserve special thanks for their contribution in this phase. Particular thanks go to the EDSC led by Dr. Tirtha Raj Khaniya for bearing the responsibility of tasks such as printing, transportation and delivery of the test papers and questionnaires to the districts; administering the test; examining the answer sheets and accomplishing the preliminary data entry works. Special appreciation also goes to the personnel from the respective DEOs for the role they played in coordinating and monitoring while accomplishing the entire process of test administration. Thanks also go to the students of respective grades for taking part in the test and to teachers including the head teachers of respective schools for their cooperation in responding the relevant questionnaires that helped in generating required information for this report.

Special thanks go to the ERO personnel for their roles during the process of item writing and finalizing the test items, conducting pre-test, developing the final items, monitoring and coordinating the entire test administration process across the country and accomplishing the data analysis and preparing this report. Among the personnel Devi Ram Acharya, Shyam Prasad Acharya and Sunita Shakya particularly deserve special thanks for their commendable works during the process of item development and finalization by analysing their difficulty level, equating and standardizing them. Similarly, they also devoted their time in tabulating, analysing, interpreting the data including the preparation of the first draft of this report. Without their tireless effort during the process, the report could not come into this shape. Due thanks also go to Prakash Kumar Kharel for his role in providing management and technical support to facilitate the entire process.

Since this report owes much from Dr Bhoj Raj Kafle, the then Under Secretary, and Jari Metsämuuronen, the Finnish consultant who took a leading role in planning, management and overall guidance in test design, administration and report writing; they deserve a heartfelt thanks for commendable works that they accomplished. Similarly, the then ERO chiefs Dr. Kamaleshwor Kumar Sinha, Jeevan Sharma Poudel, Khaga Raj Baral and Dr. Hari Prasad Lamsal equally deserve sincere thanks for providing overall guidance and support in leading the entire process without which the process would not be accomplished.

We thank Dr. Vishnu Karki for his contribution of providing some inputs on the initial draft of the report.

At the final stage, Dr. Lekha Nath Poudel and Gopal Prasad Bhattarai, worked in editing the contents of the entire report, re-writing the executive summary and conclusions including reshaping the structure and chapters. Sushil Babu Khanal co-ordinated the whole process of editing and finalising the report in this form. Similarly, the support and involvement Kul Prasad Khanal during the process is important in accomplishing the task. Dr. Binod Luitel edited the language of the entire report. This report would not have been possible to bring in this shape in time without their rigorous efforts and initiatives, therefore our sincere thanks go to each of the contributors.

Finally, special gratitude and thanks are extended to the Steering Committee headed by the Secretary for MOE in taking policy decisions promptly to facilitate the assessment process. Similarly, due thanks also go to the Technical Committee for its support in the process of assessment program design and selection of the subject to assess. Equally the Subject Committees also deserve much thanks for their support provided not only in maintaining validity and reliability of the items but also in suggesting the final approval of the test.

This will be incomplete without extending thanks to many other names from the ERO not mentioned above who have directly or indirectly been involved in accomplishing the process and provided their support and assistance to bring the report in this shape.

Constructive suggestions and feedbacks from academia, professionals and practitioners are always expected for further improvement of the products in the days to come.

December, 2015

Ministry of Education

Education Review Office

Sanothimi, Bhaktapur

Table of Contents

List of Tables

List of Figures

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
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
GLM	General Linear Modelling
IRT	Item Response Theory
Max	Maximum
Min	Minimum
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 Assessment
SD	Standard Deviation
SE	Standard Error
SES	Socio-economic status
SPSS	Statistical Package/Programme for Social Science
TIMSS	Trends in International Mathematics and Scie

Executive summary

It is generally accepted that people having literacy and numeracy skills not only able to read, write and calculate, but also think analytically and plan their activities accordingly. They have further chances of better economic opportunities, higher agricultural productivity, healthier children, and better reproductive health in comparison to those who have not acquired such skills. Moreover, the literacy and numeracy skills also form the basis for future learning. Schooling has to prepare youngsters for living better life by equipping them with these knowledge and skills needed to meet the challenges to be faced in adult life. Those knowledge and skills learnt at schools by our youngsters largely indicate the prosperity and future development of a country. Therefore, policy makers, educators, parents and society at large seek to have reliable information on to what extent their children are learning at school, whether they will be able to meet future challenge and whether there are gaps in their learning that need to be addressed. They have reason to be confident that their children are acquiring the knowledge and skills that they will need to function as family members as well as economically active persons. With the motif of seeking reliable information on students' learning at schools in terms of given knowledge and skills, assessing students' achievement by means of large-scale standardized test at various levels has evolved in the front of education. Nepal is no exception to adopt large-scale achievement test since the mid of the 1990s to determine what students have learned as a result of their educational experiences.

The practice of assessing student achievement by means of standardized test in Nepal started since 1995 that continued up to 2008 assessing achievement for grade 3 in 1995, 1997, 2001; for grade 4 in 1997; for grade 5 in 1998, 1999, 2003, 2008; for 6 and 8 in 1999 and for grade 8 in 2008 (ERO, 2013, Chap 1). However, large-scale assessment of student achievement in standardized form following Item Response Theory (IRT) was started since the establishment of Education Review Office (ERO) under the Ministry of Education (MOE). ERO conducted National Assessment of Student Achievement (NASA) for the first time in 2011 following IRT modelling to assess the learning outcomes of 8th graders in Mathematics, Science and Social Study. Next to it, this (NASA 2012) was another large-scale assessment designed to assess learning achievement in Mathematics and Nepali for 3rd graders and in English, Mathematics and Nepali for 5th graders.

Objectives of the Assessment

Like other national assessments, this study aims to generate objective, accurate, and comparative information on learning achievement of primary level students in schools being motivated to evaluate the education system so as to support in obtaining results as expected by the curricula. More specifically, the main objectives of this assessment were to determine the learning level of grade 3 and grade 5 students in English, Mathematics and Nepali against the curricula goals, to create a reliable database on the learning level in those subjects for benchmarking in order to monitor the progress over time and to generate recommendations for policy making to improve educational quality and equity.

Methodology

At the beginning of 2013, the test was conducted to assess learning achievements of students in Mathematics and Nepali for grade 3 and English, Mathematics and Nepali for grade 5. Altogether 80,232 students (38,753 in grade three and 41,479 in grade 5 from randomly stratified 1,690 sampled schools) participated in the assessment. In the sample for grade three,17,256 students were boys and 17,166 were girls. Similarly, out of the total sample for grade five, 19,617 students were boys and 19,783 were girls. Out of the 75 districts of Nepal, the dataset represents a random selection of 28 districts covering all five Development regions (Eastern, Central, Western, Mid-Western, Far-Western) and the Kathmandu Valley, as well as all Ecological Zones (Mountain, Hill, Tarai). In addition, both rural and urban schools as well as community and institutional schools are proportionally represented so that the results of the assessment can credibly be extended to the whole students and school population of Nepal.

Three versions of the items in each subject were administered and the final scores were equated by utilizing the IRT modelling. Reliability of the tests was found high and the validity was assured by applying specification grids of the national curriculum developed by the Curriculum Development Centre (CDC). From methodological standpoint, the process and practices of the inquiry has successfully followed the procedures as used in some international level test with some contextualization on them to reflect the reality of the Nepali context. Thus, this test is believed to fulfil the national and international ethical principles, criteria and standards to qualify it as credible assessment. The results were linked to the set of results from the 2008 assessments as well as to the international item banks of Trends in Mathematics and Science Study (TIMSS) and Progress in International Reading Literacy Study (PIRLS).

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 selected subjects for each grade. Thus, the students in a grade were required to participate in one of the selected subjects assigned to the school. The answer sheets were marked and achievement scores were tabulated using Optical Mark Reading (OMR) machine.

The results are reported mainly as percentages of maximum marks where 100 (%) represents all tasks solved and 0 (%) none. As a result of pre-testing of the items, the difficulty levels of the tests were set at 50—60%. For Nepali language assessment, the Common European Framework of Reference for Languages (CEFR) was also used to obtain the level of students from language achievement point of view.

Main Findings

This test has not only assessed the subject specific learning achievements and their variance by each of the cognitive domains but also has analysed the results in disaggregated form in terms of gender, types and location of schools, Ecological as well as Development regions, and the like. Similarly, it has further examined the results associating with various family and school related factors considering the extent they influenced the achievements. The key results of the assessment are summarized under the following points.

- **Clearly divided student population into three distinct groups**

Dataset reveals that there are two to three population groups among the students: low and high performing community school students, and mostly high-performing students from the institutional schools in both grades. The population in grade 5 Mathematics is closest to normal whereas it varies a lot in grade 3 with more students in low performing range. Similar scenario can be seen in English as well where majority of population is shifted towards low performing groups. The population in Nepali in both grades is also found not distributed normally along with a notable portion of low performing students. Though a large population from institutional schools lies in higher performing groups, there are also lower-performing students in grade 5. However, the high percentage of student population of low-performing students in community schools indicates that the system is not able to give sufficient support for those students who are lagging behind in the early grades. Because of the low performing large population from community schools, the main system is shifted to the lower performing level since the main population comes from the community schools.

- **Unbalanced learning across the curricula contents**

Against the expectation of curricula, the dataset is evident that certain contents of the curricula are learnt less effectively than others. For instances, in mathematics, the achievement level in algebra and numeracy is remarkably lower than arithmetic and geometry. In Nepali and English, reading and writing skills are poorer in comparison to the achievement in vocabulary. Circumscribed with such unbalanced learning in some domains, the entire system is shifted towards a low performing–making it less effective to yield better results.

- **Low capability to solve tasks requiring higher ability**

Dataset shows that the students are performing well in recognizing the correct answer and in recalling the learnt facts. They are weaker in productive type of items. In many cases, the students did not even start to answer the open ended questions and, hence, the lower score. Within all the datasets of grade 5, about 20% students are not able to solve any of the tasks requiring higher ability. Similarly in grade 3, a notable number of students (4%) were not able to solve any of the tasks requiring the ability of applying the knowledge in a novel situation. Students in institutional schools are found to have been more able to solve practical and complex problems than their peers in community schools. The same phenomenon was found also in grade 8 (ERO, 2013) which has continued from the lower grades to the end of school. For one reason or another, students in the institutional schools are seen to be relatively more able to solve complex problems than their peers in the community schools. Given context implies that existing classroom practice is seen inadequate to nurture problem solving, comparing things, analysing phenomenon, and synthesizing results from several sources to raise the achievement standard in higher ability.

- **Wider disparity in achievement between the districts and Development regions**

To look at the results from equality point of view, inequality persists between the districts and regions across the country–revealing wide differences between the districts to give equal opportunities in reaching the pre-set educational goals. The average achievement in the Kathmandu Valley is very high compared with the lowest performing districts in the sample. Differences in the mean scores between the lowest and highest scoring districts vary from 33 percent (in grade 5 Math and in grade 3 Nepali) to 43 percent (in grade 5 Nepali) and ultimately to 51 percent points in English. In English, the difference is found to be connected with the proportion of institutional schools in the district, as the medium of language for instruction in most of the institutional schools is English which has helped raise the achievement. In other subjects too, the results is seen to be the higher when there are more institutional schools. Some districts showed very high achievement without any institutional schools in the sample. Another reality, more crucial, is that, in the lowest performing districts, the average achievement is absolutely very low in the districts like Bardiya, Rolpa, Jumla, and Udayapur in Mathematics; Saptari, Achham, and Mahottari in Nepali; and Khotang, Jumla, Saptari, and Mahottari in English. These low-performing districts concentrate in two Development regions: Eastern (three) and Mid-Western (three); being the lowest achieving regions in each dataset. Among the Development regions, the Kathmandu Valley is at 21–29 percent ahead of the rest, prevailing a wider disparity among them.

- **Remarkable variations between institutional and community schools**

To compare the achievement level in terms of type of schools, a remarkable variation is seen in their performance. Students in the institutional schools outperform the students in the community schools, depending on the dataset, at the widest 20–36 percent. The number of private schools is, however, too low to have a remarkable effect on the national average. Because of the English being the medium of instruction in all grades and special thrust given to it from the early grades, the difference between the institutional schools and community schools in English is the widest at 36 percent, which is at 20 -28 in other subjects due to the rigorous teaching and special care given to the students.

- **Moderate but growing differences between the ecological belts and rural/urban schools**

Dataset shows that the students from the Mountain region are slightly better in learning achievement than the students in other regions. The lowest achievement is found in Tarai but the difference is not wide in comparison with the students from Hill when the Valley students are excluded. The Valley students outperform the other regions. When it comes to the school location, the urban community schools outperform the rural community schools by 6–12 percent; excluding the Valley, the difference is 0–6 percent. For reason or another, there is not a wide difference between the rural and urban institutional schools which is not so remarkable but the trend is serious. In comparison with the previous results, it is seen that the students from Tarai are performing lower. In Mountain, they are found performing higher than around 15 years ago. The urban schools have raised remarkably their position in comparison with the rural schools. If these trends continue in the future, it will lead to a wider inequality in society between rural and urban areas as well as between the Ecological zones. Ultimately, this will lead to an uncontrolled urbanization if the families continue to send their children to big cities to study and move later themselves to seek a better life.

- **Wider inequality in performance level between the different language groups**

The Nepali speaking majority includes all the segments of the society and, hence, their result is more or less at the average in the datasets. The datasets strongly indicates that the students from Tamang and Magar speaking groups perform the highest in all subject areas while the students from Gurung, Tharu, Limbu, and Sherpa speaking communities lag far behind the others. The result shows that the difference in mean scores between highest and lowest performing language groups is 22–53 percent depending on the subject; the largest differences are in English subject. In Mathematics,Tharu and Gurung speaking; in Nepali, Newari and Tharu speaking; and in English, Sherpa and Gurung speaking students are low performers than the others.

- **Association of poor socio-economic status with remarkably lower learning achievement**

Socio-economic Status (SES) and its components are found to have been strongly associated with the learning achievement in Nepal. The difference between students from the lowest and highest SES groups is 23–40 percent. The widest difference is found in English subject at 40 percent whereas it is at 23-30% in Mathematics and at 30-31% in Nepali. Depending on the subject, 11–26% of the students are at the lowest level of SES meeting none of seven indicators.

Results show that especially low achievement is common among the children whose parents are illiterate. As reported, 34–42% of the students have illiterate mothers and 15–19% of the students have illiterate fathers. Similarly, the lowest result is also common among families where either mother or father or both are engaged in the agricultural occupations. According to the datasets, 53–66% mothers and 35–41% fathers work in the agriculture or are involved only in household chores.

When children have very few home possessions or none of the home accessories, the achievement level is remarkably lower than the national average. According to the datasets, 2–7% of the students did not have any of the eleven home possessions including table, dictionary, peaceful place for exposition and the like and 26–45% students possess neither TV, mobile phone nor computer at home. If the issue of parents' low educational level is solved, the result in the low-SES group is likely to improve.

The same result was also found in grade 8 (ERO, 2013); and hence it indicates that the issue is a structural problem in Nepal. Especially structural problem is the high mothers' illiteracy rate and their low educational level. Right now, practically all the girls are attending school and their children will have better changes in education after 15–25 years. However, without targeted intervention from the MOE, within the next 15–25 years, a remarkable portion of students will not have an opportunity to have a mother with the reading skills at least at basic level.

- **Prolonged hours of involvement in work impeding children's learning**

The results in all subjects show that students, either working for a paid job or spending more than two hours per day for unpaid household chores beyond the school time, are found to have caused lower achievement level of the students. The dataset shows that 27–36% students work for the paid job and 15–23% of them spend more than 2 hours in household chores. Though most of the low-graders do not usually work many hours per day for the paid job, their volume is too much. The pertinent question is: Why do they need or are willing to work daily? Most probably they need to earn for pocket money or for subsistence livelihood. Whether the need for working for the paid job or need to participate more than 2 hours in the household chores is only one part of a complex knot of problems involved with the low SES affecting the low learning results. Though the child labor is prohibited by the law, something more is also required to prevent school children from working for a paid job. It is seen that, in community schools,involvement in some household chores up to 2 hours a day has not lowered children's achievement level.

- **Association of over aged schooling with lower achievement**

The highest performance is found with the students studying at their proper age with the peer in normal age group, that is, at the age of 8–10 years in grade 3 and 10–12 years in grade 5. The achievement lowers down as the age increases, or it is lower than expected. For instance, the mean achievement ranges from 61 to 64% in Nepali and 56 to 61% in Mathematics for the students studying at their proper ages, whereas it lowers down to 57 in Nepali and to 55% in Mathematics for students who are over aged which further lowers down to 56 in Nepali and to 52% in Mathematics for grade 3. More or less similar level of differences is observed in other subjects and grade. According to the dataset, 25–30% of the students are over or under the proper age for the grade. The same phenomenon was also observed in grade 8 (see ERO, 2013) indicating that delayed schooling or non-systematic entrance in schools is a structural problem for the educational system.

- **Effects of lack of textbooks in achievement**

In all datasets, the achievement level of students lacking textbook at a minimum is significantly lower than those who have access to the textbook. For instance, the students in grade 3 having textbook have achieved 65 in Nepali and 61 in Mathematics, whereas those who lack books at a minimum achieve just 54 in those subjects. Similar level of difference is seen in the subjects of grade 5 too. The dataset shows that 4–6% of the grade 3 students lack the textbook in Mathematics. The phenomenon was seen also in grade 8 (see ERO, 2013) which tells that failure in the delivery of textbooks to all areas in the country is a structural problem in the educational system, remaining a cause of lower achievement.

- **Negative effect of bullying and unfair treatment from teachers on achievement**

Two indicators which should have zero values are the student behaviour indicating the frequency of bullying in school and teacher behaviour indicating the sense of unfairness of the teachers on students. Reported data shows that consistently 54–56% of the students have encountered some kind of bullying and 5–10% of students are experiencing a severe kind of bullying in school within a month. The latter is more frequent in grade 3 (9–10%) than in grade 5 (5–6%). The figures from grade 8 (ERO, 2013) also shows that the incidences of bullying ranged from 42–46% (maximum) to 2–3% (minimum). It seems that the higher the grade, the lower, though existent and remarkable, is the bullying rate. The dataset also shows that 8–14% students feel that their teachers do not treat them fairly and the figures are lower at grade 3 (8–10%) than at grade 5 (12–14%). At grade 8 it was 19–21%, showing that the higher the grade the more is the unfairness from teachers reported by students.

To describe the relation between the types of bullying students faced and their achievement, the data evidently reveals the fact that bullying in any form has been affecting the achievement. For instance, the mean achievement in mathematics ranges from minimum 59.2 to 76% who encountered no bullying at all whereas it is just 42% for those who experienced all types of bullying in grade 3. Such notable differences are found in other grades and subjects too. Similarly, when the students feel that the actions of the teachers and the schools are ultimately good, the Mathematics results in grade 3 are better than average (59% in community school and 77% in institutional schools). At the other extreme, in feeling ultimately negative of such actions, the results are far below the average (41% in community schools and 58% in institutional schools). Similar is the situation in other subjects and in grade 5 too.

- **No remarkable difference in students' achievement because of gender and ethnicity**

The data shows that the differences between the ethnicity and genders are very small, or non-existent. The differences in achievement are practically non-existent between the genders. There are differences across ethnic groups but they are not significant in general. However, still Dalit, Madhesi, and Janjati students perform lower than Brahmin and Chhetri students even at the lowest grades. Hence, there are still lots more to do in reducing the gap between the ethnicities/castes when it comes to learning achievement. Given context implies for teachers and schools to pay special attention for the ethnic/caste issue not to widen the gap among various ethnicities.

- **Noticeable changes in leaning outcomes over the past decades**

The changes in learning achievement/test results have been remarkable in some areas while in others nothing has happened. For example, in Nepali and English datasets of grade 5, there is not much difference in comparison with the previous datasets between genders whereas in Mathematics datasets the gap between the genders has reduced, which is a positive sign. On the other hand, the English and Nepali datasets hint that the students in the urban schools have gained remarkably higher over the years. This may have been caused by urbanization, concentration of educated families, and due to the influence of private and boarding schools. It is also seen that the students in the Mountain zone have gained remarkably higher compared to the Tarai students. When it comes to Development regions, the Eastern region is seen to have lowered further down while the Far-Western has made remarkable progress both in ranking and in absolute terms. For the Far-Western region the change is positive, but for the Eastern region it is naturally not a good sign.

- **Low level performance of Nepalese students in comparison to international average standard**

When mathematics dataset is compared with the international standard, Nepali students are, on average, one year behind the international average and the 5th graders are somehow at the level of grade 4 students. In Nepali and English, the average reading proficiency of grade 5 students is much lower than the international average of grade 4 in PIRLS standards. The datasets in Nepali hint that reading proficiency is lower than one year behind the international level and the estimated level of grade 4 students is round 1.5 standard units lower than the international mean. In Nepali and English, the grade 5 students are far below even the grade 4 international average (-1.4 standard units in Nepali and -1.2 in English). The same kind of result was found also in grade 8 (ERO, 2013). However, within the dataset, there are several very highly performing students but their number in total is so low that they do not raise the national standard. Given context raises the question: "What would be the fastest and most feasible way to increase the national achievement level in Nepal?" One lesson which could be learnt from Finland of high results and low variability between the schools (see e.g., Schleicher, 2006; Metsämuuronen, Kuosa, & Laukkanen, 2013) is the strong emphasis on support for the students at the early grades.

Summing Up

Having analysed the datasets, first it is noticed that lower graders are seen higher performers than the upper graders. For instance, average achievements in Nepali and Mathematics are 63% and 60% respectively for grade 3, whereas they are 60% and 53% respectively for grade 5. Second, wide variance in average achievement can be observed among the districts from the lowest 48% in grade 3 and 37% in grade 5 to the highest 80 to 81% respectively in Nepali. Similar level of difference is also seen in Mathematics achievement which varies from 79% in grade 3 and 71% in grade 5 (the highest) to 40% in grade 3 and 37% in grade 5 (the lowest). Similarly, the level of variance in English is found at 47% points from the lowest to the highest. With regard to the Development regions, there are also wide differences in the ability of the students. The students from Mid-Western and Eastern regions are at relatively low ability level whereas those from the Western and Central regions perform better. The Valley is at the top of all. In Nepali, the differences between the Development regions are remarkable. In the Valley, which is incomparable with other zones, students outperform the students in all other regions. Eastern region is recorded as the lowest. Less or no gender gap is noted among regions. Reducing such an inequality between the lowest and highest performing students is seen challenging.

Despite the equal weightage given to all the three subjects in teaching, students' performance level varies a lot across the subjects within and between the grades. Particularly, wider difference at 7% lower in Mathematics than Nepali achievement in grade 5 is not justifiable. Similarly, the achievement variance between the types of schools and their location is also not good sign for the system.

Chapter 1: Introduction

Government of Nepal established Education Review Office (ERO) in 2010 as an integral part of the reform in quality of education 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 and quality of education so that equity and quality 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 publicize the assessment and audit reports. In this context, ERO for the first time conducted National Assessment of Student Achievement (NASA) for grade eight students in 2011. This assessment (NASA 2012), conducted among grades 3 and 5 students, was administered in March 2013.

This introductory chapter briefly presents the historical development on 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 2012. Similarly, it includes the objectives of this assessment as well as a brief indication of methods and process applied for this assessment. Finally, it briefly presents the structure of this report.

1.1 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. Such national 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 curricular 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 of and equity in the education system. The assessment of grade 8 (NASA 2011) was the first large-scale assessment 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 2012) is the second large-scale assessment targeted to grade 3 students in two subjects: Mathematics and Nepali language and to grade 5 students in three subjects: Mathematics, Nepali and English. 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 altogether 48,000 students covering 16,000 students in each subject as the sample size. When the tests were administered, the actual number of students attending the tests changed slightly which was confined to 44067 students.

1.2 Student Assessment as a Process

National Assessment of Student Achievement (NASA) in Nepal includes several nested cycles. The wider and more general cycle concerns the process from the administrative viewpoint: from the assignment from the MOE to the releasing of the final report (fig. 1.1). Within the general framework, there is another process of preparing the measurement instruments and still another process of analysing and interpreting the results.

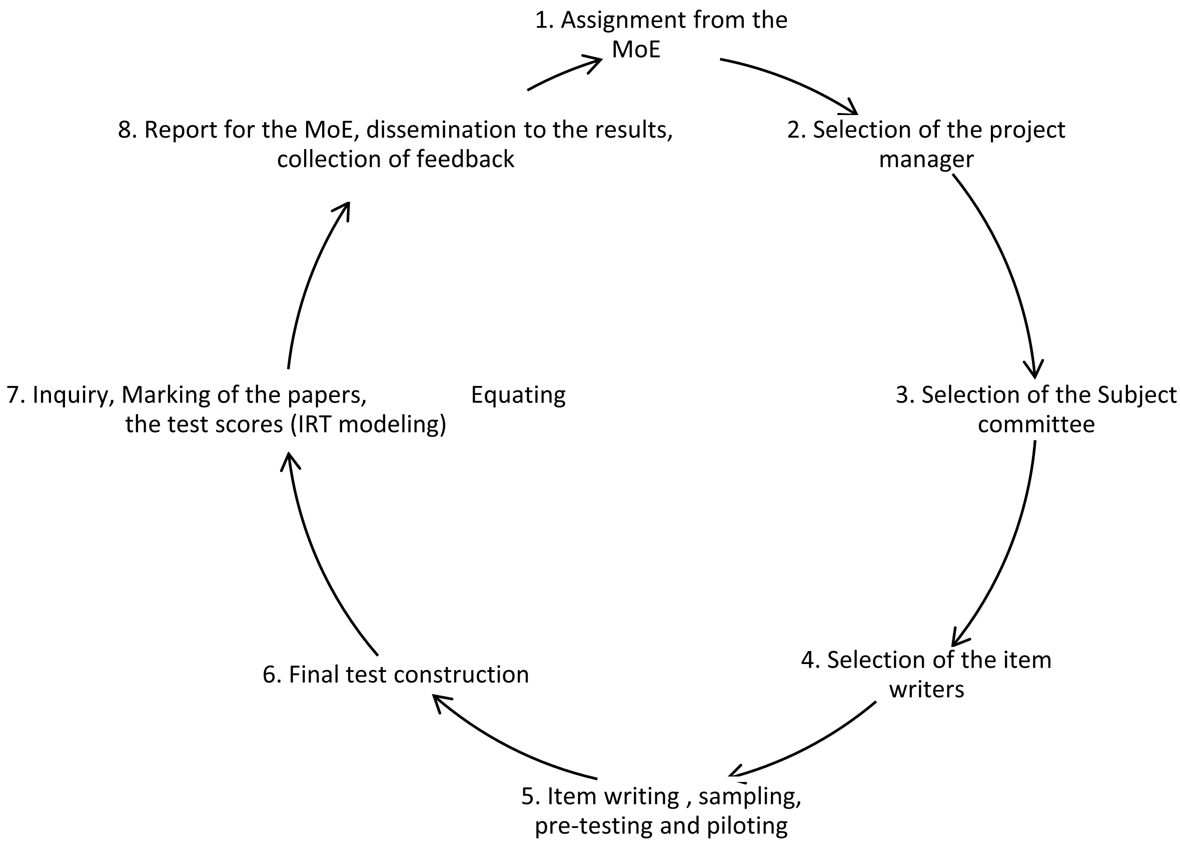

Figure 1.1 Administrative cycle of NASA 2012

The Steering Committee formed in the MOE has ratified the need to continuing student assessment programme for several years. According to the plan, the student achievement is assessed every second year at grade 8 and the next year at grade 3 and 5 at least until 2016. NASA 2012 in grade 3 and 5 is the first large-scale assessment for these grades. The subjects assessed in this assessment are Mathematics and Nepali for grade 3 and 5, Nepali, Mathematics and English for grade 5.

A number of experienced classroom teachers from schools and universities, and resource persons from the Kathmandu Valley worked for item writing. The task was to create a sufficient amount of items representing different sub-topics, various difficulty levels and cognitive levels in four versions for the pre-test and ultimately three versions for the final test to each subject and grade. The items were pre-tested at two layers: first, it was attempted to find how much the language used in the tests affect the grade 3 and 5 students' performance and second, items were pre-tested in 240 schools in several districts to find the stable item parameters for the final test.

The final versions of test were compiled using the following six principles: (1) Content's dependence on the curriculum (construct validity), (2) Content's coverage to be as wide as possible (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 results with the international results of Trends in International Mathematics and Science Study (TIMSS) and Progress in International Reading Literacy Study (PIRLS). All the tests are highly discriminatory as the reliability of the total score is higher than $\alpha = 0.93$ for all versions and subjects.

The final tests were administered in 1690 schools for each grade from 28 districts by the respective DEOs, representing various Ecological zones and developmental regions, rural and urban areas as well as community and institutional schools. Marking of students work and data entry were done centrally by outsourcing the task to a professional consulting company. Data tabulation and analysis including the equating of the test scores using the IRT modelling was done in-house at the ERO. The preliminary draft of this report was ready in the end of the year 2013. The first results were disseminated to the MOE personnel prior to sharing to the Joint Consultative Meeting (JCM) between the government and development partners on SSRP held in the beginning of December 2013. However, publication of report was delayed as report was not finalised until June 2015. From the third week of July 2015, the process of finalizing the report began and it took about two months to finalise its report, as the draft was preliminary. The tasks accomplished during the report finalisation were preparation of updated version of draft report, checking the data and editing the contents, formatting and finally, language editing and proof reading.

There were roughly 8 phases in the assessment process of students' achievement: 1) Pre-phase, 2) Preparation phase, 3) Final testing phase, 4) Data analysis phase, 5) Post-work phase, 6) Report writing phase, 7) School-wise reporting phase, and 8) Item banking phase (fig. 1.2).

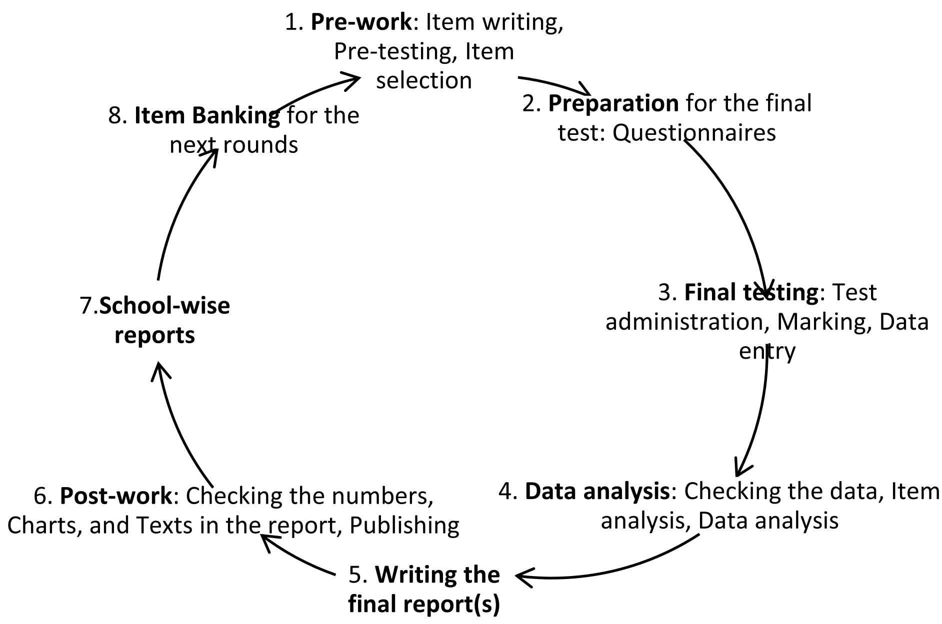

Figure 1.2. Phases of the assessment process of student achievement

The pre-phase includes item writing, reviewing, pre-testing, item analysis and selection for the final tests. At the preparation phase, the background questionnaires were prepared, relevant stakeholders like DEO and schools were informed about the process of assessment, and the test administrators were oriented on test administration process. At the final testing phase, the tests were administered at the same time in different parts of the country. The data analysis phase started with the screening of the datasets to find out and correct the possible errors in the data. At this phase, the final item analysis was done omitting a few items having very poor response rate. Finally, the data analysis produced the necessary information for report-writing. The report-writing phase took several weeks, seeking comments in several rounds. Before the report was made ready for printing, at the post-work phase, double checks were done for the data as well as interpretation of the results. At the school reporting phase, the schools involved in the assessment got feedback on their performance. At the final phase – before the next assessment round – the items and their parameters were banked for the later use.

1.3 Characteristics of NASA 2012

ERO (2013) has discussed the general characteristics of NASA – especially in connection with NASA 2011. 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 were the same as in NASA 2011. However, there are some differences between the process of NASA 2011 and 2012. One of the differences is that the NASA 2012 was administered and handled practically without international support from the very beginning of the data-analysis phase. The international consultant came once when the dataset was received back from the outsourced consultancy. The international consultant supported in test equating, data-analysis and reporting phase. Hence, the process itself showed that the in-house team of the ERO is capable of doing such tasks on its own.

1.4 Objectives of the Study

The main objective NASA 2012 was to find out whether the students in grade 3 and 5 have reached the goals set in the national curricula. The specific objectives of NASA 2012 were as follows:

i. To determine the current national level of achievement of grade 3 students in Nepali and Mathematics, and grade 5 students in Mathematics, Nepali and English;
ii. To determine variations in student achievement between different Ecological zones, Development regions, districts, school location (rural/urban), school type (community/ institutional schools), ethnicity and language groups, gender and socio-economic conditions;
iii. To examine the extent to which the home background and other pupil-related factors influence learning achievement;
iv. To compare student learning achievement in the current study with that of the previous studies;
v. To compare the student learning achievement in Nepal with that of international studies – PIRLS (Reading) and TIMSS (Mathematics).
vi. To create reliable baseline data for monitoring the progress over the period of time.

1.5 Method and Process Used in Assessment

This assessment was carried out using 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 prepared 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, the 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 the difficulty level. Because 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 to compare NASA 2011 results with some international assessments like TIMSS and PIRLS. Besides, a set of background questionnaires was also used among the students in order to identify the variables that influence the achievement of the students. Questionnaires were asked for teachers and head teachers about classroom and school management.

In this assessment, out of the total school population, 1690 school for each grade were taken as the sample from 28 sample districts and the number of students in the sample was 44067 in two subjects in grade 3 and three subjects in grade 5, 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. Development 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 score of each item was tabulated in computer 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.

1.6 Structure of the Report

This report is organized in six chapters. Chapter 1 introduces NASA 2012. This chapter mainly consists of a brief background, characteristics and objectives of NASA 2012. The second chapter presents the description of the methodology applied in NASA 2012. This chapter gives some details of the methods that were applied in NASA 2012 including sample selection and determination of size of the sample; item writing and selection procedures; reliability and validity; test administration, marking of answer papers and data entry; 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. Chapter 3 includes the analysis of results of Mathematics of both grades 3 and 5; chapter 4 includes the analysis of results of Nepali subject of both grades 3 and 5. Chapter 5 presents the analysis of assessment results of English in grade 5. Each chapter of this report discusses basic results, diversity factors and achievement, and selected explanatory factors and achievement. The final chapter concludes the report by summarising NASA 2012 process and major findings, presenting implications and conclusions of the study.

Chapter 2: Methodology

The National Assessment of Student Achievement (NASA) for grade 3 in Mathematics and Nepali, and grade 5 in Mathematics, Nepali and English, conducted in the year 2012 (hereafter, NASA 2012), is the second large-scale national level assessment administered in Nepal. Some of the key tasks of the assessment process were outsourced and the rest were carried out by the Education Review Office (ERO) own self. Tools development including test item preparation and questionnaire development, sampling and data analysis were carried out by ERO with the support of an international expert, whereas test administration, marking and data entry were outsourced to a consulting company. The assessment was conducted in 28 sample districts selected using stratified random selection method covering each of the defined strata, with altogether 1688 schools for grade three and 1690 schools for grade five which were selected proportionally from the sample districts. The total number of schools participated in assessment was 1704.[1] Standardized tests were administered for the assessment of student achievement in each sample school. Equating and calibration of tests items were done using Item Response Theory (IRT). The results are presented by analysing students' score using various statistical tools.

This chapter describes the methodology used in NASA 2012. The methodology includes- sampling, item writing and test construction, development of background questionnaires, test administration, scoring and data entry, equating of the test scores over three versions used in the final testing, and the statistical methods used in analysis. Section 2.1 deals with the sampling, section 2.2 with the item writing and test construction, and section 2.3 with the development of background questionnaires. Similarly, section 2.4 describes the variables used in the analysis, section 2.5 presents the process of equating the test scores over three versions used in the final testing, section 2.6 deals with the principles of criterion based assessment used in the assessment of Nepali and English proficiency, and section 2.7 describes the statistical methods used in analysing the data.

2.1 Sampling

Stratified random sampling method was used for sampling the districts and schools. Sample schools were selected randomly from the identified strata. This section describes various strata identified for sampling, sample size and the process of selection of sample schools.

2.1.1 Strata

The main interest of National Assessment was not the assessment of individual student but the system itself. Therefore, the basic unit for sampling was the school. The sample schools were selected in such a way that they should represent the country as widely as possible,

---

[1] It was intended to select 1700 schools, selecting both grades from the same schools, but in some schools have only one grade 3 or 5 was running and additional schools were selected for the remaining grades. As a result, the total number of schools increased slightly to 1704.

and the selected students should represent the whole student population as widely as possible. The practical sampling strategy in this context is therefore the proportional stratified sampling with random selection. For the purpose of sampling, the following strata were considered:

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

Kathmandu Valley was taken as a separate geographical stratum as 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 the mixed ethnicities, distinct weather condition, wide economic and development activities as well as the dense human capacity make the Valley a unique fourth geographical region in the analysis. Hence, there were 16 basic strata in the sampling. School size, ethnic group and language were also considered while selecting sample, but not as strata.

2.1.2 Sample size

The samples of 28 districts were randomly selected to represent each of the 16 geographical strata (see figure 2.1). The selected districts, number of schools and the students are presented in tables 2.1 and 2.2. Selection of 28 districts among 75 districts with a representative number of schools from each stratum can reasonably have a good coverage of the Development regions as well as Ecological zones.

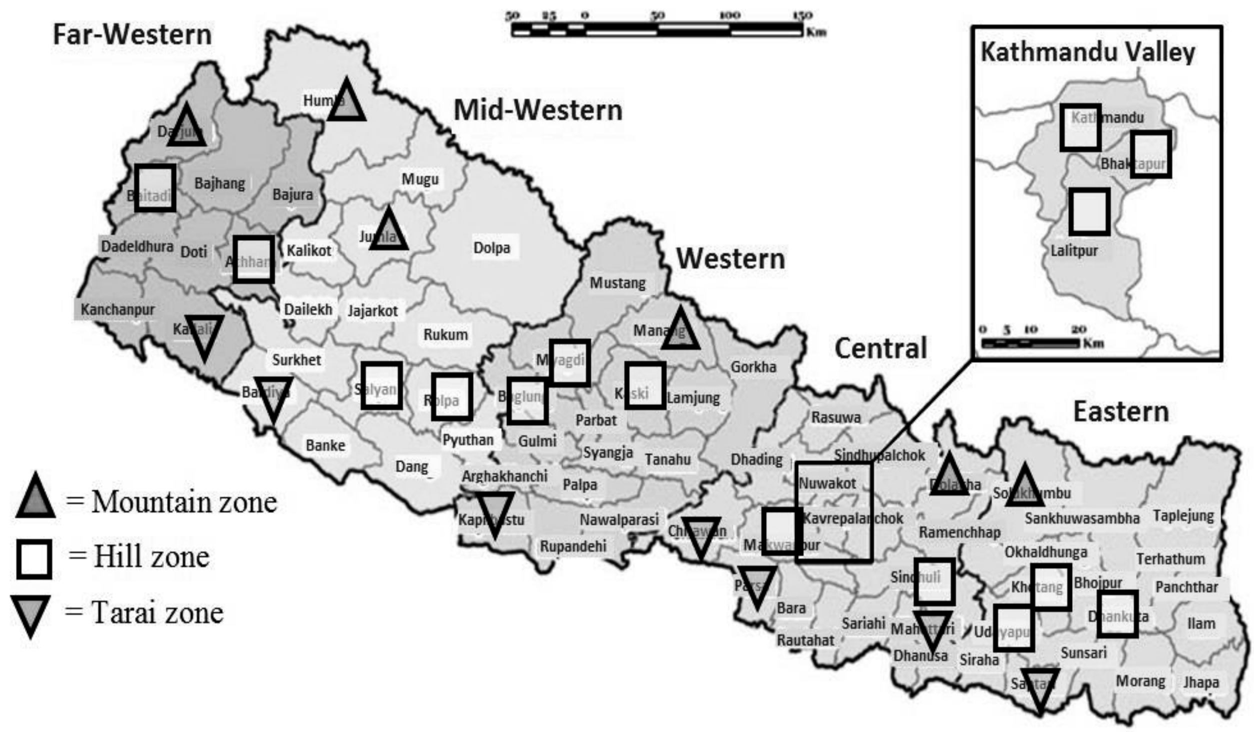

Figure 2.1 Sample districts of NASA 2012

Table 2.1 Number of schools in each sample districts in NASA 2012 for Grade 3

District	Developmental region	Ecological zone	No of Schools		No of the Students	
			Math	Nepali	Math	Nepali
Dhankuta	Eastern	Hill	22	23	407	521
Khotang	Eastern	Hill	29	31	581	659
Saptari	Eastern	Tarai	32	31	832	720
Solukhumbu	Eastern	Mountain	18	19	397	375
Udayapur	Eastern	Hill	32	33	693	768
Bhaktapur	Valley	Valley	23	23	558	544
Chitwan	Central	Tarai	36	35	831	842
Dolakha	Central	Mountain	31	28	708	587
Kathmandu	Valley	Valley	90	90	2042	2110
Lalitpur	Valley	Valley	32	31	733	721
Mahottari	Central	Tarai	27	26	660	664
Makwanpur	Central	Hill	40	40	963	979
Parsa	Central	Tarai	25	26	617	638
Sindhuli	Central	Hill	37	37	824	858
Baglung	Western	Hill	32	35	732	790
Kapilbastu	Western	Tarai	34	33	854	817
Kaski	Western	Hill	41	43	1001	974
Manang	Western	Mountain	6	4	19	16
Myagdi	Western	Hill	17	18	376	426
Bardiya	Mid-Western	Tarai	23	22	563	548
Humla	Mid-Western	Mountain	8	9	198	214
Jumla	Mid-Western	Mountain	10	10	184	190
Rolpa	Mid-Western	Hill	28	28	697	699
Salyan	Mid-Western	Hill	31	31	649	677
Achham	Far-Western	Hill	32	33	703	725
Baitadi	Far-Western	Hill	36	36	742	784
Darchula	Far-Western	Mountain	22	25	495	535
Kailali	Far-Western	Tarai	49	47	1193	1120
Total			**843**	**847**	**19252**	**19501**

Table 2.2 Number of schools in each sample districts in NASA 2012 for Grade 5

District	Developmental region	Ecological zone	No of Schools			No of the Students		
			Math	Nepali	English	Math	Nepali	English
Dhankuta	Eastern	Hill	15	15	15	374	387	348
Khotang	Eastern	Hill	21	21	21	508	504	457
Saptari	Eastern	Tarai	19	21	21	453	527	526
Solukhumbu	Eastern	Mountain	12	13	12	288	294	254
Udayapur	Eastern	Hill	22	22	21	523	495	521
Bhaktapur	Valley	Valley	15	16	15	374	408	376
Chitwan	Central	Tarai	23	24	24	590	606	622
Dolakha	Central	Mountain	19	19	18	448	417	421
Kathmandu	Valley	Valley	60	60	60	1551	1558	1546
Lalitpur	Valley	Valley	21	21	21	502	531	495
Mahottari	Central	Tarai	18	18	18	449	465	453
Makwanpur	Central	Hill	25	25	27	635	603	671
Parsa	Central	Tarai	17	17	17	448	453	422
Sindhuli	Central	Hill	25	25	25	591	562	561
Baglung	Western	Hill	22	22	23	574	564	614
Kapilbastu	Western	Tarai	22	24	21	549	599	531
Kaski	Western	Hill	29	29	27	722	711	670
Manang	Western	Mountain	4	4	5	18	9	12
Myagdi	Western	Hill	11	12	12	271	302	296
Bardiya	Mid-Western	Tarai	15	15	14	375	361	348
Humla	Mid-Western	Mountain	5	6	6	127	156	153
Jumla	Mid-Western	Mountain	7	7	6	157	159	130
Rolpa	Mid-Western	Hill	17	20	19	425	501	474
Salyan	Mid-Western	Hill	21	20	21	503	525	550
Achham	Far-Western	Hill	21	21	22	512	503	524
Baitadi	Far-Western	Hill	24	24	24	597	552	605
Darchula	Far-Western	Mountain	16	15	16	411	372	401
Kailali	Far-Western	Tarai	31	33	32	739	847	813
		Total[1]	**557**	**570**	**563**	**13714**	**13971**	**13794**

1) In several schools, more than one subjects were tested.

In the second phase of sampling, number of schools in each of the 28 districts was determined proportionally based on the number of schools in each stratum. In the sample-base for national level student assessments, the conventional maximum sample size is less than 5% of the population (see Cochran 1977; Bartlett, Kortlik & Higgins, 2001).[2] Based

2 There is no rule for 5% of population. However, the classical formula for estimating the sample size of Cochran (1977) reduces the sample size when the suggestion at the first round exceeds 5% of the population. Conventionally, over sampling is suggested when it is expected to see loss in response rates

on the latest official list of schools, the number of schools running 3rd and 5th grades were about 34,000. Thus, the number of schools per grade was fixed to 1700. For grade 5, the number of schools per subject was confined to 566, as assessment was conducted in three subjects. For grade 3, the number of schools per subject was fixed to 850 as assessment was conducted in two subjects. The schools were not selected at this phase, but only the number of schools in each stratum was determined.

Initially, 25 students were targeted to cover randomly from each selected school, but some schools have less than 25 students in the selected grade, and therefore from the big schools within the district with more than 25 students were also selected for assessment to cover the number of students of the schools having less than 25 students.

2.1.3 Selection of schools

In the third phase of sampling, list of all schools was obtained from the Department of Education (DOE) database. The list included information such as the number of students, school type (community or institutional) and address of each school. The information was used as the basis for the random selection of the schools in the district. However, there are some special cases, for example, in Manang district, more than the required number of schools were selected because students were in small number. Student samples were taken to make it reasonably comparable with the other districts to possible extent.

2.2 Item Writing, Pre-testing and Final Test Preparation

This section presents the process applied in item writing, pre-testing and final preparation of test.

2.2.1 Item writing

A team of item writing comprising school-teachers teaching in grade 3 and 5, curriculum officers, and university teachers was formed to accomplish the task of preparing a sufficient number of items for pre-test. Item writing workshop was conducted for each subject in which around 2500 items were prepared in 5 subjects. The number of items was too big for effective testing and thus there was a need for the selection of the best ones. Mathematics items were translated into English and they were screened, edited and re-written when needed. Preliminary pre-testing of grade 3 Mathematics items was conducted in Mustang district in November 2011. The result of pre-testing showed that language of test items affected the test score remarkably. Then the items were revised to include more mathematical expressions rather than linguistic substance. The final pre-test papers were printed in the secured printing press.

2.2.2 Pre-testing

During June 2012, 275-300 test items for each of the five subjects with 20 versions of test papers were pre-tested in 240 schools from 13 districts. The pre-test was administered by the personnel in the District Education Offices (DEOs) after organizing an orientation

---

(see Salkind, 1997, 107; Fink, 1995, 36). In the national level testing, this is not expected and hence over sampling was not planned.

session to the head teachers of respective schools. DEOs monitored the processes in the schools, collected the papers and sent them to ERO in Kathmandu for marking and data entry. Officers from DEOs were also oriented to the process in a one-day seminar where a basic understanding about the objective of testing was shared. ERO personnel closely 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 at DEOs.

2.2.3 Pre-test of PIRLS and TIMSS items

In February 2013, small scale pre-test was carried out in the Kathmandu Valley to acquire the item information about the released items from Trends in International Mathematics and Science Study (TIMSS) in Mathematics and Progress in International Reading Literacy Study (PIRLS) in Nepali and English language reading items. The item parameters were based on the Item Response Theory (IRT) of TIMSS and PIRLS released items, 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 selection of items.

For the pre-test, the selected PIRLS and TIMSS items were first translated into the Nepali language. Then the same items were pre-tested in English for English subject and in the Nepali language for Nepali subject in both grades. Finally, statistically appropriate items, based on reliability, were selected and included in the final test.

2.2.4 Principles followed in selecting items for the final test

Six basic principles were followed when selecting items for the final tests. The six principles are: (1) Content dependence on the curriculum (construct validity), (2) Content's coverage to be as wide as possible (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 PIRLS. Subsequent paragraphs highlight these characteristics of the final tests.

Content dependence of the curriculum

The basis of the construct and content validity of the final tests lies in the "theoretical framework", that is, in the national curricula. In the national assessment, the main idea is to test how well the objectives expressed in the national curricula are fulfilled. Specification grids for each subject and grade were prepared on the basis of the curricula. In the grids, the time allocated to 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; and the marks on the tests is proportional to those percentages in the grid.

Content coverage as wide as possible

To ensure content validity, items were selected from a broad range of topics to cover each as much as possible. 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 as a minimum length to discriminate the test takers from each other sufficiently. Thus, an attempt was made to include as many sub-topics as possible in the test. The selection was, however, proportional. When there were sub-topics of wider coverage in the curriculum than the others, more items were selected from those areas in test construction.

The content coverage was widened by using three different versions in the final testing in each subject of both grades. Several linking items linked all the versions to each other, and they were linked to some international standards by using linking items from TIMSS and PIRLS item banks.

Proper structure of cognitive levels

Bloom's taxonomy on the cognitive domain was used as the basis of cognitive levels. Bloom's original classification of Knowledge, Comprehension, Application, Analysis, Synthesis and Evaluation was shortened into four categories: Knowledge, Comprehension, Application, and Higher skills (Bloom *et al.* 1956; Metfesser, Michael & Kirsner, 1969). 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, decision was made to gear the final tests towards comprehension and application type, rather than knowledge or higher skills. This matches with the international practice in TIMSS and PIRLS tests.

It was noticed that, based on the curricular objectives and contents it was difficult to create items to measure higher skills for the grade 3 students. Because of very few items requiring higher skills, at the final phase, some items were merged with the application type of items. In the same way, there was limitation of the higher skill items in English and Nepali subjects too. In these subjects, however, it was made possible to analyse the higher skills category separately.

High discrimination power of the test

Two technical areas related to high reliability of test were addressed: 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 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 to item difficulty, that is, $\sigma^2 = p\,(1-p)$, the classical formula of alpha reliability can also be obtained using the following formula:

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

Where,
k = number of items
σ_i^2 = variance of the scores on item i
σ_i = standard deviation of the scores on item i
$\rho_{it} = \rho_{gX}$ = item test correlation

It is noteworthy that there were only two sources of information needed for estimating the reliability of the test: the item discrimination (ρ_{it}) and item variance (σ_i^2). It is also noteworthy that the alpha reliability is maximized when the sum of the elements ρ_{it} and σ is the highest. When knowing that 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 high item discrimination as possible and mediocre difficulty level in the test.

Generally, the values for reliability lower than $\alpha = 0.70$ are not taken as accurate enough to be accepted 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). The boundaries are not strict as in Knapp and 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 much lower than 50%. Especially in the Mathematics test, several 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 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 Right/Wrong and True/False type items.

The most balanced test for the national assessment of student achievement is obtained when the items are selected from the whole range of ability. Thus, there should be easy, mediocre and demanding items on the test. These kinds of items can discriminate the best and the poorest as well as the mediocre students. One possible solution, used in NASA 2012 as in NASA 2011, is to select the items so that of the total:

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

Combining the principles, by selecting the easiest items from the pre-test, the aim was to raise the average difficulty level in the tests near $\bar{p} = 0.60$, that is, 0.50 + 0.10 (from 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. The item selection was made based on the classical item difficulty parameter *p*.

Comparability with the results of NASA 2011 and with the international results

The fifth principle in item selection was that the results should be comparable with 2011 results and with international test results. In Mathematics and Nepali, no linking items were selected between grade 5 and grade 8 students. In English, the *Pokhara letter* was taken as a common item from Nepali grade 8 test to the English test.

Better comparison can be obtained when the Nepalese results are compared with the international TIMSS and PIRLS item banks. Based on the released items and their parameters (TIMSS 2007; 2009a; 2009b; Adams & Wu, 2000; PISA, 2006a; 2009), and by using IRT modelling, it is possible to find the baseline for the comparisons. 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. The items used in NASA 2012 come from the TIMSS Mathematics items released in the year 2007 and PIRLS reading comprehension texts and items released in the year 2006. Both international banks are for grade 4 students; the items were supposed to be somehow easy to grade 5 students and somehow difficult to grade 3 students. In Mathematics, the released parameters of TIMSS of the selected items were fixed at the same time in both grades 3 and 5. In Nepali, the calibration was done first separately in grade three and grade five and later they were re-scaled to make the scores comparable. In both cases the test scores are equated separately in Maths, Nepali and English and, hence, the scores are made comparable in both the grades but not over the subjects.

2.2.5 Final tests

In each of the grades 3 and 5, three test versions were administered simultaneously in the same classroom in each subject. Standards were set using IRT modelling for the assessment of Mathematics in order to keep the results comparable. The following principles were adopted from IRT modelling practices: First, no decimal number scores are allowed in the assessment during the answer sheet marking i.e., the students' responses are always marked in whole numbers. If the students are not qualified to secure full score, 0.5 score is not provided in any case. Second, the marking scheme has to be rigorously prepared to make exactly the same judgment in the years to come with the linking items. Third, IRT modelling requires that all possible marks have to be observed in the dataset. Finally, IRT modelling requires a linking procedure between the different versions of the test. All the versions were linked with each other using the identical linking items. The common items for each test version, the linking items, were carefully selected from the pre-tested items, TIMSS released items of Mathematics and PIRLS released items for Nepali and English. While selecting items it was considered that the standards set by the national curriculum should be reflected.

Mathematics test

All versions of Mathematics test (M_1, M_2 and M_3) were quite similar in content characteristics. The construct validity is quite high from curriculum viewpoint as the number of items in each subject match quite well with curricular contents. There were more Numeracy and less Arithmetic items in grade 3; and less Numeracy items in grade 5 and more Geometry items in the test compared with the curriculum content weight. Different versions include a wider variety of sub-topics under each topic.

Classical item and test analysis methods were used in pre-test phase for finding the percentage of correct answers, that is, item discrimination power. IRT was used for item calibration, finding the latent ability (Theta, θ) as well as comparing and equating the versions M_1, M_2 and M_3 and TIMSS database. SPSS software was used for the classical analysis and One Parametric Logistic Model software (OPLM) (Verhelst, Glas, Verstralen, 1995) was used for IRT modelling. The parameters of the international items were fixed during the item calibration so that all the test items of the year 2012 were calibrated in the international TIMSS scale. After the calibration of the items, all the scores in the versions M_1, M_2 and 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 comparable. The original output is the latent ability (θ) which is a standardized normal score ranging usually from – 4 to + 4 in grade 3 and – 3 to + 3 in grade 5. These values in each test versions were later transformed into equated scores and further the equated scores were converted into percentage.

Table 2.3 shows the average marks (mean) of grade 3 Mathematics calculated for three versions M_1, M_2 and M_3 separately. Based on the pre-tested items, the versions M_2 and M_3 were of the same length whereas M_1 was longer than M_2 and M_3. In grade 3 tests, 23 items were linked versions M_1 and M_2; 26 items were linked versions M_2 and M_3; and 39 items were linked versions M_1 and M3. The longer version M_1 carried maximum of 61 marks whereas the shorter versions M_2 and M_3 carried 57 and 56 marks respectively. There were four TIMSS items as the linking items all over three versions.

Table 2.3 Comparison of the characteristics of Mathematics (grade 3) test versions

Version	N	Maximum marks	Original mean score	Equated Mean[1]	SD[2]	CV[3]
M_1	6,487	61	29	60	26.1	43.7
M_2	6,183	57	33	58	25.3	43.4
M_3	6,582	56	26	60	25.5	42.2
Total	**19,251**			**59**	**25.6**	**43.1**

1)Percentage of the equated maximum score rounded in whole number
*2) SD = Standard Deviation 3) CV = Coefficient of Variation = SD/Mean*100*

In grade five, 25 items were common to all versions and 42 items were linked - either M_1 and M_3 or M_2 and M_3. Among six TIMSS items used as the linking items in M_3 four items were taken from the items used in grade 3 tests. The number of linking items is high for obtaining stable calibration between the versions and sufficient for calibrating the scales over the grades. The shorter versions M_2 and M_3 were of a maximum of 58 marks and 59 marks respectively whereas the longest version M_1 was of 61 marks. However, because of low item-total correlation, two items from M_1, two items from M_2 and one item from M_3 were discarded during the item analysis.

Table 2.4 shows the average marks (mean) of grade 5 calculated for three versions, M_1, M_2, and M_3, separately. Based on the pre-tested items, the versions M_2 and M_3 were of the same length whereas M_1 was longer than the other two versions. As all the versions of test were linked with each other by the use of the several identical linking items, the equated mean score in all three versions were almost equal.

Table 2.4 Comparison of the characteristics of Mathematics (grade 5) test versions

Version	N	Maximum marks	Original mean score	Equated Mean[1]	SD[2]	CV[3]
M1	4591	61	32	53	21.9	41.1
M2	4363	57	32	54	24.6	45.9
M3	4660	56	32	53	22.4	42.2
Total	**13614**			**53**	**23.0**	**41.1**

1)Percentage of the equated maximum score rounded in whole number
*2)SD = Standard Deviation 3)CV = Coefficient of Variation = SD/Mean*100*

The test items were classified into four categories: Arithmetic, Algebra, Geometry and Numeracy (see, tables 2.5 and 2.6). Arithmetic includes basic operations, time, money and measurement, fraction, decimal, percentage, unitary methods and simple interest, bill and budget and statistics. Similarly, Algebra includes algebra and sets. Numeracy includes the knowledge of numbers whereas Geometry includes shape, size and their measurement

Table 2.5 Characteristics of Mathematics tests of grade 3 in various content areas

Topic	Marks			Percentages			Percentages in Curriculum	Reliability			
	M1	M2	M3	M1	M2	M3		M1	M2	M3	M1–M3[1]
Total	49	57	46				100	0.94	0.94	0.93	0.936
Arithmetic	27	36	32	55.1	63.2	69.6	75	0.91	0.92	0.90	0.908
Algebra	3	2	4	6.1	3.5	8.7	9	0.52	0.37	0.44	0.445
Geometry	7	5	3	14.3	8.8	6.5	7	0.72	0.62	0.51	0.612
Numeracy	12	14	7	24.5	24.6	15.2	19	0.66	0.80	0.67	0.710

1) *Weighted mean of reliabilities in M1 to M3*

In grade 3, overall internal consistencies (given by Alpha-reliability) of the whole tests on each version were very high (α = 0.93 to 0.94), however, some of the categories (Algebra and Geometry) contains a few items, and hence reliability is somehow lower. The reliability of the score in the total sample cannot be given in a classical way because it can be estimated only version-wise.

Table 2.6 Characteristics of Mathematics tests of grade 5 in various content areas

Topic	Marks			Percentages			Percentages in Curriculum	Reliability			
	M1	M2	M3	M1	M2	M3		M1	M2	M3	M1–M3[1]
Total	61	57	56				100	0.93	0.94	0.93	0.934
Arithmetic	33	31	31	54.1	54.4	55.4	58	0.89	0.90	0.89	0.890
Algebra	11	10	10	18.0	17.5	17.9	15	0.75	0.75	0.75	0.750
Geometry	10	10	9	16.4	17.5	16.1	9	0.60	0.68	0.68	0.650
Numeracy	7	6	6	11.5	10.5	10.7	18	0.68	0.70	0.69	0.690

1) *Weighted mean of reliabilities in M1–M3*

In grade 5, reliabilities of the total scores of the different versions were over $\alpha = 0.90$, that is, the test scores could discriminate the individual pupils with high accuracy. Much less would have been sufficient because there is no intention to use the tests as the examinations. However, the more discriminating the tests, the more accurate will be the outputs. Because of the limited number of items, the reliability of Algebra test in grade remains low.

The items used in the tests varied from objectively scored items (that is, the multiple choice items, fill in the blank, true or false, very short answer items) to subjectively scored, usually productive items (short answer type and long answer type items). The test also met all the cognitive levels of Bloom's taxonomy (see Bloom *et al.* 1956; Metfesser, Michael, & Kirsner, 1969). Higher skills are measured mainly by the open-ended, productive type of questions. In Mathematics, there were very few items measuring higher skills. Almost all the items that needed higher skill were close to the level of Application. Hence, the questions of both the categories combined together as application and higher skills.

Table 2.7 Characteristics of various domains of Mathematics tests of Grade 3

Levels of cognitive domain	Marks			Percentages			Reliability			
	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1–M3[1]
Knowledge	17	8	11	34.7	14.0	23.9	0.86	0.82	0.79	0.823
Comprehension	14	22	13	28.6	38.6	28.3	0.79	0.84	0.75	0.792
Application and Higher skills	18	27	22	36.7	47.4	47.8	0.83	0.88	0.83	0.846

1) *Weighted mean of reliabilities in M1–M3*

Table 2.8 Characteristics of various domains of Mathematics tests of Grade 5

Levels of cognitive domain	Marks			Percentages			Reliability			
	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1–M3[1]
Knowledge	14	11	14	23.0	19.3	25.0	0.73	0.73	0.78	0.748
Comprehension	16	17	16	26.2	29.8	28.6	0.80	0.82	0.80	0.805
Application and Higher skills	31	29	26	50.8	50.9	46.4	0.87	0.87	0.86	0.867

1) *Weighted mean of reliabilities in M1–M3*

In grade 3 Mathematics (table 2.7), 33% of the items were of comprehension type, 39% items were of application type, 5% of higher ability and 23% of knowledge type. Similarly, in grade 5 Mathematics (table 2.8), 23% of items were of knowledge, 28% were of comprehension and 49% were of application and higher ability level. It shows that more focus was given to comprehension and application.

Nepali language tests

As in Mathematics test, classical item and test analysis methods were used in the pre-test phase of Nepali subject to find the percentage of correct answers, item discrimination power, and the test reliability. IRT was used for item calibration, finding the latent ability (Theta, θ) as well as comparing and equating the versions N_1–N_3 and PIRLS database. SPSS software was used for the classical analysis, and OPLM software (Verhelst, Glas, Verstralen, 1995) was used for IRT modelling. The parameters of the international items were fixed during the item calibration so that all the test items of the year 2012 were calibrated in the international PIRLS scale. After calibration of items, all the scores in versions N_1–N_3 were transformed into the same scale, that is, the scores were equated. This means that all the scores in each test version were made comparable. The original output is the latent ability (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.

Table 2.9 shows the average marks (mean) of grade 3 calculated for three versions N_1, N_2 and N_3 separately. All the versions (based on pre-test) were parallel when it comes to their maximum values (77 in each); N_1 looks to be somehow easier than N_2 and N_3. All the versions were linked with each other by using the identical linking items. There were four PIRLS items as the linking items in all the three versions.

Table 2.9 Characteristics of the Nepali test of grade 3

Version	**N**	**Number of max. marks**	**Original mean score**	**Equated Mean**	**SD**	**CV**
N_1	6307	77	48	62	24.4	39.1
N_2	6438	77	47	63	23.2	36.7
N_3	6756	77	47	62	24.4	39.2
Total	**19,501**			**63**	**24.0**	**38.7**

Table 2.10 shows the average marks of grade 5 calculated for three versions N_1, N_2 and N_3 separately. Based on the pre-tested items, all the versions were parallel whereas N_1 was relatively shorter. All the versions were linked with each other by using the identical linking items. The N_3 had the highest score 91 and N_1 had the lowest score 89. There were six PIRLS items as the linking items over the three versions.

Table 2.10 Characteristics of Nepali test of grade 5

Version	N	Number of Max. Marks	Original Mean Score	Equated Mean	SD	CV
N_1	4,570	89	55.1	60	22.7	37.7
N_2	4,519	90	53.4	58	22.6	39.1
N_3	4,882	91	54.6	59	23.1	39.1
Total	**13,971**			**59**	**22.8**	**38.7**

The Nepali language test was planned to take one hour more to complete than the other tests. The reason for this is that it was decided to include 30–45 minutes just in reading the texts used in reading comprehension test. In grade 3 tests, 17 items were common for all versions and additionally 7 items were linked - either N_1 and N_2 or N_1 and N_3. In grade 5, 25 items were common for all versions and 42 items were linked - either N_1 and N_3 or N_2 and N_3. Large number of linking items were included in each test in order to obtain stable calibration between the versions.

The test items were classified into four categories: Reading, Writing, Grammar and Vocabulary. It is worth noting that Grammar and Vocabulary are not mentioned explicitly in the curriculum for grade 3 though “functional grammar” is used as a term to cover these contents. However, Listening and Speaking are explicitly mentioned in the curriculum. For grade 3, in the final versions of test, the weight of Reading and Writing items are 38% and 42% respectively. The items used in the tests varied from objectively scored items - that is, the multiple choice items, fill in the blank, true or false, very short answer items (72% of the total items) - to subjectively scored, usually productive, items (short answer type and long answer type items, which were 28% altogether).

Table 2.11 Characteristics of Nepali tests of Grade 3 in various content areas

Content Areas	Marks			Percentages			Percentages in Curriculum	Reliability			
	N_1	N_2	N_3	N_1	N_2	N_3		N_1	N_2	N_3	N_1–N_3[1]
Total	77	77	77					0.95	0.95	0.95	0.95
Reading	27	30	30	35.1	39.0	39.0	25	0.87	0.90	0.89	0.89
Writing	33	34	32	42.8	44.2	41.5	25	0.89	0.87	0.88	0.88
Grammar[2]	12	9	10	15.6	11.6	13.0	5	0.89	0.81	0.84	0.85
Vocabulary	5	4	5	6.5	5.2	6.5		0.73	0.65	0.68	0.69
Listening							20				
Speaking							25				

1) Weighted mean of reliabilities in N_1–N_3 2)Functional grammar

In grade 3 Nepali, overall internal consistencies (given by reliability) of the whole tests on each version were very high (α = 0.95). The reliability of the score in the total sample cannot be given in a classical way because it can be estimated only version-wise.

Similarly with grade 3,the test items of grade 5 also were classified into four categories: Reading, Writing, Grammar, and Vocabulary. It is worth noting that the curriculum for grade 5 students do not explicitly include Grammar and Vocabulary though weightage is allotted for the Functional Grammar integrating Vocabulary in it. Instead, Listening and Speaking are explicitly mentioned. For grade 5, overall internal consistencies (given by reliability) of the whole tests in each version were very high (α = 0.95). Due to the few items in some of the categories (Grammar and Vocabulary), however, the reliability is somehow lower. The reliability of the score in the total sample cannot be given in a classical way because it can be estimated only version-wise. From validity point of view, it is notable that, in the Nepali curriculum, the weighting of reading items (30%) is less than Writing (35%) and much less in Grammar (5%). In the final version, the counting over the versions, the difference in the weight of Reading (28.1%) and Writing (38.9%) is somehow wider than in the curriculum, and both Grammar (18.5%) and Vocabulary (14.9%) are weighted more than in curriculum. The items used in the tests varied from objectively scored items (that is, the multiple choice items, fill in the blank, true or false, very short answer items, 54.8% of items) to subjectively scored, usually productive items (short answer type and long answer type items, which were 45.2% of the total items).

Table 2.12 Characteristics of Nepali tests of Grade 5 in various content areas

Topic	Marks			**Percentages**			**Percentages in Curriculum**	**Reliability**			
	N_1	N_2	N_3	N_1	N_2	N_3		N_1	N_2	N_3	N_1–N_3[1]
Total	89	90	91					0.95	0.96	0.96	0.96
Reading	22	30	24	24.6	33.3	26.4	30	0.85	0.89	0.87	0.87
Writing	35	35	35	39.3	38.8	38.5	35	0.88	0.88	0.88	0.88
Grammar[2]	16	18	16	18.0	20.0	17.5	5	0.85	0.85	0.85	0.85
Vocabulary[3]	16	8	16	18.0	8.9	17.6		0.82	0.70	0.83	0.80
Listening							15				
Speaking							15				

1) *Weighted mean of reliabilities in N1–N3*
2) *Grammar is mentioned as Functional grammar in grade 5 curriculum.*
3) *Vocabulary is integrated in Functional grammar.*

In grade 5 Nepali, overall internal consistencies (given by reliability) of the whole tests in each version were very high (α = 0.95 to 0.96). The reliability of the score in the total sample cannot be given in a classical way because it can be estimated only version-wise.

As the cognitive levels of Bloom's taxonomy (see Bloom et al. 1956; Metfesser, Michael, & Kirsner, 1969) are met though weighted toward Comprehension (57% and 47% of the items in grade 3 and 5 respectively), Application and Higher Skills (23% and 23%) on expense of Knowledge (8% and 10%) and Higher skills (12% and 21%) (Tables 2.13 and 2.14).

Table 2.13 Characteristics of Nepali tests of Grade 3 in various cognitive domain

Hierarchical level	Marks			Percentages			Reliability			
	N_1	N_2	N_3	N_1	N2	N_3	N_1	N_2	N_3	N_1–N_3[1]
Knowledge	5	7	7	6.4	9.0	9.0	0.66	0.72	0.70	0.694
Comprehension	42	45	46	53.8	57.7	59.0	0.92	0.91	0.92	0.917
Application	19	17	18	24.4	21.8	23.0	0.89	0.81	0.82	0.840
Higher skills	12	9	7	15.4	11.5	9.0	0.51	0.46	0.48	0.483

1) Weighted mean of reliabilities in M_1–M_3

Table 2.14 Characteristics of Nepali tests of Grade 5 in various cognitive domain

Hierarchical level	Marks			Percentages			Reliability			
	N_1	N_2	N_3	N_1	N2	N_3	N_1	N_2	N_3	N_1–N_3[1]
Knowledge	6	11	9	7.0	12.5	9.9	0.66	0.77	0.73	0.721
Comprehension	42	40	42	48.7	45.5	46.2	0.93	0.93	0.93	0.930
Application	21	18	21	24.4	20.5	23.0	0.89	0.89	0.90	0.893
Higher skills	17	19	19	19.8	21.6	20.9	0.76	0.78	0.80	0.780

1) Weighted mean of reliabilities in M_1–M_3

English language test

Table 2.15 shows the average scores (mean) calculated for three versions E_1, E_2 and E_3 separately. Based on the pre-tested items, the versions E_1 and E_2 were parallel, whereas E_3 was shorter. All the versions were linked with each other by the use of the identical linking items. The longer versions E_1 and E_2 were scored out of a maximum of 71 and 75 maximum marks and E_3 had a maximum score of 60 maximum marks. There were four PIRLS items as the linking items all over three versions.

Table 2.15 Comparison of the characteristics of English test versions

Version	N	Number of max. marks	Original mean score	Equated Mean	SD	CV
E_1	4,524	71	44	54	24.1	44.6
E_2	4,657	75	42	53	24.5	46.0
E_3	4,621	60	38	54	24.1	45.0
Total	**13,802**			**54**	**24.2**	**45.2**

Classical item and test analysis methods were used in the pre-test phase for finding the percentage of correct answers and the item discrimination power. IRT was used for item calibration ,finding the latent ability (Theta, θ) as well as comparing and equating the versions E_1–E_3 and PIRLS database. SPSS software was used for the classical analysis and OPLM software (Verhelst, Glas &Verstralen, 1995) was used for the IRT modelling. The parameters of the international items were fixed during the item calibration so that all the test items of the year 2012 were calibrated in the international PIRLS scale. After the calibration of the items, all the scores in the versions E_1–E_3 were transformed into the same scale, equating the scores. This means that all the scores in each test version are comparable. The original output is the latent ability (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 the equated scores were further converted into percentage.

The test items were classified into four categories: Reading, Writing, Grammar and Vocabulary (table 2.16). It is worth noting that the curriculum for the grade 5 students do not explicitly include Grammar and Vocabulary, though "functional grammar" is used as a term. Instead, Listening and Speaking are explicitly mentioned. From the beginning, it was thought that Speaking and Listening should be taken into account in the assessment. However, organizing these kinds of tests with objective and comparable manner would have been practically impossible in the situation where schools are lacking electricity and necessary equipments for organizing and scoring the listening test for all. Hence, they were omitted from the final test. In order to compare the results in years to come with 5th and 8th grade students, Vocabulary and Grammar were included into the test battery.

Table 2.16 Characteristics of English tests of Grade 5 in various content areas

Topic	Marks			Percentages			Percentages in Curriculum	Reliability			
	E_1	E_2	E_3	E_1	E_2	E_3		E_1	E_2	E_3	E_1–E_3[1]
Total	71	75	60					0.95	0.96	0.94	0.95
Reading	28	23	24	39.4	31.1	40.7	25	0.92	0.89	0.89	0.90
Writing	26	31	17	36.6	41.8	28.8	25	0.81	0.82	0.79	0.81
Grammar	8	11	7	11.3	14.9	11.9		0.75	0.79	0.64	0.73
Vocabulary	9	9	11	12.7	12.2	18.6		0.71	0.82	0.80	0.78
Listening							25				
Speaking							25				

1) Weighted mean of reliabilities in E_1–E_3

Overall internal consistencies (given by Alpha-reliability) of the whole tests on each version were very high (α = 0.94 to 0.96), however, some of the categories (Grammar and Vocabulary) contain fewer number of items and hence the reliability is somehow lower. The reliability of the score in the total sample cannot be given in a classical way because it can be estimated only version wise.

From validity point of view, it is notable that, in the English curriculum, the weighting of Reading items is the same as of writing (25%). In the final version, counting over the versions, the weight of Reading and Writing items are nearly equal (37% and 36% respectively). The items used in the tests varied from objectively scored items (that is, the multiple choice items, fill in the blank, true or false, very short answer items, 52% of items) to subjectively scored, usually productive items (short answer type and long answer type items, 48% of items). All the cognitive levels of Bloom's taxonomy (see Bloom et al. 1956; Metfesser, Michael & Kirsner, 1969) are met though weighted towards Application (41%)

and Comprehension (44%) on expense of Knowledge (6%) and Higher Skills (9%) (see, table 2.77). Higher skills are measured mainly by open-ended, productive type of questions.

Table 2.17 Characteristics of English tests of Grade 5 in various cognitive domains

Hierarchical level	Marks			Percentages			Reliability			
	E_1	E_2	E_3	E_1	E_2	E_3	E_1	E_2	E_3	E_1–E_3[1]
Knowledge	5	5	5	6.0	6.0	5.4	0.61	0.63	0.60	0.613
Comprehension	38	37	38	45.8	44.6	41.3	0.90	0.91	0.89	0.900
Application	32	37	38	38.6	44.6	41.3	0.90	0.88	0.81	0.863
Higher skills	8	4	11	9.6	4.8	12.0	0.66	0.33	0.70	0.562

1) *Weighted mean of reliabilities in E_1–E_3*

2.2.6 Marking Schemes

The tasks of data collection, marking, and data entry were outsourced to a consultancy firm. Because the company has neither been involved strictly in the item writing, pre-testing, nor in marking of the pre-test papers[3], it was important to prepare a marking scheme for each subjective type of item[4] so that marking would be as reliable as possible. A marking scheme was prepared by adding strict examples of what kind of correct answers should be allowed for the marks and what kind of answers should not be credited for marks.

2.3 Test Administration, Marking, and Data Entry

2.3.1 Test administration

The test was administered on the same day, the end of February in all the 28 sample districts. The administration was outsourced to a consultancy firm, and the officers from the MOE/ERO monitored the process. Monitoring was done also by DEO and RED personnel. Actual information was collected by the district offices. Test administration was smooth in almost all the cases. The total number of the schools participated in assessment was 1704.

2.3.2 Marking and data entry

The marking of the papers as well as the data entry were outsourced. Most of the questions were objective – each questions carrying one mark. Subjective items, including short answers, carried up to 3 marks in mathematics tests. In Nepali and English tests, the subjective items were somehow long answer type, carrying up to 5 marks.

Altogether 80232 students' papers, 1,704 teachers' questionnaires and 1704 head teachers' questionnaires were collected into a "marking centre" in Kathmandu. Within three months of April, May, and June 2013, the outsourced firm marked the answer papers and prepared dataset. Teachers' dataset was entered two times because several mistakes were detected.

[3] Some of the markers in the company were involved in the processes but most were not.

[4] Most of the objective type of items – True/False, Multiple-choice, and Matching type of items – are not marked.

Because of human errors, the datasets included a notable number with incorrect data units[5]. Detecting these and correcting them for the school-wise reports took a considerable length of time. The student datasets were screened and, except for some minor mistakes in language groupings, no mistakes were found in the item and background-wise datasets. This was assured in advance by programming the maximum values of the items to the data entry software and requiring double entry and cross checking of the input.

2.4 Background Questionnaires

In student assessment, as important as describing the achievement results is trying to find the factors which explain the differences between the students and schools. Hence, one needs background questionnaires including the relevant contextual information related to the students, students' family as well as the school and teaching. This kind of variable is, for example, the socioeconomic status (SES) of the pupil's family. It is known that these characteristics of the pupil seem to explain the achievement level quite well, but it cannot be changed easily as the poor families cannot be made rich and it may be unethical to try to make the rich families poor. Hence, one cannot do much for the phenomenon. On the other hand, if the reason for poorer results is the low educational level or illiteracy of the parents (part of the SES), the government can do something to improve the adult literacy rate in the country.

2.4.1 Conceptual model for the background questionnaires

A sketchy modelling of the complex phenomenon of learning is in use in the Finnish National Board of Education(FNBE) (Metsämuuronen, 2009) as well as in NASA 2011 (ERO, 2013). The same model, with contextual modification, was used as the basis for compiling the background questionnaires in grade 3 and 5 (fig. 2.2). 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 close to the pupils are the **peer group factors**: social support to studies, bullying, atmosphere in the classroom, and so on. All the information related to these factors is based on the students' background questionnaires. **Teacher factors** evidently play some role in student achievement, however, usually much less than what is expected. Such factors as classroom activities, teaching skills, use of teaching materials, for example, probably affect the student achievement. These are asked to the teacher. School factors can be divided into two: **managerial factors** and **physical factors**. These factors, such as atmosphere in the school, the condition of the school premises, safety, absence, and so on, are asked to the head teacher or principal. **Economic factors** and

[5] The datasets included, among other things, numerous incorrect school- and district codes (alone in Mathematics, there was found 1777 students with a systematical error in the district code), several doubled students (with exactly the same numerical arrow), several head-teacher papers from the same school, and so on. Additionally, about 100 teachers were not able to be identified whether they were from grade 3 or 5.

demographic factors are available in national statistics – in the NASA activities, the demographic factors are part of the sampling scheme.

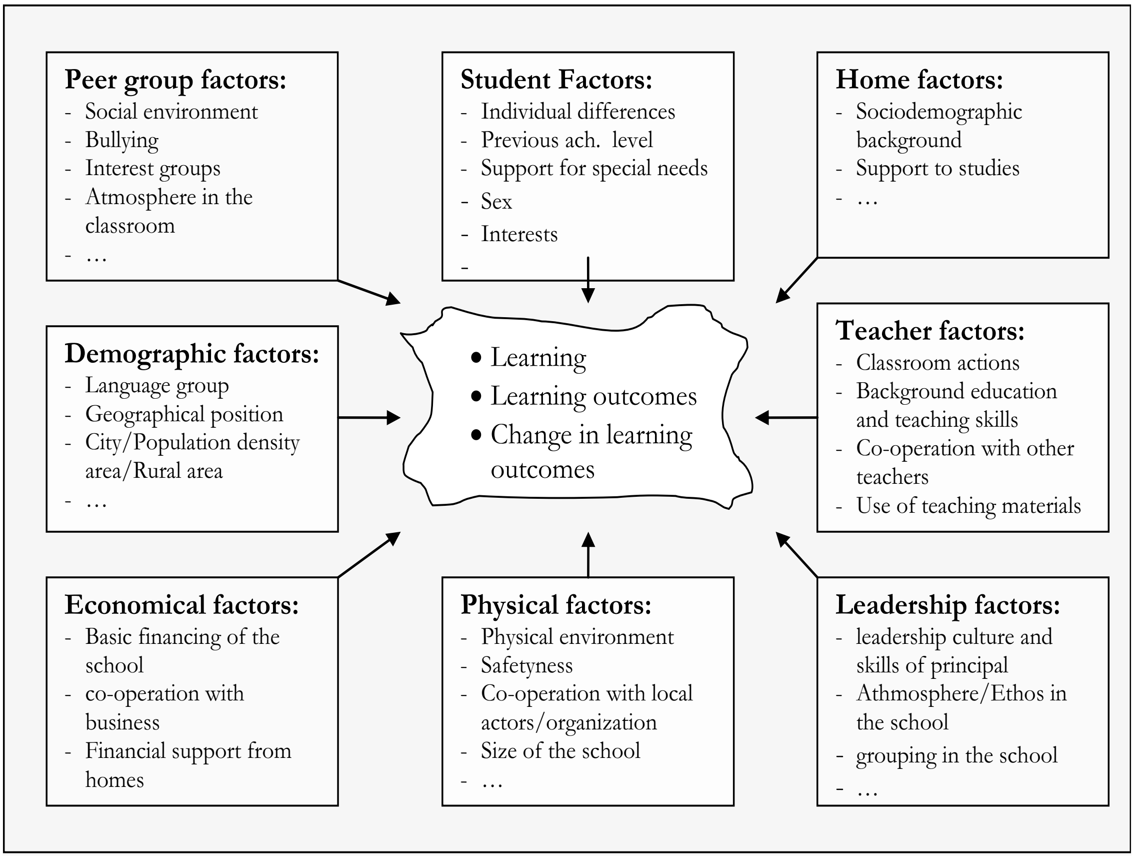

Figure 2.2 Conceptual framework for the background information

(Adapted and contextualized from Metsämuuronen, 2009)

2.4.2 Selection of questionnaire

The variables for the background were selected in several phases. At the first phase, the statisticians were familiarized with the 2008 background questionnaire and with both TIMSS and PISA questionnaires (PISA 2003a; 2003b; 2006b; 2006c; TIMSS, 2003; 2006). In the second phase, the regression questionnaire models prepared in 2008 were studied. Two things were decided: 1) the variables which showed statistically significant relevance with the student achievement in 2008 were decided, at least to include in the questionnaire, and 2) there would be three sets of questionnaires: questionnaires for students, teachers and head teachers. The family questionnaire was omitted and, instead, the questions related to the family were asked to the students. At the third phase, the 2008 background questionnaire was taken as the basis for the question selection; relevant items were kept and intuitively (or statistically) non-significant variables were omitted. This checked version was then added with relevant items from TIMSS and PISA questionnaires. Many SES variables and motivational variables were borrowed from the international questionnaires. At fourth phase, a number of national experts and researchers went through the questions to discard some of the selected ones or to add some new and more relevant variables. At the final phase, the subject committees took the final stand for the questionnaires. Altogether three background questionnaires were prepared based on NASA 2011 questionnaire: one for the students (student and family related questions), one for the teachers (teacher and teaching related questions), and one for the head teacher (teacher, teaching, school and resource related questions). After having discussed with the local experts, some changes were made in the items from 2011 background questionnaire in which some items were made easier to understand, some were added as alternatives (like the items of occupation of the parents), and some were added (like the future aim of the students). Practically, though, items of 2011 were used after some minor change in all subjects.

Though the questions were known to be somehow demanding for grade 3 students, the background questionnaire was the same for grade 3 and 5 students. The teacher was informed to help the grade 3 students when needed, for example, with the questions such as parents' education and occupation. It is notable, though, that there are quite many missing values in the grade 3 datasets and in some variables, almost 25% students did not answer the background questionnaire. This seemed to be systematic: in many cases, not all the students in a school answered the background questionnaire. Most probably, the reason lies in the teacher in the classroom.

2.5 Specific Variables Used in Analysis

Within the analysis, three sets of variables are worth handling more carefully in order to fully understand the results: the concept of equated scores, Fennema-Sherman attitude scale (Fennema & Sherman, 1976), and the indicator for socioeconomic status (SES). Because three different versions with different lengths and possible different difficulty levels were used in student achievement, the scores are not automatically comparable. Hence, the scores were to be equated before the analysis; the logic and procedure of these equated scores are handled in this section. The Fennema-Sherman test is a widely used test (e.g. in PISA and TIMSS questionnaires) to measure attitude towards school subjects.

2.5.1 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 to each other. Thus, it was possible to equate the test scores with IRT modelling (Lord, 1980; of equating, see Béguin, 2000) and finally to acquire the comparable latent ability of each student over the different versions. IRT modelling is the very tool for equating test scores in the well known international comparisons of PISA and TIMSS studies. As the modern test theory, IRT modelling replaced the classical test theory when it comes to complex testing settings with different test versions and a need to compare the results over the years. The testing procedure used in NASA 2012 is typically this kind of complex endeavour, which benefits from IRT modelling. IRT modelling is the only credible way to assess the achievement level in grade 3 and 5 at national level, and to compare the results in Nepal with the international standards (such as PIRLS and TIMSS results).

The need for equating comes from four facts. First, to widen the number of items and thus the range of asked topics and sub topics in the testing process, it was natural to use several versions in testing. In NASA 2012 testing, three versions were used. As a comparison, in PIRLS and TIMSS testing procedures more than one versions are administered in each school. Second, in order to compare the results of grade 3 to 5 and grade 5 to 8, IRT modelling is, by far, the most accurate method in order to perform the comparison when equal (or parallel) tests are not used. Third, the comparison with the international level results unquestionably requires IRT modelling because the only knowledge available of the items was the IRT difficulty parameter β. Fourth, using IRT modelling made it possible to free one of the equal lengths of the test papers. When the tests were of unequal length and deliberately of somehow 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 preconditions 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 so called 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 more exhaustively in Béguin, 2000, 17–36):

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

The success of the equating depends on three things. First, the linking items should represent a sufficient range of ability level and too easy and too difficult items should, however, be avoided. Second, the linking items should represent a short test inside the test, which should cover the different content areas as widely as possible. Third, the stable parameters in the equation 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 metrics (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 of zero ($\theta = 0$). The better the student, the higher is his/her θ above the zero line and parallel and conversely the weaker the student, the lower is θ below the zero line. Now, 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 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 in comparison to the international standard.

At the final phase, the Mathematics scores were calibrated into the TIMSS Mathematics scale of grade 4, Nepali and English reading scores were calibrated into the PIRLS grade 4 reading scale. In Section 3, the results produced by the IRT modelling are hidden, and the original scores are transformed into the equated scores and the equated scores are changed to percentages of the maximum score. Hence, the score 100 means that the student was able to solve, in theory, all the tasks of the total score or sub-scores of the maximum marks.

2.5.2 Fennema-Sherman attitude scale

A shortened version of Fennema–Sherman Attitude Scales (Fennema & Sherman, 1976) are used in several international comparisons, like in TIMSS 2007 (Mullis, Martin, & Foy, 2008) and its predecessors (1995, 1999, and 2003) as well as in PISA. Original scales include nine dimensions, but in these international comparisons, only three dimensions with four items in each (see table 2.18) and two negative items in each of the first two dimensions are used. The names of the factors can be "Liking Math", "Self-Efficacy in Math", and "Experiencing utility in Math" (see Metsämuuronen 2012a, 2012b; compare naming in, e.g., Kadijevich, 2006; 2008). Because of students' inconsistent manner in answering the attitude-scale in NASA 2011, only the dimension of "Experiencing utility in Math/Nepali/English" was taken into the measurement instrument of grade 5 students. The item-total correlations and reliability of the score are presented in tables 2.18 and 2.19. Reliability of the score of five items is sufficient ($\alpha = 0.74$).

Table 2.18 Reliabilities of the FSAS "Utility" scale in NASA 2012 (grade 3)

Item	Item-total correlation	
	Mathematics (N = 13,687)	**Nepali (N = 18,024)**
q19a English subject helps me in the daily life,	0.477	0.695
q19b I need to do better in English to read the subject in the upper grades.	0.562	0.785
q19c English is necessary in order to work abroad.	0.584	0.768
q19d English is necessary in order to study abroad.	0.57	0.738
q19e I need to do better in English to get a desired job.	0.516	0.776
Reliability	**0.769**	**0.899**

Table 2.19 Reliabilities of the FSAS "Utility" scale in NASA 2012 (grade 5)

Item	Item-total correlation		
	Mathematics (N=11,562)	**Nepali (N=13,534)**	**English (N=11,170)**
q19a English subject helps me in the daily life.	0.482	0.617	0.425
q19b I need to do better in English to read the subject in the upper grades.	0.54	0.76	0.535
q19c English is necessary in order to work abroad.	0.542	0.763	0.499
q19d English is necessary in order to study abroad.	0.546	0.715	0.551
q19e I need to do better in English to get a desired job.	0.508	0.741	0.481
Reliability	**0.756**	**0.882**	**0.736**

2.5.3 SES variables

According to Bradley and Corwyn (2002), socioeconomic status (SES) is one of the most studied constructs in social studies. The construct seems to have interested the researchers because of the belief that high SES families provide an array of services, goods, parental actions, and social connections for their children that potentially rebound to the benefit of them 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 to cognitive and academic attainment (see a convincing literature in, for example, Bradley & Corwyn, 2002; APA, 2007).

In the literature, social status is commonly conceptualized in terms of socioeconomic standing, taking into account the various combinations of income, education, and occupation (APA 2007, p 5). The challenge in measuring SES is that there has not been a complete consensus on precisely what represents economic position or social status (Liberatos, Link& Kelsey, 1988; McLoyd, 1997), economic position or social status and hence there is not a single measure for SES (Bradley & Corwyn, 2002; APA, 2007:5). Bardley and Corwyn (2002: 373) put it as follows: Although there is general consensus 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, a statistical procedure that includes each indicator, or each indicator singly; or (c) how best to measure each component (Krieger et al., 1997).

The following seven indicators of SES were selected into the final SES indicators:

- Father's education
- Mother's education
- Father's occupation
- Mother's occupation

- Home possessions
- Home accessories
- Attending to private school

The indicators and their cut-offs are presented in table 2.20. Note that there are some minor changes in SES indicators of NASA 2012 compared to the NASA 2011. One is that there are less home accessories in the questionnaire in 2012. In home possessions also, the 12 possessions of NASA 2011 were shortened to 11 possessions in NASA 2012. Hence, lower maximum scores.

Table 2.20 Indicators of SES in NASA 2012 (English grade 5)

Variable	**Cut-off**[1]	**Effect on total score**[2]
Father's education	Less than grade 10-passed = 0, other = 1	+17% points, $\eta^2 = 0.096$
Mother's education	Less than grade 10-passed = 0, other = 1	+19% points, $\eta^2 = 0.085$
Father's occupation	Agriculture = 0, other =1	+12% points, $\eta^2 = 0.057$
Mother's occupation	Agriculture = 0, other =1	+10% points, $\eta^2 = 0.039$
Home possessions	5 or less out of 11 possessions=0,6 or more=1	+13% points, $\eta^2 = 0.072$
Home accessories		
Mobile phone	2, 3 = 1, other = 0	+13% points, $\eta^2 = 0.067$
Television	1–3 = 1, other = 0	+14% points, $\eta^2 = 0.079$
Computer	1–3 = 1, other = 0	+3% points, $\eta^2 = 0.004$
All together	0–1 out of 3 = 0, other = 1	+13% points, $\eta^2 = 0.071$
Attending to private school	No = 0, yes = 1	+35% points, $\eta^2 = 0.399$
Total SES		+39% points, $\eta^2 = 0.311$

1) Based on DTA 2) Based on one-way ANOVA

Because the variables were of different scales (from nominal to ordinal scales) and because of incomparable scores (from 0 –1 to 0 –11), all the variables were rescaled first to fit with each other. At the first phase, the variables were analysed with respect to educational outcomes. Decision Tree Analysis (DTA)(see section 2.6), the data mining tool in SPSS software, and ANOVA – the basic tool for analysing the differences between the group means – were used to find the best classification of each variable with regard to the statistical differences in learning outcomes. At the second phase, eleven variables comprising the home possessions and three variables comprising the home accessories were summed up and dichotomized based on DTA and ANOVA. At the third phase, all seven variables for SES were dichotomized based on DTA and ANOVA. Hence, all the variables – regardless of their original scale – were scaled as 0 or 1, where 1 indicates the higher SES (and maximization of learning outcomes). This makes all the individual indicators equal weighted. At the final phase, seven indicators were summed up as the final SES indicators.

It is worth noting that the final SES indicators, (1) are strictly geared towards educational outcomes (and not health, for example), (2) are balanced with education and occupation though somehow over-representation of the economic dimension (3 indicators), (3) have moderately high reliability (0.65) – indicating that it can separate, at least, the extremes quite nicely, and (4) can be changed when the society changes – this indicator reflects society at the end of the fiscal year 2012.

2.6 Criterion-based Assessment in Nepali and English and CEFR Levels

An additional possibility in the Nepali and English language test was to use the criterion based assessment. In the testing of Mathematics, one is bound to use norm-referenced testing because no such internationally accepted general criteria are formed which would be the basis of assessing the real proficiency level of Mathematics. Instead, the final testing in Mathematics produces a norm with which the different groups (such as Ecological zones, developmental regions, or gender) are compared. Thus, one may get to know that in a certain Ecological zone the results are better than in another zone. However, one does not know how good the pupils are, that is, what is the real achievement level; it may appear that all the students in the population are good or poor though there still may be significant differences between the groups. In Mathematics, the only external norm that can be usable is the international norm coming from the large international population (TIMSS and PISA datasets).

Contrary to the situation in Mathematics, in the Nepali and English languages it is possible to apply criterion-based assessment, because in the language testing processes there are several common frameworks for setting the standards. One of those, the Common European Framework of Reference for Languages (CEFR) was selected for the basis of the standard setting because the standards and procedures are well described in the literature (for example, FNBE, 2004; Takala, 2009; Kaftandjieva, 2009; Mitzel et al., 2001; Van der Schoot, 2009) and the levels are transformable into other standards(see http://en.wikipedia.org/wiki/ Common_European_Framework_of_Reference_for_Languages). One advantage of the CEFR classification is that there is a connection of CEFR standards with other standards such as TOEFL. Thus, there are two external criteria available to assess what is the general language proficiency level in Nepal: the international reading test (PISA scale) and CEFR level. The background of the CEFR levels, the method used for the standard setting (Metsämuuronen, 2013) and its application in Nepal are handled in detail in NASA 2011 report (ERO, 2013). Here only the categories are given.

In NASA 2012 as well in NASA 2011, an adaptation of CEFR, more precise than the original, is used. The adaptation was prepared in the Finnish National Board of Education (FNBE) for assessing the language proficiency in school. The original scale is categorized into five groups – A_1 level, elementary proficiency level (limited communication in the most familiar situation); A_2 level, first stage of basic proficiency (basic needs for immediate social interaction and brief narration); B_1 level, functional basic proficiency (dealing with everyday life); B_2 level, first stage of independent proficiency (managing regular interaction with native speaker); C_1 level, first stage of fluent proficiency (managing in a variety of demanding language use situations). In the adaptation of FNBE, every level is further classified into two to three sub-levels (see table 2.21).

Table 2.21 Abridged adaptation of the CEFR levels in the FNBE (FNBE, 2004)

CEFR level	Reading comprehension	Writing comprehension
A1	Elementary proficiency level: limited communication in the most familiar situation	
A1.1	• Is familiar with the alphabet, but understands little of the text. • Recognizes a small number of familiar words and short phrases and can tie these in with pictures.	• Can communicate immediate needs using very brief expressions. • Can write the language's alphabets and numbers in letters, write down his/her basic personal details and write some familiar words and phrases.
A1.2	• Can understand names, signs and other very short and simple texts related to immediate needs. • Can identify specific information in simple text, provided he/she can reread it as required.	• Can communicate immediate needs in brief sentences. • Can write a few sentences and phrases about him/herself and his/her immediate circle (such as answers to questions or notes).
A1.3	• Can read familiar and some unfamiliar words. Can understand very short messages dealing with everyday life and routine events or giving simple instructions. • Can locate specific information required in a short text (postcards, weather forecasts).	• Can manage to write in the most familiar, easily predictable situations related to everyday needs and experiences. • Can write simple messages (simple postcards, personal details, simple dictation).
A2	First stage of basic proficiency: Basic needs for immediate social interaction and brief narration	
A2.1	• Can understand simple texts containing the most common vocabulary (personal letters, brief news items, everyday user instructions). • Can understand the main points and some details of a few paragraphs of text. Can locate and compare specific information and can draw very simple inferences based on context.	• Can manage in the routine everyday situations in writing. • Can write brief, simple messages (personal letters, notes), which are related to everyday needs, and simple, enumerated descriptions of very familiar topics (real or imaginary people, events, personal or family plans).
A2.2	• Can understand the main points and some details of messages consisting of a few paragraphs in fairly demanding everyday contexts (advertisements, letters, menus, timetables) and factual texts (user instructions, brief news items). • Can acquire easily predictable new information about familiar topics from a few paragraphs of clearly structured text. Can infer meanings of unfamiliar words based on their form and context.	• Can manage in routine everyday situations in writing. • Can write a very short, simple description of events, past actions and personal experiences or everyday things in his/her living environment (brief letters, notes, applications, telephone messages).
B1	Functional basic proficiency: Dealing with everyday life	

CEFR level	Reading comprehension	Writing comprehension
B1.1	• Can read a few pages of a wide variety of texts about familiar topics (tables, calendars, course programmes, cookery books), following the main points, key words and important details even without preparation. • Can follow the main points, key words and important details of a few pages of text dealing with a familiar topic.	• Can write an intelligible text about familiar, factual or imaginary topics of personal interest, also conveying some detailed everyday information. • Can write a clearly formulated cohesive text by connecting isolated phrases to create longer sequences (letters, descriptions, stories, telephone messages). Can effectively communicate familiar information in the most common forms of written communication.
B1.2	• Can read a few paragraphs of text about many different topics (newspaper articles, brochures, user instructions, simple literature) and can handle texts requiring some inference in practical situations of personal relevance. • Can locate and combine information from several texts consisting of a few pages in order to complete a specific task.	• Can write personal and even more public messages, describing news and expressing his/her thoughts about familiar abstract and cultural topics, such as music or films. • Can write a few paragraphs of structured text (lecture notes, brief summaries and accounts based on a clear discussion or presentation).

B2	First stage of independent proficiency: Managing regular interaction with native speaker	
B2.1	• Can read a few pages of text independently (newspaper articles, short stories, popular fiction and non-fiction, reports and detailed instructions) about his/her own field or general topics. Texts may deal with abstract, conceptual or vocational subjects and contain facts, attitudes and opinions. • Can identify the meaning of a text and its writer and locate several different details in a long text. Can quickly identify the content of text and the relevance of new information to decide whether closer study is worthwhile.	• Can write clear and detailed texts about a variety of areas of personal interest and about familiar abstract topics, and routine factual messages and more formal social messages (reviews, business letters, instructions, applications, summaries). • Can express information and views effectively in writing and comment on those of others. Can combine or summarize information from different sources in his/her own texts.
B2.2	• Can read independently several pages of complex text written for a variety of purposes (daily newspapers, short stories, novels). Some of these may be unfamiliar or only partially familiar, but deal with areas of personal relevance. • Can identify the writer's attitudes and the function of the text. Can locate and combine several abstract details in complex texts. Can understand enough to summarize or paraphrase the main points.	• Can write clear, detailed, formal and informal texts about complex real or imaginary events and experiences, mostly for familiar and sometimes unfamiliar readers. Can write an essay, a formal or informal report, take notes for future reference and produce summaries. • Can write a clear and well-structured text, express his/her point of view, develop arguments systematically, analyze, reflect on and summarize information and thoughts.

CEFR level	Reading comprehension	Writing comprehension
C1	First stage of fluent proficiency: Managing in a variety of demanding language use situations	
C1	• Can understand lengthy and complex texts from a variety of fields in detail. • Can adapt his/her style of reading as appropriate. Can read critically, assessing stylistic nuances, and identify the writer's attitudes and implicit meanings in the text. Can locate and combine several abstract details in complex texts, summarize these and draw demanding conclusions from these.	• Can write clear, well-structured texts about complex subjects and express him/herself precisely, taking the recipient into account. Can write about factual and fictional subjects in an assured, personal style, using language flexibly and diversely. Can write clear and extensive reports even on demanding topics. • Shows command of a wide range of organizational means and cohesive devices.
C2	Mastery or proficiency: understand with ease virtually everything heard or read.	
C2	• (no description in FNBE) • Can understand with ease virtually everything heard or read. Can summarize information from different spoken and written sources, reconstructing arguments and accounts in a coherent presentation. Can express him/herself spontaneously, very fluently and precisely, differentiating finer shades of meaning even in the most complex situations.	

2.7 Statistical Methods Used in Analysis

While analysing the results, some statistical tools and concepts have been used which are mainly the mean, standard deviation, percentage, frequency and so on. Similarly, statistically significance, effect size and explanatory power are the concepts. Each of them is described here follow.

2.7.1 Analytical tools used in the statistical analysis

The basic tools of statistical description (means, standard deviations, percentages, and frequencies), correlations (Pearson's product moment correlation coefficient), and comparison of two means (t-test) as well as statistical inference (p-values, effect sizes) are used when appropriate in the analysis and reporting. These methods are described in all standard textbooks of statistical description and inference (e.g. Metsämuuronen, 2013). Analyses are done using SPSS programme. The Analysis of Variance (ANOVA) is used in the General Linear Modelling (GLM) when several means are compared. All the p-values are corrected by using Multilevel modelling (or Hierarchical Linear modelling) by using SPSS Linear Mixed models module when they are not obvious (as $p < 0.001$ is obviously significant with or without correction).

Somehow more exotic method or a set of methods called Decision Tree Analysis (DTA) are used in some cases when willing to find the best predictors of the achievement out of hundreds of possible meaningful variables. DTA is one of the methods used in data mining which is very effective when it comes to finding statistically the best groupings of the independent variables. DTA produces a chart such as the one below:

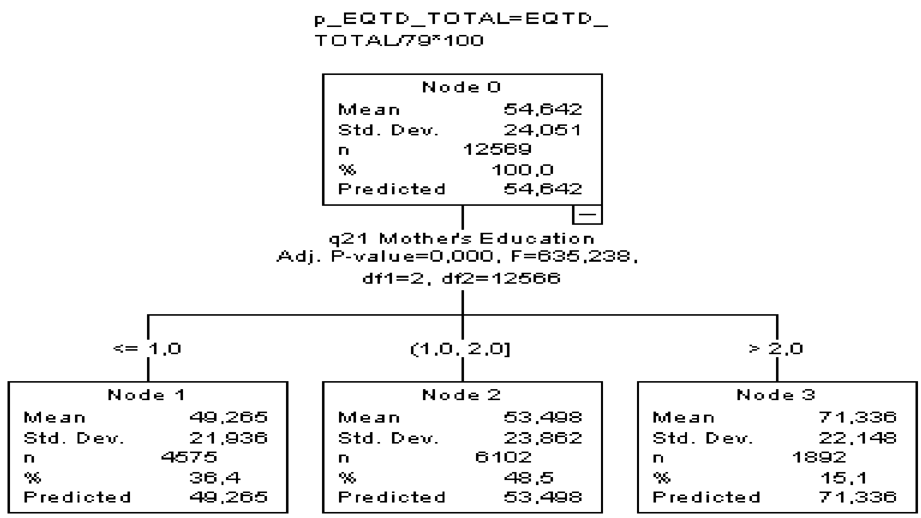

Figure 2.3 An example of DTA modelling in English

The chart shows that the original "mother node", node 0, where the students average achievement is 54.6% of the maximum score, can be divided into three categories of mothers' education which all differ from each other in a statistically significant manner ($F(2, 12566) = 635.24$, $p < 0.001$). Just below the node 0, there are indicators for the statistical test: the *p*-values are adjusted by using Bonferroni adjustment (Adj. *P*-value) and the test used was the *F*-test with the degrees of freedom of 2 (which is the number of groups minus 1) and 12566 (which is N minus the number of groups). In node 1 (here "< = 1", that is, the lowest class of mother's education, that is, "illiterate"), the students' achievement level is 49.3 and in the highest group (node 3, here "> 2.0", that is, "grade 10-passed mothers or higher") and the average students' achievement level is 71.3. In the boxes, there is also information on standard deviation (Std. Dev.), sample size (n), the percentage of cases that are in this node (%), and the predicted mean in the node (predicted) which appears to be the same as the mean.

2.7.2 Some statistical concepts used in the text

Within the text, three important concepts have been used: Statistical Significance, Effect Size and Explanatory Power of the variable. **Statistical significance** is the *p*-value which refers to the possibilities to generalize the result to the population. Behind the *p*-value (from "probability") is the fact that there is always a measurement error while measuring human mental processes. 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 as well as 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 $p < 0.001$, this means that the difference would be found at risk of less than 0.1% – only less than one sample out of 1000 samples from the same population, the results differ from those obtained. If the p-value is $p = 0.002$, the risk for a faulty decision (or difference) is 0.2%.

When the sample size is huge – the sample of 19,000 students, for instance – the *p*-value easily gives a signal that the difference between the groups is real in population. In this context, *p*-value does not tell whether the difference is small or big. For this purpose, there is another statistical concept, which is Effect Size (ES). **Effect size** indicates how far the lowest and highest groups are from each other. The especially used indicators of ES are Cohen's *d* for two means and Cohen's *f* for several means (Cohen, 1988). Cohen has given boundaries for small, medium and large effect sizes. 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 large effect sizes are collected in table 2.22.

Table 2.22 Rough boundaries of Effect Sizes

Size	**Cohen's *d***	**Cohen's *f***
Small	< 0.2	< 0.1
Medium	Round 0.4	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 as "the difference between boys and girls is statistically significant at $p < 0.001$, 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 ANOVA as an analytic tool, the output allows 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 actually is a correlation coefficient 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 $f = \sqrt{(0.3/0.7)} = \sqrt{0.43} = 0.65$ showing high effect size (see Cohen, 1988, p. 284).

Chapter 3: Assessment Results in Mathematics

Mathematics as a school subject has been assessed systematically and frequently in the National Assessments of Student Achievement (NASA) in Nepal. In the assessment of 2011 (see ERO, 2013), the grade 8 students were assessed, now the grade 3 and 5 students are in focus. The frequent assessment is motivated by the fact that the value of mathematical skills is ranked high in the modern society. The “Arithmetic skills” in Mathematics, that is, Numeracy are utmost important in the societies where an increasing number of information is given in a numeric form, in tables, graphs and plots, as well as in strict numbers – nothing to tell about the required ablity to count percentages and value of money in the everyday life. It is expected that the modern citizens should be able to handle such information to survive in the information flow.

Mathematics proficiency at grade 3 has been assessed once in grade 3 and couple of times in grade 5 within the national assessments of student achievement (NASA) in Nepal. The results of the previous national assessment (see BPEP, 1995 & 1997; 1998; EDSC, 1997;1999; 2001; 2003; 2008; PEDP, 1998; CERID, 1999; CERSOD, 2001; Fulbright, 2008) are not fully comparable with NASA 2012 because of the missing linking procedure between the tests. Though the proficiency levels are not comparable with others National Assessment results in the absolute sense (as, for example, percentages of correct answers are not), the proportional differences between the groups are compared in what follows. Besides, there is also a possibility of comparing grade 3 and 5 results because of some linking items from international TIMSS test used in both the grades.

Assessment results in Mathematics are based on the achievement test conducted among 19,252 students of grade 3 in 841 sample schools and among 13794 students of grade 5 in 557 sample schools from 28 sample districts. Sample schools represent all Ecological zones and Development regions, rural and urban areas as well as community and institutional schools. Basic and disaggregated results of assessment based on various strata and diversity factors are included in the analysis. Besides, the extent to which a number of related factors influenced student achievement has been scrutinized. In this chapter, analysis of results for each grade are presented in 3.1 and 3.2 separately and the findings of of both the grades are summarized together in a separate section 3.3.

3.1 Assessment Results in Mathematics for Grade 3

This section analyses the assessment results for grade 3 Mathematics. It starts with analyzing basic results including overall distribution of scores, results in the different content areas in general, and goes to describe the effects of different diversity factors from equality point of view. It then analyses the influencing factors explaining the differences in the achievement of Mathematics

3.1.1 Basic Results of Assessment in Mathematics for Grade 3

As the basic results of assessment in Mathematics, this sub-section analyses the overall distribution of scores, results in the various content areas and various levels of cognitive domains, result variations based on item types, and comparison of results with previous assessments and international test results.

Distribution of overall scores

When assessing student achievement, the population is usually normally distributed in a large sample. The 3rd grade Mathematics sample is big enough to form the normal distribution. However, figure 3.1.1 shows that the total score is not distributed normally. There are clearly distinctive clusters of sample populations in the dataset: low and high performing students.

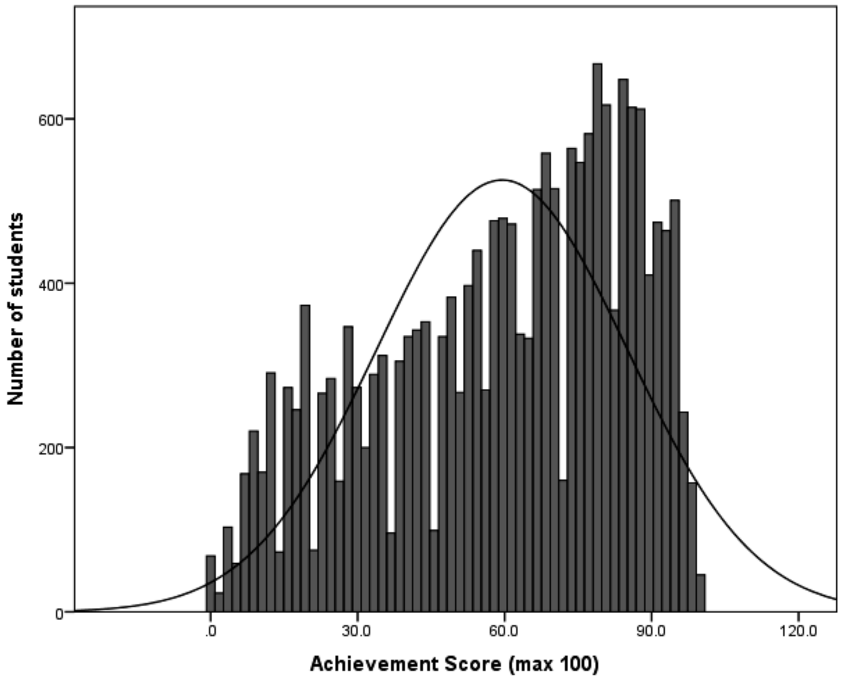

Figure 3.1.1 Distribution of achievement scores in Grade 3

In figure 3.1.2, the left-hand side distribution shows the achievement of community school students and the distribution on the right-hand side shows that of the institutional school students. The main system is shifted to the lower performing level because the main population comes from the community schools. The main population in the institutional schools performs very well; the selection of the students may explain the high result. There are quite a number of the students in community schools getting equally high marks as gained by students in institutional schools. On the basis of figure 3.1.2 it is evident that the students in community schools vary from low-performering to the highest performing, whereas most of the students from institutional schools are performing relatively high.

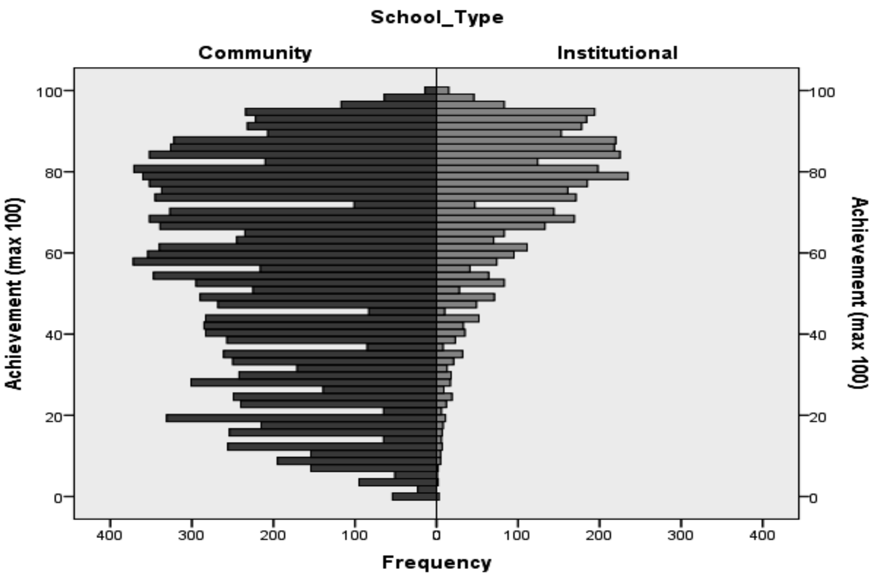

Figure 3.1.2 Distribution of the students'achievement scores by school type

Figure 3.1.2 shows the distribution of the main score of schools. Both populations are slightly skewed: community schools are skewed to include more low-performing schools and institutional schools are skewed to include more high-performing schools. On the basis of school mean score of the student performance, there are two categories of schools. One (wide) population seen on the upper histogram in figure 3.1.2 shows the average of around 50–55% score and the other on the right-hand side (the lower histogram) with the mean of around 75%. One may see three sub-populations in the histogram for the institutional schools: (1) very high performing schools (the average around 90%), (2) high performing schools (around 70%), and (3) medium performing schools (around 50%). The difference between the populations is remarkable.

By analysing the data further with the scatter plot, and combining the socio-economic status (SES) with the average achievement in school, figure 3.1.3 shows that two types of schools fall into two groups: a) institutional schools (triangle) are mostly performing very well and the average SES is also very high, and b) community schools (circle) vary from a very high performing to a very low-performing.

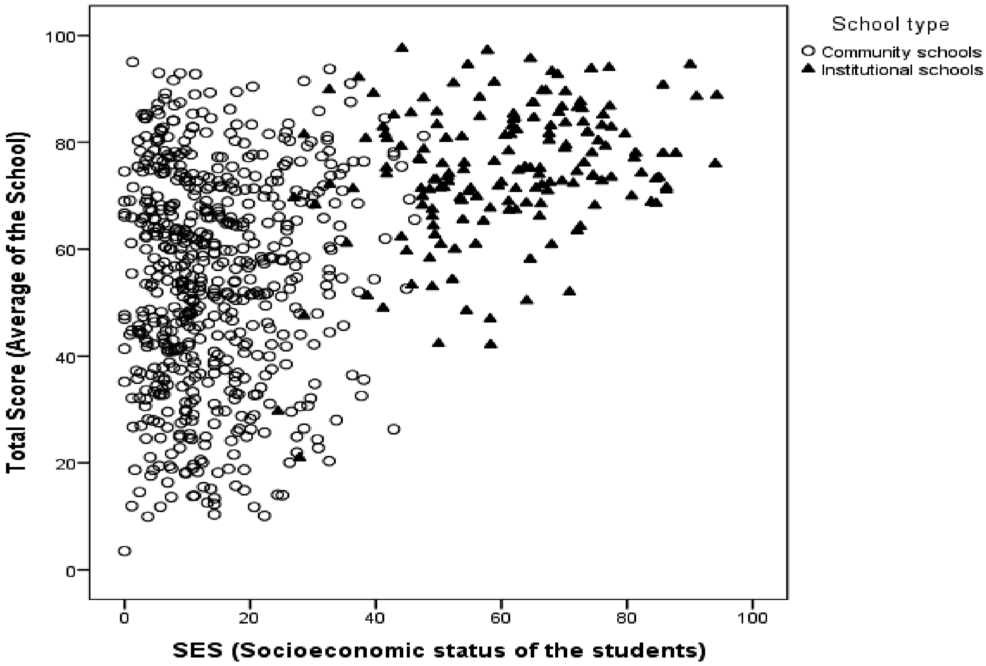

Figure 3.1.3 Relation between achievement and socio-economic status by school type

The dataset is evident that the grade 3 population in mathematics is not distributed normally. There are three distinctive student populations: low medium and high performing students from community schools and almost high performing students from institutional schools. The variation between the community schools in their performance is remarkable.

Achievement in various content areas in Mathematics

The whole Mathematics test was a combination of four content areas: 1) Arithmetic, 2) Algebra, 3) Geometry, and 4) Numeracy. In the curriculum, more weight is given to arithmetic skills compared to the others. Figure 3.1.4 shows the students' achievement in Mathematics as a whole and the achievement level in four content areas.

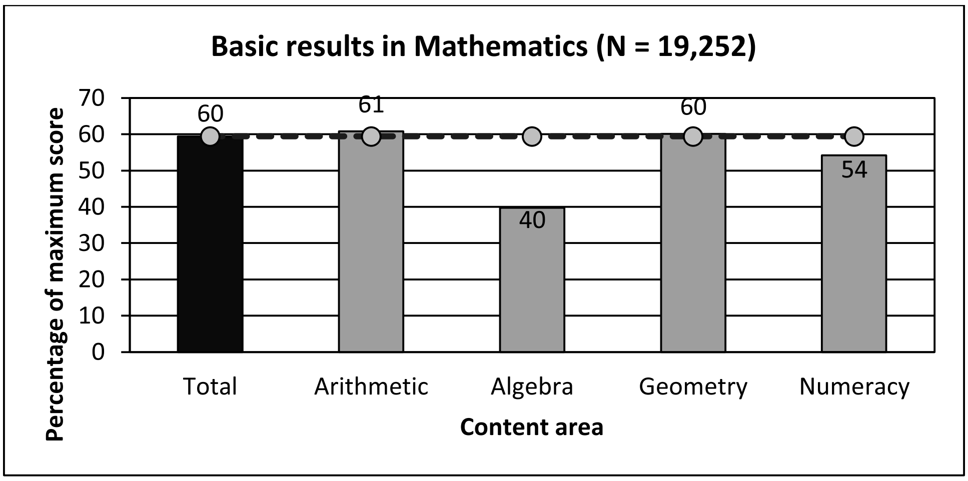

Figure 3.1.4 Comparison of score in various content areas

The percentage of achievement score shows that the national average of Mathematics is 60%. Among the different content areas, students are remarkably weaker than the average in Algebra (40%) and Numeracy (54%). They perform better in Arithmetic (61%) and at an average level in Geometry (60%). While looking at the range of the scores of students from the database, students' scores range from 0 to 100.

Because of the difference in achievement between community and institutional schools, it is interesting to know whether there is proportional difference in the content areas between the students. Table 3.1.1 illustrates the differences.

Table 3.1.1 Achievement in various content areas by the school type

Content area	Community schools (N = 14,476)			Institutional schools (N = 4,775)		
	Mean	**SD**	**CV**	**Mean**	**SD**	**CV**
Arithmetic	55.9	25.5	45.7	75.8	17.7	23.4
Algebra	34.4	33.2	96.5	55.8	32.7	58.6
Geometry	55.1	33.7	61.1	75.3	25.9	34.3
Numeracy	49.2	30.5	61.9	69.4	24.0	34.6
Total[1]	**54.4**	**25.8**	**47.4**	**74.7**	**17.9**	**24.0**

1) *Note that the total score mean is not the mean of the content areas because it has been equated independently from the separate content areas.*

The profile of achievement in different content areas seems to be of the same kind in community schools compared to the institutional schools. In both types of schools, the level in Algebra is about 20 percent lower than the level in Arithmetic, and the same proportions are seen in all other content areas.

The dataset indicates that the learning achievements are the lowest in the content areas of Algebra (40%) and Numeracy (54%), the highest in Arithmetic (61%), and at the range of national average in Geometry (60%). The achievement in Algebra is remarkably lower compared to Arithmetic and Geometry. The differences between the content areas are similar in community and institutional schools.

Achievement in various cognitive domains

Mathematics test as a whole was constructed based on Bloom's taxonomy for cognitive domain (Bloom *et al.,* 1956; Metfesser, Michael & Kirsner, 1969) – that is, *knowledge*, *comprehension*, *application*, and *higher ability* (reasoning/problem solving). Because the number of items requiring higher ability, for the IRT modeling purposes, the classes of application and higher ability were combined. The achievement of the students at the various cognitive levels is shown in figure 3.1.5.

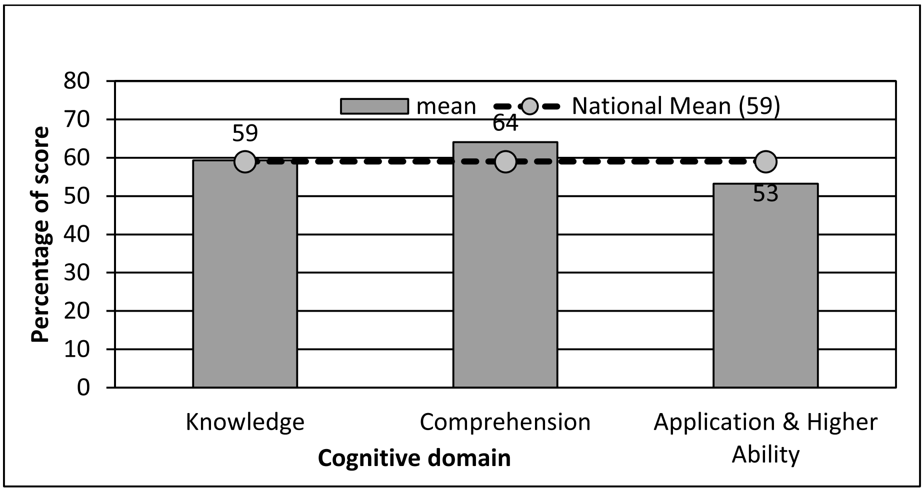

Figure 3.1.5 Comparison of achievement at various cognitive domains

Compared to the grade 5 dataset (see section 3.2) as well as the other subjects, it is noteworthy that the grade 3 students are better in the comprehension type of items than in the knowledge type of items. The technical reason for this is that quite many comprehension type of items were very easy; some of the items (like comparison of higher and lower value of the numbers given) could be categorized as knowledge type. Remarkably high number of students (13%) were able to solve only 15% or less practical problems, that is, the application (and higher ability) type of items. Around 4% of the students are unable to solve even any of these tasks.

The dataset informs that the students' ability to solve practical problems is quite low (53%). Students are good in recalling type of questions (59%) and comprehension (64%). Notable number of students (4%) were not able to solve any of the tasks requiring the ability of applying the knowledge in a novel situation. Students in institutional schools are found to be more able to solve practical and complex problems than their peers in community schools.

Type of item and achievement

There were basically 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 of a question, or one explicit piece of information was needed to get the correct answer. 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 made on each of the hierarchial levels of cognitive domain (knowledge, comprehension, application, and higher ability) and a wide range of difficulty levels, though the subjectively scored items tend to be more demanding because of the higher level of cognitive domain. Table 3.1.2 presents the basic statistics of the item type-wise achievement levels.

Table 3.1.2 Mean score by item type

Type of items	Mean	SD	Minimum	Maximum
Objective	65.5	23.9	0	100
Subjective	54.4	30.3	0	100

It is obvious that the subjectively scored tasks – usually those with more demanding requirements for the correct answer – are solved much lower (54%) than the objective items (65%). Most of the objective items wereof knowledge and comprehension level, in which average score was higher (see, fig 3.1.4) whereas subjective items were application and higher ability type. It is notable that in many cases the students did not even start attempting the open-ended questions; on an average, 20% of the students give the missing data in these items – in some items, over 30% of the students; hence, the lower mean.

Dataset shows that the students are performing well in objective type of items (65.5%) which were mostly recalling and comprehending type. Students weaker subjective types of items (54.4%), which were mostly productive type. In many cases, the students did not even start to answer the open ended questions and, hence, the lower score.

Comparison of achievements with previous datasets

The national assessment carried out in various years aims to assess the change in the achievement level and the progress over a period of the years. The datasets of previous Mathematics assessment for grade 3 are, however, somehow sparse (see, BPEP, 1995). The previous datasets also carry two challenges hindering the comparison with the present dataset. First, the National Assessment of grade 3 students carried out by the Basic and Primary Education Project (BPEP, 1995:14) shows that the national average of the students was 38%. This National Assessment in 2012 shows that the national average of Mathematics in grade 3 is 60%. These figures are coming from Classical Test Theory and they are not comparable with each other because of the lack of a proper linking procedure. The differences between the scores can easily be explained by different difficulty levels of the test. Second, the previous datasets of grade 3 are not available and, hence, any IRT modeling-based procedures for comparison could not be made.

Though the comparison cannot be made in absolute sense, proportional comparisons can be made, with caution, on the basis of the previous results. The proportional differences are collected in table 3.1.3.

Table 3.1.3 Comparison of scores of 1995 and 2012 in Grade 3

Mean Level	1995(BPEP, 1995)	NASA 2012
	% of cases	% of cases
ABOVE the mean	44.4	54.1
AT the mean	4.0	2.5
BELOW the mean	51.6	43.5

A rough comparison of distributions of 1995 and 2012 datasets shows that in both years the distributions are not normal.

Table 3.1.4 Comparison of achievement scores of 1995 and 2012 by gender

Indicators	1995 (BPEP, 1995)		NASA 2012	
	Boys	Girls	Boys	Girls
Mean	15.0[1]	14.9[1]	59.5[2]	59.5[2]
SD	8.28	7.65	25.4	25.1
CV	55.1	51.2	42.7	42.2
N	1,038	824	8,304	8,670
T	0.257		0.036	
Sig.	n.s.		n.s.	
Cohen's d	0.00		0.00	

1)Raw scores, maximum 40 2)Percentages of the maximum marks, maximum 100,
3) n.s.= not significant

Compared to the 1995 dataset, the difference between the boys and girls has not changed radically; the difference has disappeared (non-significant) in 2012; there is no difference between boys and girls.

Table 3.1.5 Comparison of achievement scores of 1995 and 2012 by Ecological zones

	1995 (BPEP, 1995)			NASA 2012		
	Mountain	Hill[1]	Tarai	Mountain	Hill[1]	Tarai
Mean	16.7[2]	16.3[2]	14.0[2]	60.1[3]	55.4[3]	54.7[3]
SD	7.2[4]	8.4[4]	7.8[4]	25.1	25.8	25.4
CV	42.9	51.6	55.7	41.8	46.5	46.5
N	666	681	1,184	2,001	8,368	5,550

1) Students from Kathmandu Valley are excluded 2)Raw scores, maximum 40, weighted mean
3) Percentages of the maximum marks 4)Weighted mean

Compared to the 1995 dataset, the difference between the Ecological zones is found to have expanded slightly. The students in the Mountain zone have raised their achievement in comparison with the Hill and Tarai zone (no difference in 1995; notably higher result in 2012). One may note, though, that there was only one district from Mountain zone in 1995.

Table 3.1.6 Comparison of achievement of 1995 and 2012 by Development regions

	1995 (BPEP, 1995)					NASA 2012[1]				
	Eastern	Central	Western	Mid-Western	Far-Western	Eastern	Central	Western	Mid-Western	Far-Western
Mean	12.6[2]	16.6[2]	21.8[2]	15.4[2]	14.0[2]	51.9[3]	56.3[3]	63.9[3]	45.7[3]	58.0[3]
SD	7.7[4]	8.0[4]	9.3[4]	7.4[4]	7.8[4]	26.1	25.2	22.3	25.1	26.1
CV	60.6	48.5	42.5	47.8	56.1	50.3	44.8	34.9	54.9	44.9
N	652	295	275	964	345	2910	4603	2982	2291	3133

1)Students from Kathmandu Valley are excluded. 2) Raw scores, maximum 40, weighted mean.
3) Percentages of the maximum marks, maximum 100. 4)Weighted mean

While comparing the Development regions, two things should be kept in mind: First, only one district had represented Central and Western Development regions in 1995, and second, the Valley was not included in the 1999 sampling. However, while taking the datasets comparable, changing the raw scores to the percentages of the total score (31.6, 41.4, 54.5, 38.4, and 34.9 for the regions respectively), and keeping the Western region (with the highest score) as a reference point for comparison; it is notable that, according to the 1995 dataset, the students from the Far-Western region had performed 20 percent lower than the students from the Western region; in 2012 the difference is only one third of that; i.e, 6 percent. The effect size has decreased from $d = 0.96$ to $d = 0.28$ showing that the difference has narrowed remarkably. The same happened with Eastern region in comparison with Western region: in 1995 the difference was 23 percent, but in 2012 it is half of that; 12 percent. The effect size has decreased from $d = 1.04$ to $d = 0.43$ showing that the difference has reduced from high to moderate. These signs are positive from equality point of view. On the other hand, the variance has increased in the Mid-Western region at the same time as the difference to the Western region has increased. This is not a good sign from the educational equality point of view.

Table 3.1.7 Comparison of achievement of 1995 and 2012 by caste/ethnicity

Caste/Ethnicity	1995 (BPEP, 1995)			NASA 2012[1]		
	N	Mean (%)	Difference[2]	N	Mean (%)	Difference[2]
Brahman and Chhetri	909	38.1		10191	61.4	
Newari	53	42.6	-4.5	656	52.6	4.9
Gurung/Magar/Tamang	326	37.6	0.6	718[3]	73.7	-13.7
Tharu	107	38.4	-0.3	901	50.5	7.0
Muslim	6	48.8	-10.6	725	58.6	3.9
Others	440	35.3	2.8	2383[3]	59.1	-0.8

1) Students from Kathmandu Valley are excluded 2)Difference in comparison with the Nepali-speaking group 3) Combined

Compared to the 1995 dataset by the ethnicity of the student from table 3.1.7 that the Nepali speaking majority and especially the Gurung/Magar/Tamang speaking minority have raised radically their position than the other ethnic groups. While, in the 1995 dataset, both the groups were at the same level and radically lower (38%) than the small‒ Muslim community (49%), the reality has turned opposite with the years. In mathematics at grade 3, the Gurung/Magar/Tamang students are far behind (74%) the other groups (50–61%).[6] The change in the phenomenon is notable. Table 3.1.8 presents the information provided in the above tables.

Table 3.1.8 Summary of comparison of achievement between 2012 in to 1995 datasets

	Selected background variables			
	Sex	**Ecological zone**	**Development region**	**Ethnicity/language**
Main finding	No change in difference; still no difference between the gender	Students in the Mountain zone score higher. Differences have expanded slightly.	Students in the Eastern, Mid-Western and Far-Western regions score higher. Differences have reduced remarkably.	The Nepali speaking majority and especially the Gurung/Magar/ Tamang speaking minorities have raised their position among the other ethnic groups.

[6] If the Valley schools were included in the analysis, the mean would be the same.

Compared to the 1995 results, no change is found between boys and girls. However, students in the Mountain zone and Far-Western region scored higher, and students from Gurung/Magar/Tamang community scored remarkably higher in 2012.

Comparison with the international standard

The NASA 2012 was made comparable with the international TIMSS Mathematics assessment. Four of the released TIMSS items were used as linking items to the international item bank. Items of TIMSS were selected so that they fit with the Nepalese curriculum and items are familiar to the Nepalese students.[7] 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 2012 is zero, it corresponds to the average level of the international students of Grade 4.

Figure 3.1.6 shows the comparison of students' achievement with the international standard. In the figure, the x-axis shows the content areas of mathematics and y-axis shows the ability shown by the students. The middle horizontal line of $\theta = 0.00$ indicates the international average. When 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. Parallel, when the bars are going upwards, it indicates that the ability level is higher than the international mean.

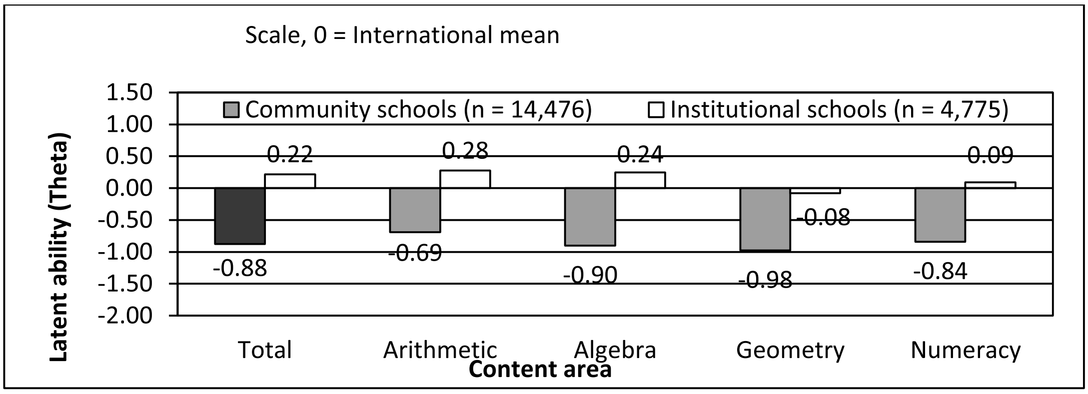

Figure 3.1.6 Comparison of student achievement to TIMSS mean scale

Figure 3.1.6 shows that the average ability shown by the grade 3 students in Mathematics is generally below the international average. This indicates that the students in Nepal score remarkably lower than their international peers (at grade 4). The achievement level of an average grade 3 student in the community schools ($\theta = -0.88$) is very low compared to an average international student of grade 4. The achievement level of an average student in the private schools ($\theta = +0.22$) is much higher than that of his/her peer in the community school and somehow higher than the average in the international item bank. It is good to remember three things. First, all the linking items came from the content area of Numeracy, Arithmetic and Geometry and hence there actually is no real equating in Algebra. Second,

[7] This principle is the same as used in NASA 2011 with grade 8 (ERO, 2013). This causes, most probably, that the results are better than if selected the items randomly from the international item bank (see discussion in NASA 2011 report).

all the linking items were selected to fit the Nepalese curriculum and hence the real achievement level is most probably lower than seen in figure 3.1.6[8]. Third, the difficulty level of the items (suitable for 4th graders) was not the best suiting to the 3rd graders' Mathematics proficiency. Hence, the lower proficiency is expected. From that point of view, the high result in institutional schools is somehow surprising.

While comparing the results to international assessment, it is evident from the dataset that the average Mathematics proficiency of 3rd graders in Nepal is lower than the international average in TIMSS standard. The students in institutional schools are at the middle or even higher level in achievement than international average.

3.1.2 Assessment Results Based on Diversity Factors

Diversity is a relative and contextual term. In the context of Nepal, some of the experts have observerd eight diversities: geographical/ecological, linguistic, gender/sex–related, religious, ethnic, cultural, disability and economic related diversity (see ERO, 2013). The NASA 2012 background information questionnaire included six of the above diversities; two of these (the cultural and religious background of the students) were not asked. Additionally, however, three other diversities are handled in this section: district-wise, school type-wise (community/institutional), and school location-wise (rural/urban). These factors can be taken as equality factors; as all children regardless of their sex, language, birth place, or family background should have equal opportunities to reach the same educational goals.

District variations in student achievement

It is good to keep in mind that there may be other lower or better performing districts within those not selected in the sample. The district-wise differences are presented in table 3.1.9 and figure 3.1.7. The table shows the achievement in the selected districts in descending order. The mean represents the average achievement percentage of the particular district.

Of the randomly selected districts in the sample, student performance was very low in Udyapur (46%), Khotang (47%), and Saptari (49%) from the Eastern region; in Mahottari (48%) from the Central region; and in Jumla (40%), Rolpa (40%), Bardiya (41%), and Salyan (50%) from the Mid-Western region. Except for Parsa district (73%), the outperforming four districts come from the Central region, particularly the districts are from the Kathmandu Valley: Kathmandu (79%), Bhaktapur (77%), and Lalitpur (72%). Similarly, Kaski (70%), Humla (70%), and Solukhumbu (70%) districts also got high scores. It is worth noticing that Parsa, Humla, Solukhumbu, and Manang are able to achieve a high achievement with a negligible influence from institutional schools, whereas 60 – 68% of the students in the Kathmandu Valley were from the private schools with much higher socio-economic status (see section 3.1.3). Out of the ten lowest performing

[8] The same challenge was seen in 2011 datasets. The high results in mathematics and social studies were caused, most probably, by the familiar type of items. No genuinely new content areas or item types were introduced from the international item bank. This is not challenge in Nepali and English where the linking text and related items were selected from outside the curriculum.

districts, five had no institutional school. From this perspective, interesting districts are those where the number of institutional schools is low but the achievement scores are higher than the national average.

Table 3.1.9 Average achievement score in selected districts

Districts	N	Mean	SD	CV	Districts	N	Mean	SD	CV
Kathmandu	2,042	79.1	16.3	20.6	Myagdi	376	55.2	22.1	40.0
Bhaktapur	558	77.3	15.8	20.4	Kailali	1,193	55.2	26.5	48.0
Parsa	617	72.9	16.9	23.1	Chitwan	831	53.4	23.8	44.5
Lalitpur	733	71.9	17.4	24.2	Dolakha	708	52.3	25.5	48.7
Kaski	1,001	70.3	17.5	24.9	Achham	703	51.2	26.1	50.9
Humla	198	70.3	19.6	27.9	Sindhuli	824	49.8	25.5	51.2
Solukhumbu	397	70.0	19.1	27.3	Salyan	649	49.7	25.4	51.1
Manang	19	67.1	22.3	33.3	Saptari	832	48.7	23.8	48.8
Darchula	495	66.5	23.5	35.3	Mahottari	660	48.0	25.2	52.6
Baitadi	742	63.3	24.4	38.6	Khotang	581	46.9	26.9	57.4
Makwanpur	963	62.5	24.1	38.6	Udayapur	693	45.8	26.9	58.7
Kapilbastu	854	61.8	23.1	37.3	Bardiya	563	41.3	25.0	60.6
Baglung	732	61.8	25.0	40.4	Rolpa	697	40.3	21.8	54.0
Dhankuta	407	58.6	24.9	42.5	Jumla	184	39.6	23.2	58.5
					Total	**19,252**	**59.5**	**25.6**	**43.1**

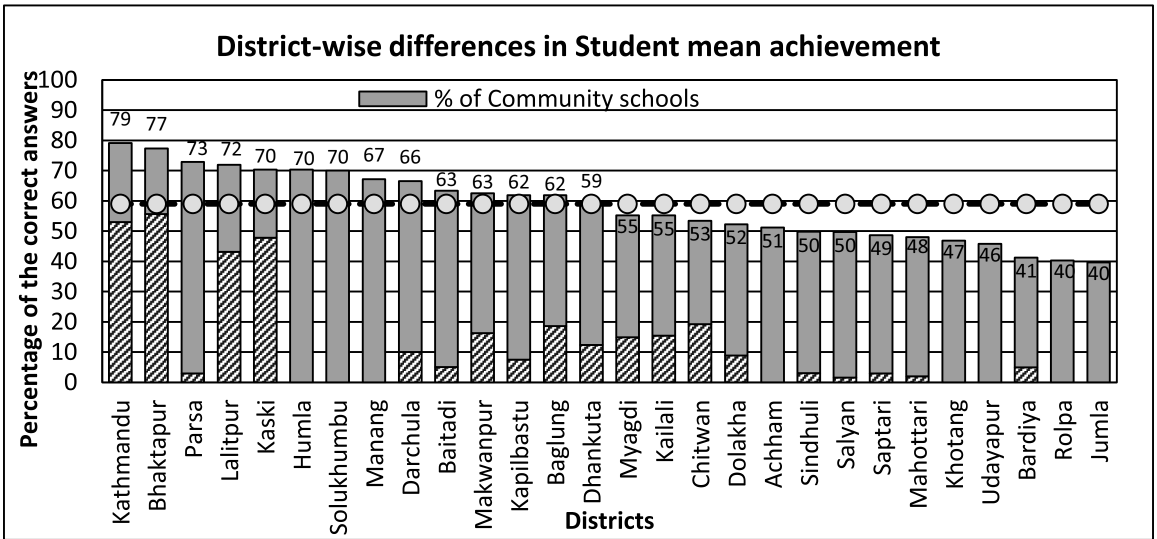

Figure 3.1.7 Average achievement in the sample districts by school types

The difference in achievement due to the district is statistically significant ($p < 0.001$). The variation explained in achievement due to the district is $\eta^2 = 0.208$, that is, the district explains 21% of the variation in the data which is quite high. Effect size is $f = 0.51$ – indicating that the difference between the lowest performing district (40%) and highest performing district (79%) is remarkably high.

The dataset suggests that there is a wide difference between the districts when it comes to the equal opportunity for children to reach the pre-set goals in Mathematics. The results are bound to the 28 districts selected randomly in the sample; even lower-performing districts could be found if other districts had been selected. The results are very high in the districts where the proportion of institutional schools with high socio-economic status is high.

Ecological zone and student achievement

The Mountain, Hill and Tarai are three geographical features in Nepal though the Valley is taken as a special geographical feature because of being the most densely populated area in the country with more opportunities than other areas. Not only from the population point of view, also the mixed ethnicities, favourable climatic conditions, concentration of economic activities, highest development index, as well as the dense human population make the Valley a unique geographical area in the analysis. The variation in the Ecological zones in NASA 2012 is presented in table 3.1.10.

Table 3.1.10 Achievement in various Ecological zones

	Community schools				Institutional schools			
Ecological zone	**N**	**Mean**	**SD**	**CV**	**N**	**Mean**	**SD**	**CV**
Mountain	1,808	60.8	24.7	40.6	193	53.0	27.6	52.1
Hill	6,902	51.4	25.7	50.0	1,466	74.2	15.9	21.4
Tarai	4,643	52.1	25.6	49.0	907	67.6	20.3	30.0
Valley	1,123	72.1	19.8	27.5	2,210	79.8	14.2	17.8
Total	**14,476**	**54.4**	**25.8**	**47.4**	**4,776**	**74.7**	**17.9**	**24.0**

The data shows that, first, on average, the students from the Valley outperform the students from other Ecological zones. Second, it is seen (and this differs from the other datasets) that the students in institutional schools in Mountain zone perform much lower (53%) than in the other zones (Terai: 68%, Hill: 74%). The difference in achievement between the Valley students (80%) and Mountain students (53%) is much wider than the Community schools (72% and 61% respectively). One may note also the exceptionally low value for the Coefficient of Variation in the Valley, which in community schools, is about half of that in the other areas. The obvious reason for this is the systematically high score in the Valley compared to the other areas.

The achievement among the zones would have differed significantly in both types of schools ($p < 0.001$) if the Valley was excluded from analysis. Tukey's *post hoc* test tells that, in community schools, there is no difference between Hill and Tarai, but the students from the Mountain zone differ from the students of both Hill ($p < 0.001$) and Tarai zone ($p < 0.001$). In institutional schools, all the zones differ from each other ($p < 0.001$ for all comparisons). Ecological zone explains 5% of the variance in community schools (η^2 =

0.053) and 13% in institutional schools (η^2 = 0.127).[9] As a comparison, we should remember that the district explains more than 20% of the variation. The effect size is f = 0.24 in the community schools and f = 0.38 in the institutional schools, showing moderate difference (or high in the institutional schools) between the highest and lowest performing Ecological zones. The effect sizes are smaller if the Valley is excluded from the analysis (f = 0.12 and f = 0.31 respectively). This means that, in the community schools, the real differences are not remarkable across the Ecological zones, but the Valley differs radically from the other areas. From equality point of view, this can be taken as a possibly good sign.

Dataset indicates that there is a moderate difference in students' performance among the four Ecological zones in both community and institutional schools. Students in the Kathmandu Valley outperform those in other regions.

Development region and student achievement

Students' achievement varies according to the Development regions which are divided into Eastern, Central, Western, Mid-Western, and Far-Western. Additionally, the Kathmandu Valley is taken as the 6th Development region though administratively it falls under the Central Development region. The mean achievements in the Development regions are given in table 3.1.11.

Table 3.1.11 Achievement in various Development regions

Development	**Community schools**				**Institutional schools**			
region	N	Mean	SD	CV	N	Mean	SD	CV
Eastern	2,769	50.6	25.8	51.1	141	78.5	15.5	19.7
Central	3,843	54.9	25.4	46.2	760	63.3	23.2	36.7
Western	1,877	57.7	23.3	40.3	1,105	74.4	15.8	21.2
Mid-Western	2,200	44.9	25.0	55.6	91	65.5	18.8	28.7
Far-Western	2,664	55.8	26.6	47.6	469	70.5	18.6	26.4
Valley	1,123	72.1	19.8	27.5	2,210	79.8	14.2	17.8
Total	**14,476**	**54.4**	**25.8**	**47.4**	**4,776**	**74.7**	**17.9**	**24.0**

The highest performance is found in institutional schools in the Kathmandu Valley (80%) and in the Eastern region (79%). The performance is the lowest in community schools in the Mid-Western (45%) and Eastern (51%) regions; the Mid-Western result shows remarkably low achievement compared to other regions, nothing to say with the Kathmandu Valley. The difference between the regions is statistically significant both in the community and institutional schools ($p < 0.001$). Tukey's *post hoc* test shows that, within the community schools, the average achievement level in the Mid-Western and Eastern regions is significantly lower than in any other region ($p < 0.001$) and in the Valley the achievement is higher than in any other region ($p < 0.001$). There is no difference between Central, Western, and Far-Western regions when it comes to the achievement

9 If the Valley is excluded from the analysis, the values for Eta squared would be 0.015 and 0.089 respectively, that is, explaination of only 1% and 9% variation. The role of the Kathmandu Valley students in the whole national mean is remarkable.

level in Mathematics. In the institutional schools, Tukey's *post hoc* test shows that the students in the Kathmandu Valley outperform those in all other regions. In Eastern ($p < 0.001$), Central and Mid-Western regions the students perform lower than in the other regions ($p < 0.01$).

Development region explains 6% of the variance within the community schools ($\eta^2 = 0.064$) and 11% within the institutional schools ($\eta^2 = 0.114$).[10] This is somehow the same proportion as found in the Ecological zones. One remembers that the district explains more than 20% of the variation, which means that within the Development regions there are lower and higher performing districts. The effect size is $f = 0.26$ in the community schools and $f = 0.36$ in the institutional schools, showing a moderate or wide difference between the highest and lowest performing regions. The effect sizes are moderate if the Valley is excluded from analysis ($f = 0.17$ and $f = 0.27$ respectively). The differences are wider among the Ecological zones than the Development regions.

From the dataset, it is evident that there is inequality across the Development regions in terms of children's level of competency in Mathematics. Especially the wide difference between the community schools in the Valley and in the rest of the country (27 percent as the highest) is a strong sign of inequality in learning Mathematics. There are also wide differences between the regions in institutional schools; the difference in student performance in the private schools between the Valley and Central region is the highest, i.e,17 percent.

School type and student achievement

All the schools are categorized into community and institutional (that is, private schools). The differences in Mathematics achievement have been handled in the sections above. Here the main differences are presented in table 3.1.12.

Table 3.1.12 Type of school and the average achievement

Content areas	Community (N = 14,476)			Institutional (N = 4,776)			Mean difference	Cohen's *d*
	Mean	SD	CV	Mean	SD	CV		
Arithmetic	55.9	25.5	45.7	75.8	17.7	23.4	19.9	0.84
Algebra	34.4	33.2	96.5	55.8	32.7	58.6	21.3	0.64
Geometry	55.1	33.7	61.1	75.3	25.9	34.3	20.2	0.63
Numeracy	49.2	30.5	61.9	69.4	24.0	34.6	20.2	0.70
Total	**54.4**	**25.8**	**47.4**	**74.7**	**17.9**	**24.0**	**20.3**	**0.84**

The achievement levels in community schools and institutional schools differ from each other remarkably as presented above. The average performance in the total score in private schools is 75% whereas, in community schools, it is 54% with a 21 percent difference,

[10] If the Valley is excluded from analysis, the values for Eta squared would be 0.031 and 0.074 respectively, that is, only 3% and 7% explanation – one third and half of those with the Valley included in the analysis. The role of the Kathmandu Valley students in the whole national mean is remarkable.

which is remarkable. The difference is statistically significant (p < 0.001) and the effect size is high (d = 0.84) – showing that the community schools are far below the institutional schools. Difference is the highest in the content area of Arithmetic (d = 0.84) and quite high also in Numeracy (d = 0.70). Division of the students into community and institutional schools explains 11% of the student variation in Arithmetic (η^2 = 0.115) and 8% in Numeracy (η^2 = 0.083). The deviation in the community schools is remarkable ranging from near 0% to near 100%; contrarily, most private schools in the sample show relatively high performance. This may be explained partly by much higher socio-economic status in the institutional schools and strict selection of the students.

One may note the comparatively high value of Coefficient of Variance with the content area of algebra (96.5% and 58.6%). In community schools, this indicates just that the mean value was relatively low in this content area. In institutional schools, the question is more interesting; also the variance is higher compared to the other content areas. This means that, in this content area, the differences are relatively wider in the institutional schools than in community schools.

The dataset reveals that, on average, the students in institutional schools outperform the students in community schools. The difference is highest in the areas of Arithmetic and numeracy. This deviance can be explained partly by much higher socio-economic input into students' life and strict selection of the students in institutional schools.

School location and student achievement

The schools were divided into two groups based on their location: rural and urban. This information was based on the response given by the head teacher though some of the head teachers did not inform about the school location. The achievements of the students in rural and urban schools are presented in table 3.1.13.

Table 3.1.13 Student achievement on the basis of location of school

Location of school	**Community**				**Institutional**			**CV**
	N	Mean	SD	CV	N	Mean	SD	
Rural	12,766	53.5	25.7	48.1	1,787	74.2	18.7	25.2
Urban	1,256	62.6	25.8	41.2	2,819	74.8	17.5	23.4
Mean difference		9.1				0.6		
Cohen's d		-0.35				-0.03		
Total[1]	**14,476**	**54.4**	**25.8**	**47.4**	**4,776**	**74.7**	**17.9**	**24.0**

1) *The total is calculated by using the whole dataset while the Urban and Rural include some missing values.*

In the urban community schools, the achievement level of the students (63%) is 9 percent higher than that of rural community schools (54%). The difference is statistically significant (p < 0.001) and the effect size is moderate (d = 0.35). If the community schools of the Valley are excluded, the score of the urban community schools lowers to 54; the difference (1 percent) is not statistically significant and the effect size is low (d = 0.04). The main difference in the community schools is, hence, caused by the high achievement level of the students in the Valley schools. The division into rural and urban schools explain 1% of the student variation in the community schools (η^2 = 0.010 and excluding the Valley schools it will be less than η^2 = 0.000). The latter is a good sign from equality point of view. Excluding the Valley, there is no difference between rural and urban community schools. In the institutional schools there is no difference between the rural and urban areas in any case.

Data indicates that the students in the urban community schools have gained 9 percent more than the students in the rural areas. Excluding the Valley schools, the difference is practically zero. In institutional schools, there is no difference between the rural and urban areas. From educational equality point of view, this is a good sign.

Home language and student achievement

In the context of Nepal, students' achievement is found to be depending on the language spoken in their homes i.e., the mother tongue of the students. Mother tongue reflects, in many cases, the ethnic background and hence any difference may be taken as a possible source for inequality in society.

On the basis of Mathematics data, 35% of the 3rd graders speak a language other than Nepali as their first language. These "other" languages are quite fragmented; the largest groups in the Mathematics dataset are Tharu (5.8%), Urdu (4.7%), and Newari (4.2%). After dividing the languages into ten groups excluding Nepali, there were still 14.0% of the students classified into the group "other". Because the languages are very fragmented and the Nepali speakers are the majority of the students, for the purpose of the statistical analysis, all the other languages are first grouped into "non-Nepali". The results are presented in tables 3.1.14 and 3.1.15 and illustrated in figure 3.1.8.

Table 3.1.14 Student achievement on the basis of home language

Language group	Community				Institutional			
	N	Mean	SD	CV	N	Mean	SD	CV
Nepali	4,776	56.7	24.7	43.7	607	75.2	19.4	25.8
Non-Nepali	6,924	55.2	24.9	45.0	3,268	74.6	17.1	23.0
Mean difference		1.5				0.7		
Cohen's *d*		0.059				0.039		

When combining all the minor language groups as "Non-Nepali", there is not a notable difference between the language groups in the community schools (1.5 percent points favoring the Nepali speakers). Though the difference is statistically significant (p = 0.002), the effect size is low (d = 0.06). In institutional schools, the difference is less than 1 percent, which is not notable (p = n.s., d = 0.04).

On the basis of the original categorization of the minor languages, the issue looks quite much interesting. It is evident that the Tamang and Maithili speaking students are at quite much higher level in Mathematics than the Nepali students (73% and 68% compared to 55% of community schools). On the one hand, the students from Tharu (48%) and Gurung (44%) background perform much lower than the average. On the other hand, the few Tharu students in the institutional schools perform the highest (83%) and the Magar (82%) and Urdu (80%) students are very close. The Nepali students represent the major population, and hence it includes both the lowest and highest extremes and their result is at the average.

Table 3.1.15 Achievement in the different language goups

Language/ Ethnicity[1]	Community				Institutional			
	N	Mean	SD	CV	N	Mean	SD	CV
Tamang	484	73.1	15.7	21.4	0	-	-	-
Maithili	32	68.0	22.8	33.5	26	74.2	16.9	22.8
Magar	36	63.7	19.9	31.2	168	81.8	11.1	13.6
Rai	121	59.7	25.9	43.4	5	36.8	18.8	51.0
Urdu	685	57.4	23.8	41.4	40	80.1	12.7	15.9
Nepali	6924	55.2	24.9	45.0	3268	74.6	17.1	23.0
Limbu	7	54.1	33.2	61.3	7	58.6	23.8	40.5
Newari	607	51.3	25.2	49.2	49	69.1	20.3	29.3
Tharu	831	47.8	23.1	48.3	70	83.3	10.6	12.7
Gurung	24	43.9	31.4	71.5	6	78.7	22.5	28.6
Other	1948	57.5	25.1	43.6	235	69.9	23.5	33.6

1) *Those language groups in which number of the students was less than 5 are excluded.*

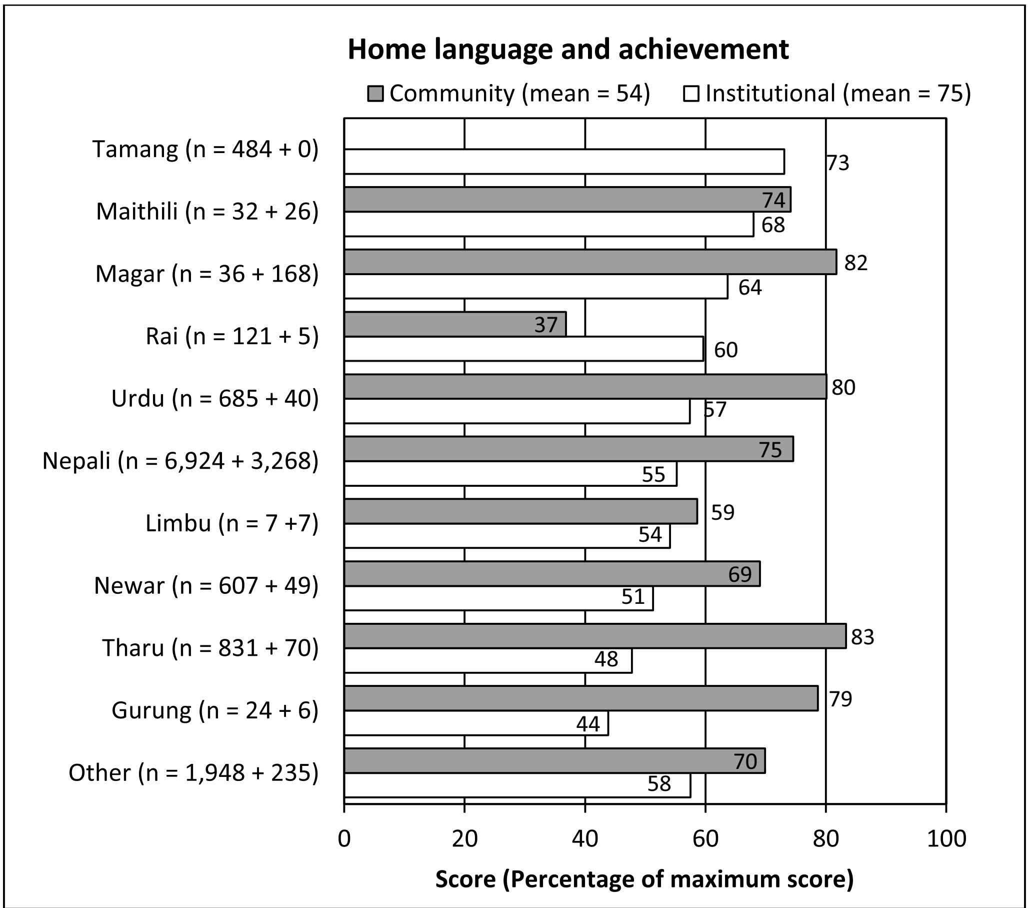

Figure 3.1.8 Relation between language at home and achievement

The difference between the students in the highest and lowest performing language groups are statistically significant ($p < 0.001$) and notable; the effect sizes are moderately high ($f = 0.18$ in community schools and $f = 0.16$ in institutional schools). The division into smaller language groups explains about 3% of the variation in the data ($\eta^2 = 0.032$ in community schools and $\eta^2 = 0.026$in institutional schools). Though the differences are wide between the extreme groups, it is good to keep in mind that the number of students is quite small in some of the language groups –hence, the moderate effect size. When analyzing only the minority languages and hence, excluding the Nepali speakers and the group "Other", the effect size is high ($f = 0.39$) in the community schools – indicating really remarkable difference between the highest performing minority group (Tamang, 73%) and the lowest performing group (Gurung, 44%).

Language and Development region

When combining the results for the Development region and mother tongue, the achievement score of the students within a certain language groups varies remarkebly among the different regions.

Table 3.1.16 Achievement across the different language groups in various regions

Development Region	**Nepali[1]**	**Magar**	**Tharu**	**Newari**	**Urdu**	**Tamang**	**Rai**	**Maithili Awadhi**	**Gurung**	**Limbu**
Eastern	55.3[2]	40.4	52.4	52.9	57.8		59.1		75.4	94.7
Central	59.4	51.3	48.8	52.5	53.1	73.1	54.8	54.8	29.2	41.5
Western	64.8	65.8		65.8	67.1	77.2		72.9	66.2	58.6
Mid-Western	48.0	46.1	61.4	42.7	38.6	59.6			14.0	7.0
Far-Western	56.1		48.2	56.0	87.7				36.1	
Valley	76.3	81.1	78.5	93.9	78.1	52.6	66.2	89.0	77.9	94.7
Total	**61.4**	**78.6**	**50.5**	**52.6**	**58.6**	**73.1**	**58.8**	**70.8**	**50.8**	**56.4**
N[3]	10,192	204	901	656	725	484	126	58	30	14

1) *The language groups of less than 10 students are not included in the table.*
2) *The main population is highlighted by the gray shade. In some nonhighlighted cases there is only one student behind the mean.*
3) *The language groups are ordered on the basis of their frequency.*

All language groups except for Tamang and Rai have a high score in the Valley. Almost all language groups perform lowest in the Mid-Western Development region. Especially, students' performance having mother tounge Gurung and Urdu is low in the Mid-Western region.

Language and Ecological zone

In all the specified language groups (except Tamangs and Rai), the highest scores tend to be found in the Valley (table 3.1.17) whereas it is found lowest in many of the language groups in the Tarai Zone than in the other zones.

Table 3.1.17 Achievement in various language groups in different zones

Eco zone	Nepali[1]	Magar	Tharu	Newari	Urdu	Tamang	Rai	Maithili[4]	Gurung	Limbu
Mountain	60.2[2]	37.4	57.9		63.3		72.2	55.8	52.9	64.0
Hill	57.0	63.5	54.0	51.8	51.9	71.3	49.2	74.8	48.5	71.1
Tarai	57.3	47.1	50.3	52.5	46.0	73.1	43.9	60.5	30.4	40.4
Valley	76.3	81.1	78.5	93.9	78.1	52.6	66.2	89.0	77.9	94.7
Total	61.4	78.6	50.5	52.6	58.6	73.1	58.8	70.8	50.8	56.4
N[3]	8370	606	578	499	327	219	87	35	33	16

1)The language groups of less than 10 students are not included in the table. 2)The main population is highlighted by shading. In some un-highlighted cases, there is only one student behind the mean. 3)The language groups are ordered on the basis of their frequency. 4)Includes Awadhi and Bhojpuri.

The dataset reveals that there is an educational inequality within the language groups in mathematics. In community schools, the students from Magar (79%) and Tamang (73%) backgrounds perform very high in mathematics while the students from Tharu (50%) and Gurung (51%) background perform much lower. The differences between the language groups are remarkable.

Ethnicity/Caste and student achievement

The latest household survey (CBS, 2012) shows that the participation rate of Hill Dalits has increased remarkably in the lower level of education, but their participation in the secondary and higher education is still very low. The results concerning the castes and achievement are condensed in table 3.1.18 and illustrated in figure 3.1.9.

Table 3.1.18 Achievement of the students from various ethnicities by school types

Caste/ Ethnicity	Community				Institutional			
	N	**Mean**	**SD**	**CV**	**N**	**Mean**	**SD**	**CV**
Brahman	837	60.4	23.8	39.4	985	75.6	15.8	20.9
Cheetri	2,342	60.0	24.7	41.2	813	75.3	17.0	22.6
Janjati	4,099	56.2	24.4	43.4	1,393	75.6	17.1	22.7
Madhesi	1,346	54.4	24.6	45.1	224	76.1	18.6	24.4
Dalit	1,901	53.0	24.6	46.4	213	68.8	19.9	28.9
Others	831	50.3	26.0	51.7	179	68.5	21.9	31.9
Total[1]	**14,476**	**54.4**	**25.8**	**47.4**	**4,776**	**74.7**	**17.9**	**24.0**

1) The total is calculated by using the whole dataset while the others also include missing values.

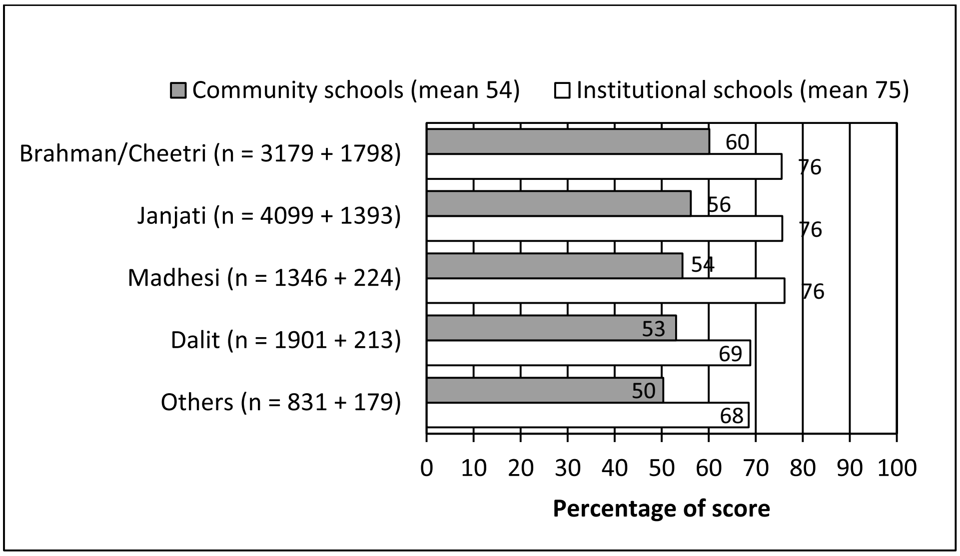

Figure 3.1.9 Relation between ethnicities and achievement

In the community schools, the students from "Other" castes are performing the lowest (50%) in Mathematics, followed by Dalit (53%), Janajati (56%) and Madhesi (54%) students. Dalit students perform below the average also in institutional schools. The overall difference between the groups is statistically significant ($p < 0.001$) though the effect size is small ($f = 0.12$) in community schools. Dividing the students according to their ethnic background explains just 1.4% of the student variation ($\eta^2 = 0.014$). In institutional schools, the effect size is also small ($f = 0.12$); dividing the students according to their ethnic background explains just 1.5% of the student variation ($\eta^2 = 0.015$). From equality point of view, this is a good sign though there is still a lot to do to reduce the gap between the castes.

Dalit students have been followed up because they have historically remained deprieved from education. A positive sign from equality point of view is that the Dalit students perform better than the national mean (53%)[11] in the Eastern (66%), Central (61%), Western (88%), and Far-Western (63%) Mountain areas as well as Western Tarai (63%). However, it is seen that results are much lower than average in the Eastern (45%) and Mid-Western Tarai (39%), and Mid-Western Hill (43%).

Generally speaking, the few Dalit students in institutional schools (n = 213) perform always lower than the average. Especially low performance is found in Central Mountain area (35%).

[11] A number of students did not tell their caste/ethnickty – hence, the missing values. Because of the missing values, the total score (52.6 > 53) is somehow lower than that of the total student sample (54.4 > 54).

Table 3.1.19 Dalit students' achievement by Ecological zones/ Development regions

Schools	**Ecological zone**	**Development Region**					
		Eastern	**Central**	**Western**	**Mid-Western**	**Far-Western**	**Total**
Community schools	Mountain	65.7	61.5	87.7	40.5	62.6	59.2
	Hill	48.4	56.2	55.4	42.7	55.8	51.8
	Tarai	45.3	54.7	62.6	39.0	52.1	52.2
	Total	**50.3**	**56.2**	**56.6**	**41.8**	**55.4**	**52.6**
Institutional schools	Ecological zone	Eastern	Central	Western	Mid-Western	Far-Western	Total
	Mountain		35.5			82.5	38.0
	Hill	74.7	71.7	67.6		87.7	68.8
	Tarai	82.9	58.1	60.1	67.0	76.8	71.0
	Total	**78.4**	**48.6**	**67.3**	**67.0**	**77.9**	**65.5**

Dataset suggests that there are statistically significant difference in achivement though not necessarily remarkable across the ethnicities/castes in Mathematics. Dalit (53%) and Madhesi students (54%) as well as "Other" castes (50%) are performing significantly lower than Brahmin, Chhetri and Janjati. Dalit students perform lower especially in the Central Mountain area (35%) in institutional schools and in the Eastern (45%) and Mid-Western Tarai (39%), and Mid-Western Hill (43%) in community schools.

Gender and student achievement

Efforts have been put globally into reducing the difference between boys' and girls' school achievement. Since the sex or gender-wise equality is considered important in the modern discourse, the matter is handled somehow more extensively than the previous sections of equality. Basic results are presented in table 3.1.20 and figure 3.1.10.

Table 3.1.20 Student achievement of boys and girls by school type

Gender	**Community**				**Institutional**			
	N	**Mean**	**SD**	**CV**	**N**	**Mean**	**SD**	**CV**
Boys	6,514	54.8	25.5	46.5	2,156	73.9	17.6	23.8
Girls	6,384	55.0	25.6	46.5	1,921	74.6	18.2	24.4
Total	**12,898**	**54.9**	**25.5**	**46.5**	**4,077**	**74.2**	**17.9**	**24.1**

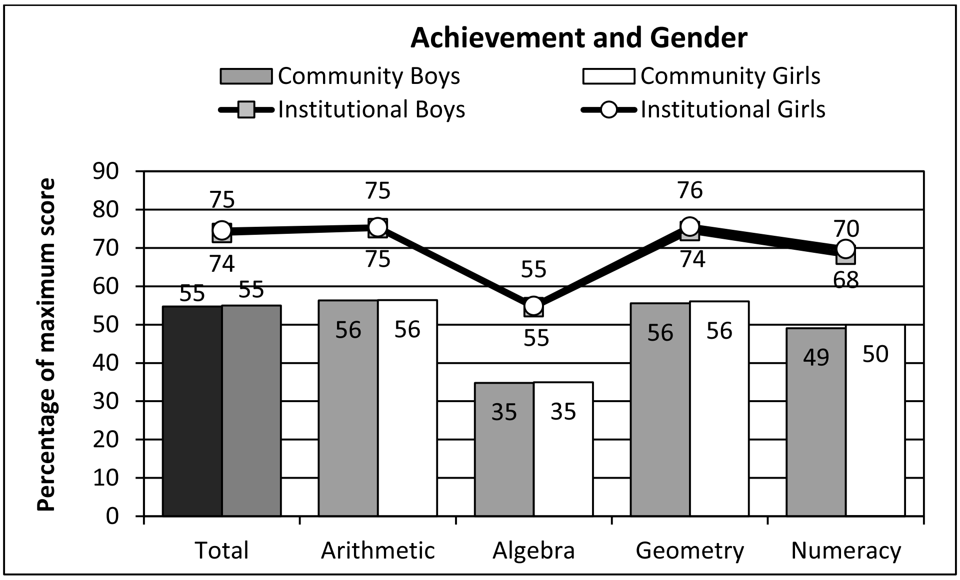

Figure 3.1.10 Comparison of achievements of boys and girls in various content areas

There are no statistically significant differences between boys and girls in both the community and institutional schools in any of the content areas of Mathematics (p = n.s.). From equality point of view, this is a positive signal towards gender parity.

Gender and Ecological zone

In community schools, there is no difference (p = n.s.) between gender among the ecological belts though, in the Valley, the sampled girls are seen to be performing somehow better (73%) than boys (70%). In institutional schools, boys are found to be performing better in the Mountain zone. When it comes to the Ecological zones, the differences between boys and girls are very small except for institutional schools in the Mountain zone.

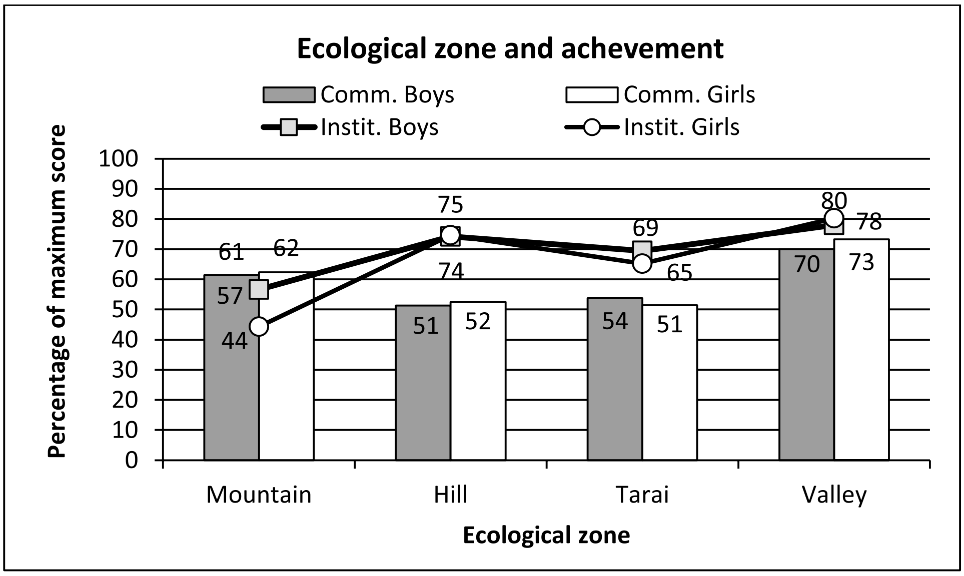

Figure 3.1.11 Comparison of achievement among Ecological zones by school type

Gender and Development regions

There are no notable differences between the Development regions when it comes to boys' and girls' equal opportunities to reach the same educational goals (figure 3.1.12). The difference between boys and girls seems to be somehow wider in institutional schools in the Central (5 percent) and Eastern region (4 percent) and in community schools from the Kathmandu Valley (3 percent) and Eastern districts (3). The differences are narrowing, which is a good sign from equality point of view.

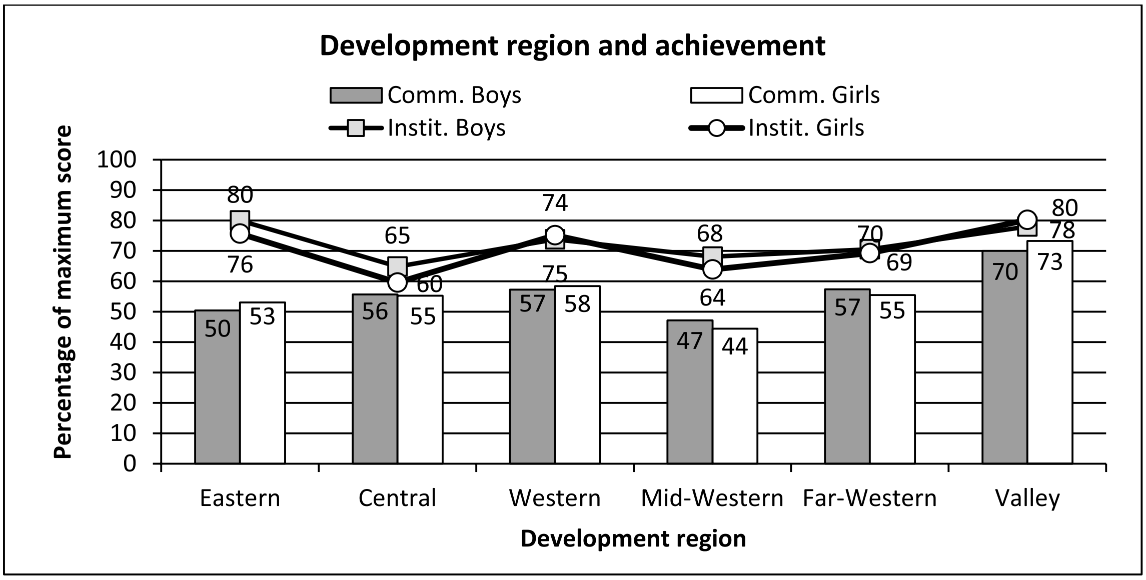

Figure 3.1.12 Comparison of achievement among Development regions by school type

Dataset shows that the differences between boys and girls in Mathematics proficiency are medium level. In institutional schools from Mountain zone, boys outperform girls by 13 percent. Also, in the "Other" castes/ethnicities, girls outperform the boys by 7 percent. Otherwise, the differences are very small. From equality point of view, this is a positive sign.

3.1.3 Selected explanatory factors and achievement

The simplistic model in section 2.4 represents several possible factors, which may explain the differences in student achievement. Many of the factors have already been handled above: geographical factors, such as districts, the Ecological zone, and Development region, as well as school-related factors such as school type and school location. Also some individual related factors were handled, such as home language, ethnicity/caste and sex/gender. In this section, some other factors are taken into consideration. The socio-economic status (SES) of the students' families, paid work before and after school, students' attitude towards Mathematics as a school subject, age of the student, and support provided to the studies are mainly family and individual related factors. As a sample of deepening school and teacher-related factors, also the availability of school books, homework assigned and checked by the teacher, and selected activities in the school are handled.

Two things are worth mentioning. First, some of the questions in the students' background questionnairs were too demanding for grade 3 students to be answered meaningfully alone. Hence, the teacher was asked to help the students fill in the questionnaire. These kinds of questions include, for example, information about parents' education. The questions were included in the questionnaire to obtain comparable information over the different grades. Second, may be because of the first point, there are quite many missing values in the background questionnaires. For example, in the question of mother's education, 4986 students (25.9%) did not answer the question. This evidently has an effect on the analysis and therefore the reader needs to be critically aware of the crudeness of the information when it comes to the grade 3 students. In most cases, however, the result is seen to be credible and comparable with grade 5 datasets.

Parents' education, occupation, home possessions and student achievement

There are many variables indicating the socio-economic status. In NASA 2012, these were categorized into parents' education, parents' occupation, home possessions, home accessories, and whether the student attends a private school or not. Finally, the SES is estimated on the basis of seven indicators related to the economic, educational, and occupational background of the family (see section 2.5). In this section, the education of the parents is further elaborated, so that the illiteracy of the parents is analyzed in relation to the achievement of Mathematics.

Several SES related variables were analyzed by using a data mining tool of SPSS and DTA. The method is very effective in finding the cut-offs of the predicting variable, such as mothers' education, and classifying the factor into several groups, which differ statistically in the most significant way from each other in relation to student achievement. Some examples of this are handled with parents' educational background and its relation with students' achievement in mathematics.

Parents' education

In NASA 2012 background questionnaire, parents' education is divided into eight categories: 1) illiterate, 2) just literate, 3) grade 10, 4) SLC, 5) IA/ grade 12, 6) BA, 7) MA, and 8) above MA. The question was asked to the students and hence there may be some impurities embedded in the data; the number of (just) literate mothers in the dataset seems too high (see figure 3.1.14). However, with the huge dataset the result is seen to be credible.

DTA classifies mother's education into four groups with statistically significant differences in students' achievement levels (figure 3.1.13, of the explanation of the elements in the figure, see Section 2.5): illiterate (students' average is 59), just literate (61%), grade 10 passed (68%), and SLC passed or higher (72%). The difference between each group is statistically significant ($p < 0.001$). In practical words, the results means that when the mother has passed at least grade 10, she can give, on average, + 9 percent point advance for her child in the national test compared to an illiterate mother, and if she has passed SLC, she can give + 12 percent advance for her child. These figures are much lower than in the case of English subject, for example, where the advance is + 22 percent point just by being grade 10 passed. One may note that, when knowing that the national mean is

62%, the students' average level is higher than the mean when mothers' education is grade 10 passed or higher.

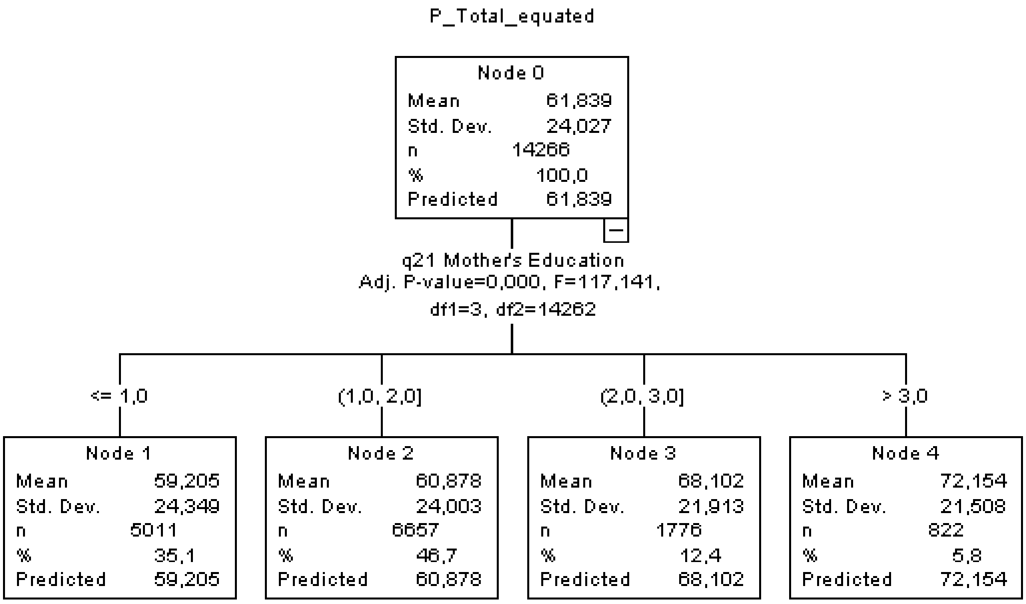

Figure 3.1.13 DTA of mothers' education and students' achievement

Figure 3.1.14 shows that if the mother was IA passed, the advance was + 14 percent over greater than the illiterate mother. Mothers' education explains 2.6% of the student variation (η^2 = 0.026), which indicates a small or moderate effect size (f = 0.16). Obviously, the result means that the children of the highly educated mothers are mainly found in the private schools.

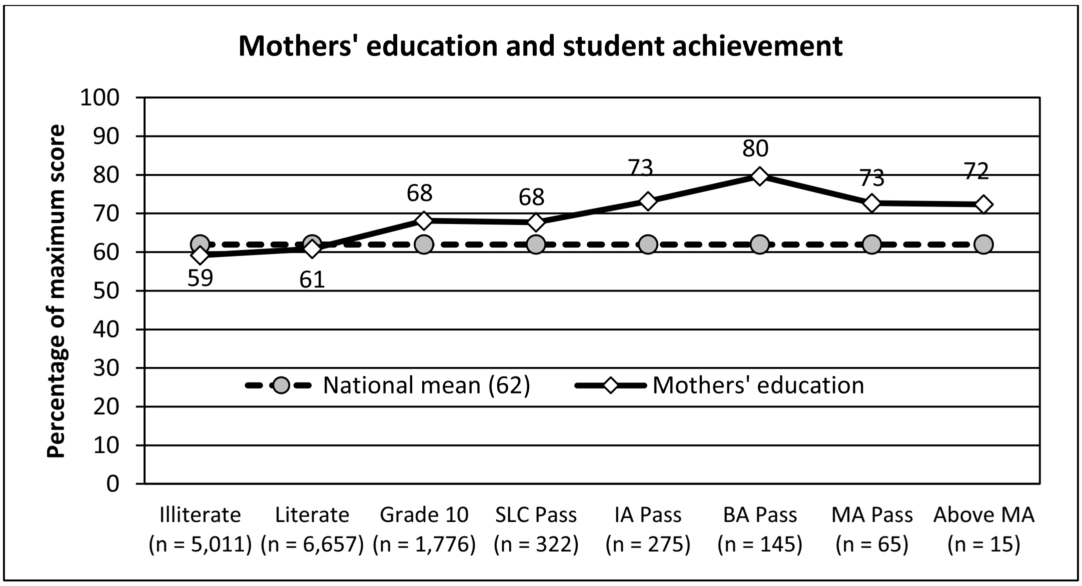

Figure 3.1.14 Mothers' education and students' achievement

In parallel, DTA divides fathers' education into three categories: illiterate or just literate (59%), *grade 10 passed* (67%), and *SLC passed or higher (71%)* (figure3.1.15).The difference between each group is statistically significant ($p < 0.001$). In practical words, the results show that when the father has passed grade 10, he can give, on average, + 8 percent advance for his child in the national test compared to an illiterate father and, if he has passed SLC, he can give + 12 percent advance.

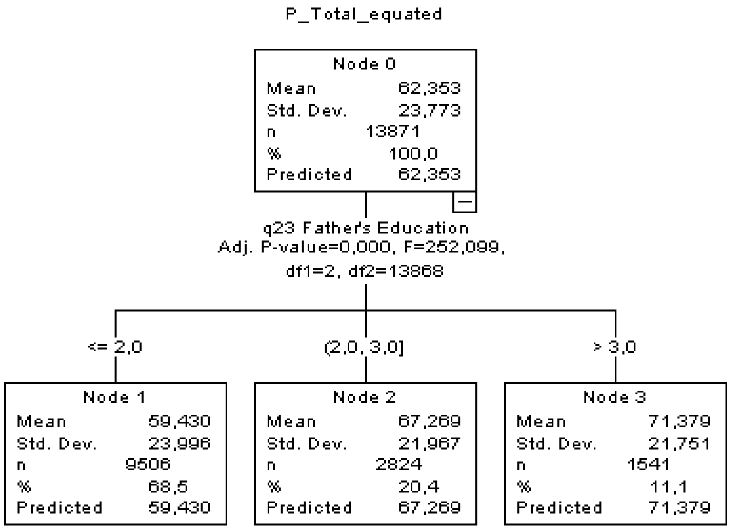

Figure 3.1.15 DTA of fathers' education and students' achievement

Figure 3.1.16 shows that if fathers' education is higher than MA, the advance was + 20 percent more than the illiterate father. The high average means that the children from the highly educated fathers (as well as mothers) are mainly found in the private schools. Fathers' education explains 4% of the student variation (η^2 = 0.038) which indicates a moderate effect size (f = 0.20).

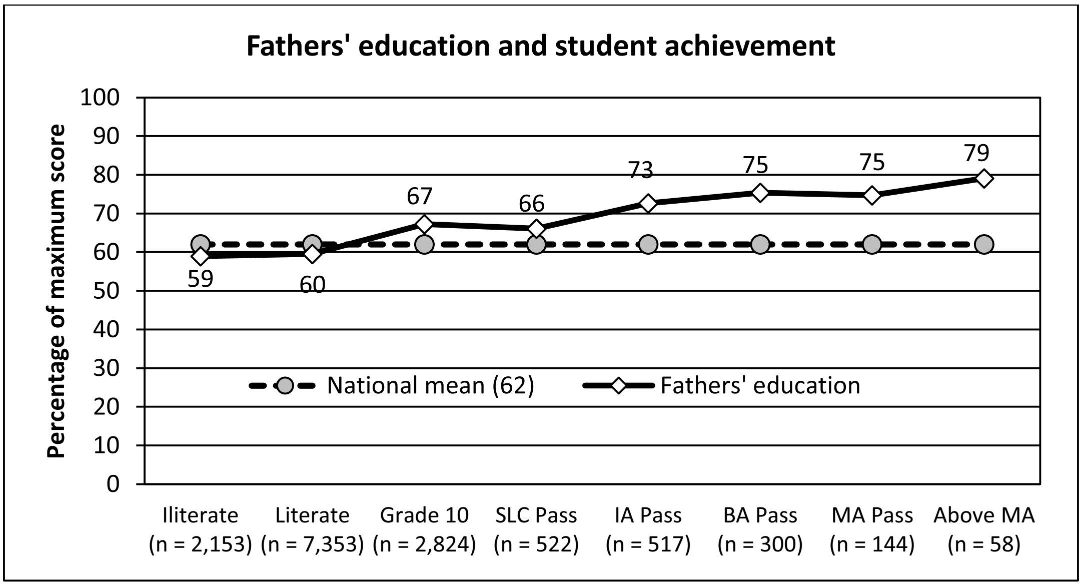

Figure 3.1.16 Fathers' education and students' achievement

After combining the mothers' and fathers' education, the poorest prediction in DTA for the children's future achievement in mathematics comes when fathers' education is not known (49%) or the father is illiterate or (just) literate but the literacy of the mother is not known (52%). The highest results are in the group where both the father and mother have passed (at least) the SLC (76%) or when the father has passed SLC and mother (at least) is grade 10 passed (73%). It is evident that the educational capacity provided by the parents can be utilized by the students: the higher the parents' education the better results will be gained by the children.

In what follows with the final SES variable, grade 10 was determined as the cut-off for parental education, that is, when being passed the grade 10 (or higher), the indicator for mothers' (and fathers') education for SES was set to 1, and the lower education than grade 10 passed gave the value 0.

Parents' occupation

The occupation of parents was categorized into eight groups: 1) working abroad, 2) farming and working at home, 3) only working at home, 4) teaching, 5) services, 6) business, 7) daily wages, and 8) working at other's home. The result related to mothers' occupation is seen in figures 3.17 and 3.18 and to fathers' occupation in figures 3.19 and 3.20.

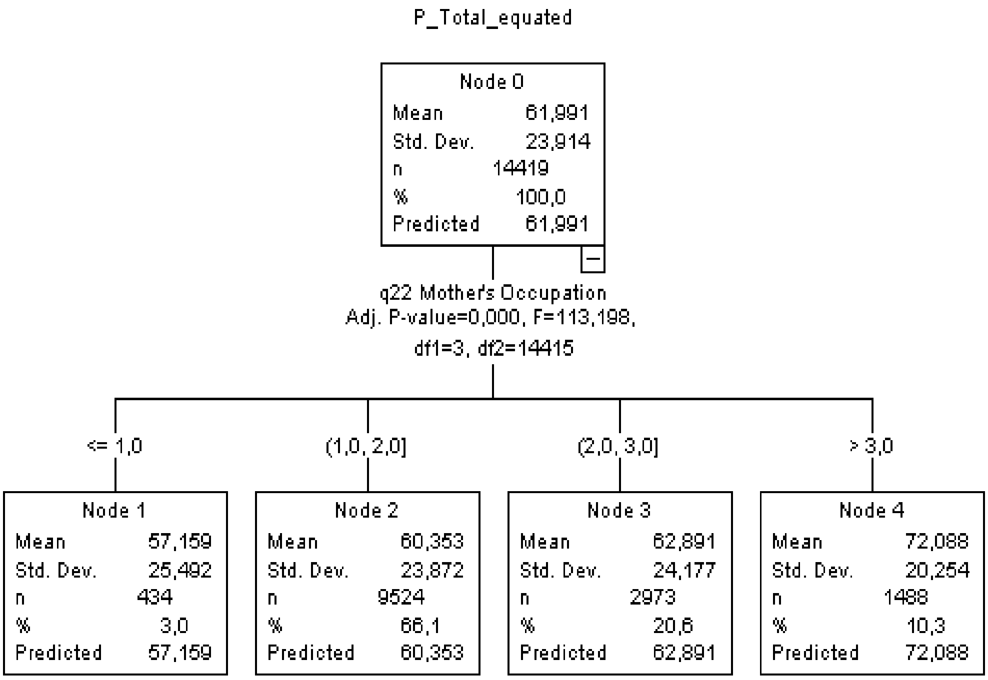

Figure 3.1.17 DTA of mothers' occupation and students' achievement

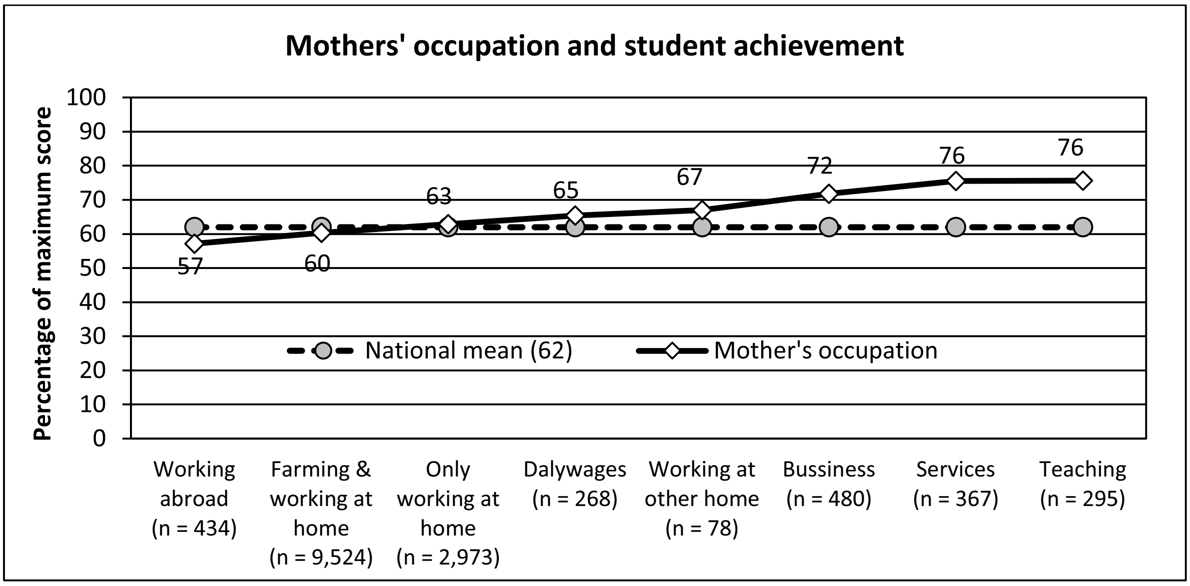

Figure 3.1.18 Mothers' occupation and students' achievement

From figure 3.1.18 and comparative students' mean scores by DTA (figure 3.1.17), it is seen that the achievement is the lowest when the mother works abroad (57%) or mothers' occupation is agriculture (60%). It is statistically significantly lower than when the mother works only at home (63%) nothing to say when mother works as a teacher or in services (76%). Mothers' occupation explains less than 3 percent of the student variation (η^2 = 0.026), which indicates a small or moderate effect size (f = 0.16).

When it comes to fathers' occupation, on the basis of DTA (figure 3.1.19 and ANOVA), the main division is whether or not the father works for agriculture, only at home (that is, is probably unemployed), or abroad (58–61%) in comparison with the other occupations (65–74%). More precisely, if the father is invoved either in agriculture or works only at home, the children's Mathematics skills are remarkably lower (58%) compared with the possibility that the father were in a business (71%), teaching (73%) or service profession (74%). The highest achieving students are from the groups whose parents have a regular income source. The difference between the lowest and highest groups is 27 percent, which is a wide gap. Fathers' occupation explains 6 percent of the student variation (η^2 = 0.064), which indicates a high effect size (f = 0.45).

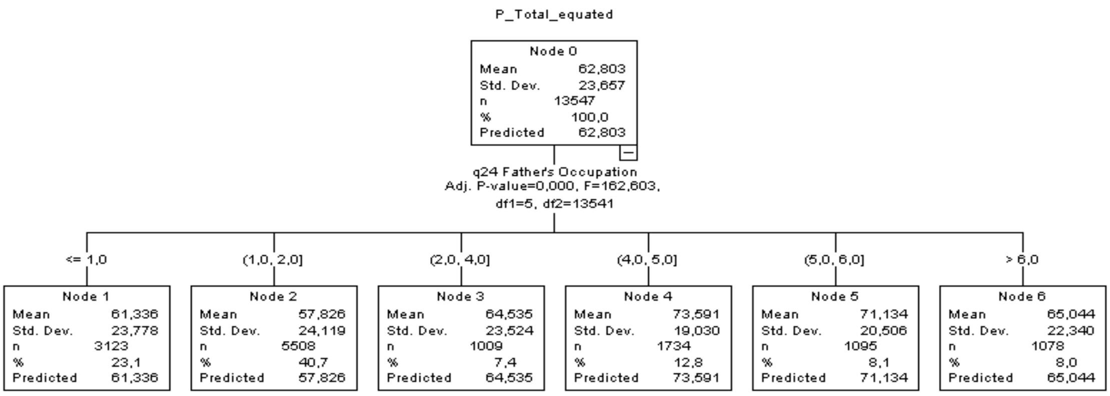

Figure 3.1.19 DTA of fathers' occupation and students' achievement

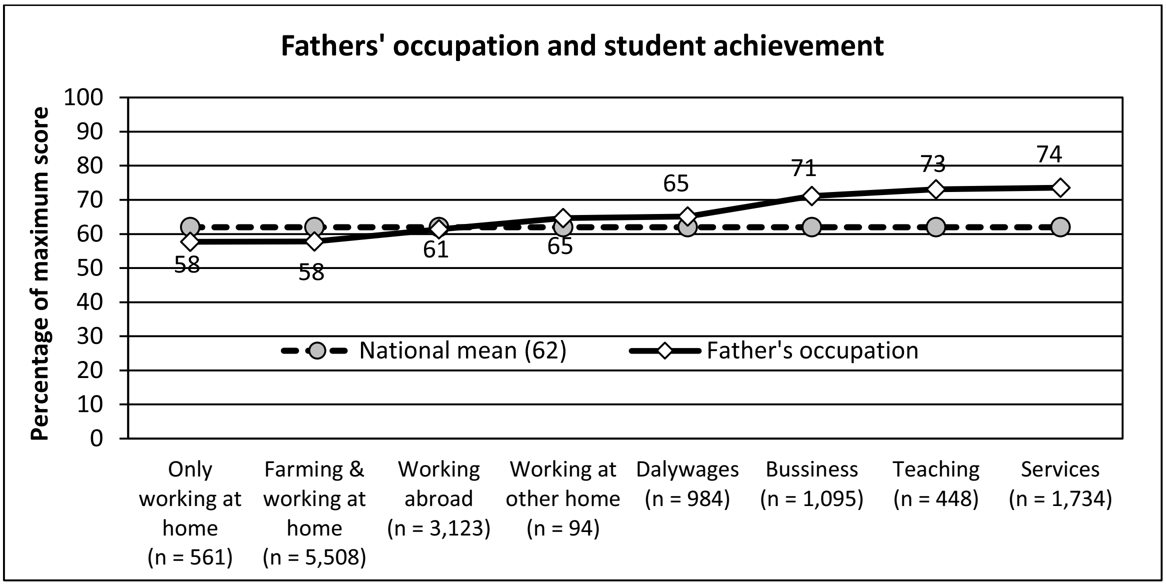

Figure 3.1.20 Fathers' occupation and students' achievement

When combining the mothers' and fathers' occupation, DTA shows that the lowest student achievement is found in the families where the father is working abroad and the mothers' occupation is not known (47, n = 136) , fathers' occupation is unknown regardless what the mother's occupation is (50, n = 1,343), or both parents come from an agricultural background (52, n = 309). The highest achieving students are from the the families where the father is in services (74, n = 1,734) or business (71, n = 1,095) regardless the mother's occupation. It is worth noting that the service and business occupations are more probably urban than rural,contributing to the better results for children.

For the later use as a SES indicator, the cut-off for the parents' occupation was made so that being in agriculture gives 0 and all other options give 1.

Home possessions and accessories

Facilities and resources available at home are found to have some effect 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 help in studying at home: 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 and poetry books, or artistic things like pictures, and books that help them for study, such as a dictionary. Another type of home possession includes different types of normal home accessories (and hence, in what follows these are called home accessories to differentiate them from home possessions) such as the number of mobile phones, televisions, and computers.

There are 11 questions in the student background questionnaire related to home possessions. Each was scored 1 if the student had an access to the possession. Adding these items up, the maximum score was 11 indicating that the student responded to have access to all of the possessions and the lower the score the fewer possessions they have at home. Figure 3.1.21 shows the relation between home possessions and achievement level. Except for the highest category, the achievement level of the students rises logically as there is a greater access to home possessions. Pearson correlation between the achievement level and the factor ($r = 0.16$) is statistically significant ($p < 0.001$) and indicates moderate effect size ($d = 0.35$).

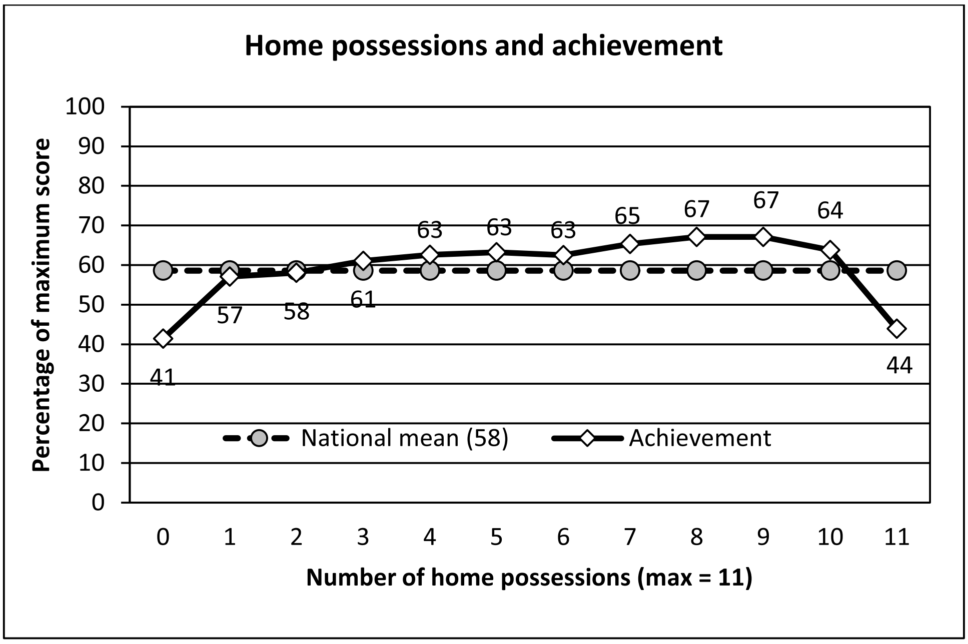

Figure 3.1.21 Association between home possessions and achievement

For the later use in SES, the cut-off for the factor was set on 6 possessions:[12] if 6–10 items were met in the background questionnaire, the student was given 1 otherwise 0.

The same pattern – the more possession, the better results – can be seen also with home accessories, as seen in figure 3.1.22. The question in the background questionnaire was set differently compared with home possessions; with the accessories it was asked *"How many of the following accessories do you have in your family?"* with the options 0 – 3 (or more). For the indicator, the availability of the home accessories is dichotomized in the same way as the home possessions. After dichotomizing the items individually by using meaninful cut-offs found with ANOVA and DTA (and maximizing the differences in achievement level, see figure 3.1.22), all three indicators were summed up.[13] The maximum score was 3 – indicating that the student possessed a set all of the accessories at home.[14]

---

[12] The cut-off was selected to be 6 because of the willingness to keep the boundaries comparable over the subjects. In grade 3 dataset, the cut-off could have been three possessions.

[13] There was also the fourth item in the questionnaire – the number of radios at home. However, this item behaved pathologically in the analysis: the more the radios at home the less the achievement. Hence, it was not taken as an indicator of SES.

[14] The analysis is bound to the fact that the values were given by the students – they are, in many cases, credible. However, as with the home possessions (see figure 3.1.21), here also is a doubt that some of those students who marked all the possessions and accessories either did not understand the question or were just willing to fool with the questionnaire. When it comes to accessories, the effect is not noticeable (compare figures 3.1.21 and 3.1.22).

Table 3.1.21 Dichotomizing the indicators for home accessories

Accessory	Cut-off for 1	Cut-off for 0
Mobile phone	2, 3	0, 1, missing
Television	1–3	0, missing
Computer	1–3	0, missing

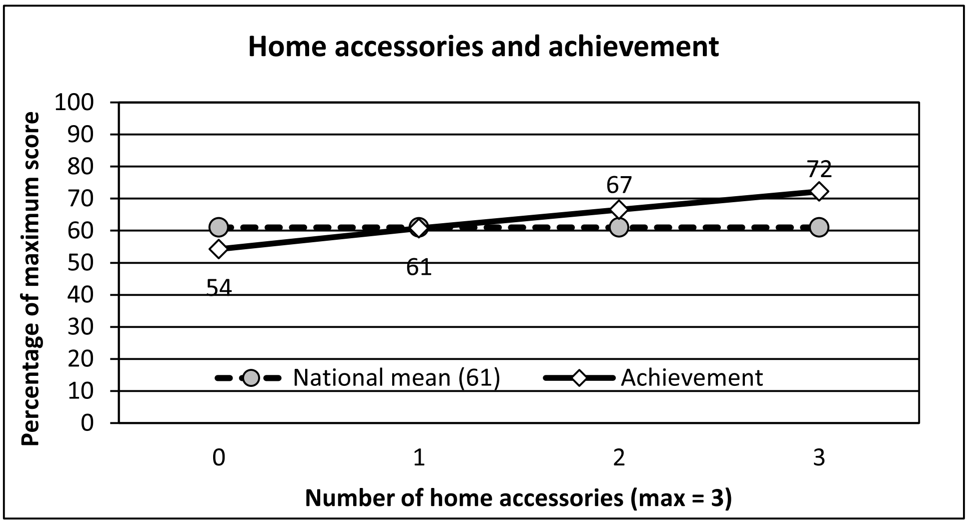

Figure 3.1.22 Association between the number of home accessories and achievement

Figure 3.1.22 clarifies how the increase in number of home possessions or accessories increases in student achievement scores. Students' achievement increases from 45% (if none of them are available) to 64% (if all three of them are available). Availability of all the stated facilities indicates the higher SES of the family. Correlation between the number of home accessories and achievement is $r = 0.25$ ($p < 0.001$), which is certainly positive and indicates moderate effect size ($d = 0.57$).

For the later use in SES, the cut-off for the factor was set on 2 accessories out of 3: if 2–3 items were met in the background questionnaire, the student was given 1 otherwise 0.

The dataset strongly suggests that parents' educational level predicts the children's future achievement level in Mathematics. Especially harmful for the achievement level is seen to be the situation where either the father or mother or both are illiterate or just literate. As amany as 35.1% of the students had an illiterate mother and 15.5% had an illiterate father.

The dataset further indicates that either economic or intellectual capacity or both at home helps children to increase their Mathematics proficiency. If either the father or mother or both are coming from an agricultural or related occupation or they are working abroad, the students' achievement in Mathematics is significantly lower than with the other occupational groups. As many as 66.1% mothers and 40.7% fathers worked in agriculture or only at home.

Data also shows that when children have very few home possessions – none of the 11 – the achievement level is remarkably lower than the national average (41%). With 4 – 10 possessions, the average score is much higher (> 63%) and, in any case, it is higher than the national average. The same is true of home accessories: when there is no accessory, the results are remarkably lower (54%) than when all three were met (72%). Of the sample, 13.7% of the students did not have any of the home possessions and 37.5% had no accessories.

Socio-economic status (SES) and achievement

The socio-economic status of the family was formed on the basis of seven indicators which were all first dichotomized. The variables (mothers' education, fathers' education, mothers' occupation, fathers' occupation, home possessions, home accessories, and type of school where students were studying) were summed up (as SES) and changed into the percentage of the maximum score (P-SES). Deeper description of the transformations is seen in section 2.4. The P-SES represents the percentage of SES of the students' family; 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 are positive), and 0 refers to the lowest possible SES (that is, all the seven indicators of SES are negative). The analysis of the P-SES by using Univariate GLM (that is, the Regression modelling) shows the strong relation between SES and achievement. Figure 3.1.23 presents the relationship between SES of the students and the achievement.

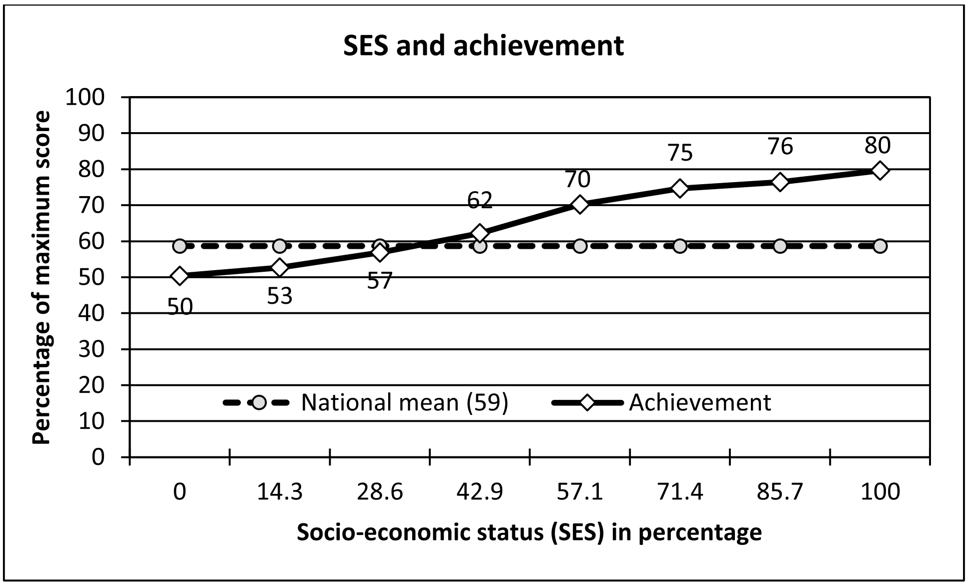

Figure 3.1.23 Relation between SES and achievement

Figure 3.1.23 shows a strict positive relationship between SES and achievement in Mathematics – the higher the SES the higher is the achievement. Pearson correlation between the variables is $r = 0.35$ which is a high value ($p < 0.001$) and indicates high effect size ($d = 0.88$). The difference in achievement between the lowest SES groups (53%) and the highest one (80%) is remarkable. SES explains about 13% of the student variation ($\eta^2 = 0.129$) which is not very high percentage compared to, for example, the English dataset ($\eta^2 = 0.311$), but it seems to be at the same level as was found in Nepali language in the grade 3 dataset ($\eta^2 = 0.144$).

It is worth noting that SES as a variable is more school-related than student-related factor. The correlation between SES and achievement is $r = 0.35$ in the student dataset but $r =$

0.46 in the school-wise dataset. It is also worth noting that even though the SES is controlled in the student-wise dataset[15], there is still a statistically significant difference between the community and institutional schools ($p = 0.001$). However, the effect size is reduced from $f = 0.36$ to $f = 19$, that is, from high to moderate.

From sociological point of view, it might be interesting to know which of the individual indicators of SES are *not* met in those families where the children perform the lowest. Figure 3.1.24 illustrates the fact that in the families with less than four SES indicators met, the challenge lies mainly in three factors marked in figure 3.1.24 with dotted line: both mother's and father's education is low and the child does not attend the private school. The last is difficult to change in the community schools but the low literacy level of parents (and especially mother's low educational level) would be possible to tackle by an educational policy.

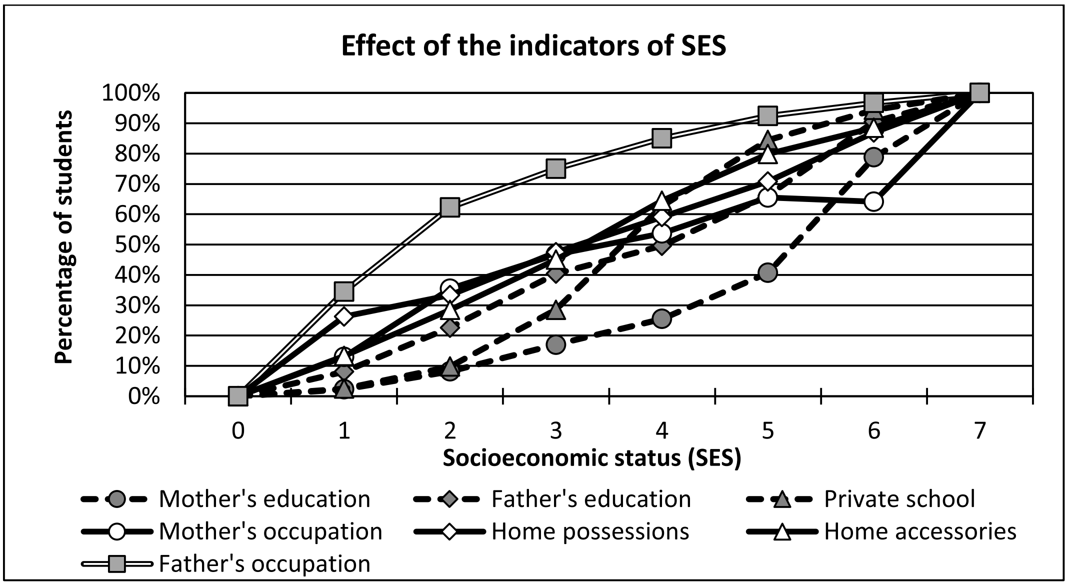

Figure 3.1.24 Effect of individual SES indicators in achievement

The dataset suggests that the socio-economic status plays a vital role in Mathematics achievement in Nepal. The difference between the lowest and highest SES groups is remarkable (30 percent). This means that if the SES of the lowest performing students raises into a decent level, that is, in practice, that the problems of parents' low educational level is solved, the results in these groups will raise remarkably. Especially challenging is the situation in the families where the father or both parents are illiterate or just literate and they work in agriculture or abroad. It is found that 24.8% students are at the lowest level of SES.

---

15 Because the attending to private school is imbedded in the SES, the school type does not explain the achievement in ANCOVA when controlling the SES. For the ANCOVA, another SES – without the school type – was created.

Working beyond school hours and achievement

Several questions were set in the student background questionnaire regarding the students' activities outside the school. Two of them are briefly handled here: Working after or before the school for a paid job, and participating in household work/chores. The values of the variables are divided into five categories: 0 (no time at all), 1 (less than 1 hour per day), 2 (1 – 2 hours per day), 3 (2 – 4 hours per day), and 4 (more than 4 hours per day).

The DTA indicates that, when it comes to working after or before school, the cut-off is whether the students work for a paid job or not. When the children have no paid job at all, the results are above the national average in both community and institutional schools. The connection is straight in institutional schools in comparison to community schools. If the students are working for a paid job – even less than one hour – the results are statistically significantly lower than the average. The ANOVA shows that the relationship is straightforward ($p < 0.001$) though mild ($f = 0.23$ in institutional schools) negative when students need to be engaged in paid work before and after school. It is notable, though, that most of the grade 3 children do not need to be engaged in paid work (63% students in community schools and 75% in institutional schools). Working after or before school indicates that the family is poor and the extra income is needed for living. It is obvious that when the student needs to work for more than 4 hours per day there is no time or energy to handle school homework; in institutional schools, the achievement of children working 2 – 4 hours per day (64, n = 108) is notably lower than that of those with no need at all for paid work (77, n = 2,560).

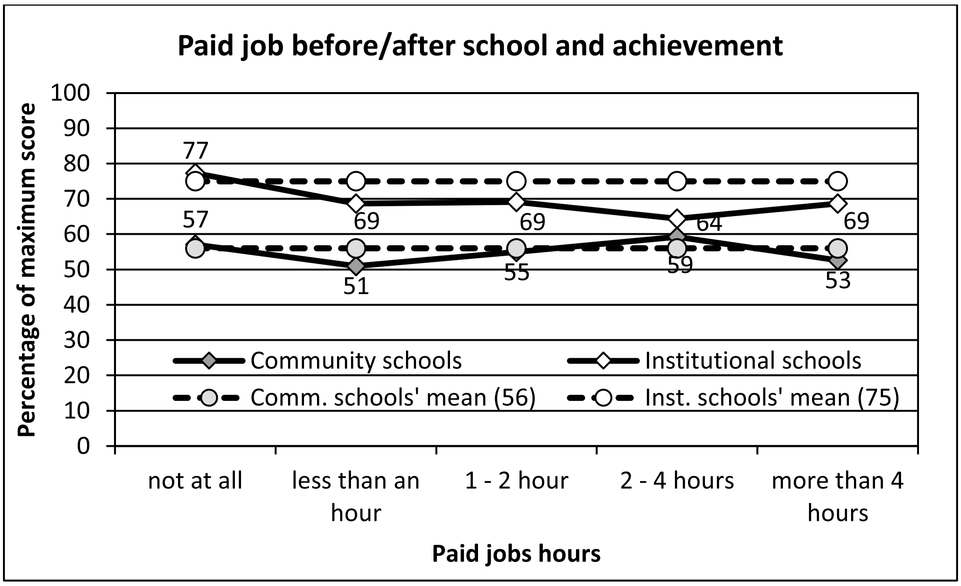

Figure 3.1.25 Relation between achievement and paid job before and after school

When it comes to the unpaid participation in the household work, it is a usual practice in families that the children take part in household chores at home as a part of the socializing process of the children. The DTA shows that when the child spends some time (less than two hour) for the household chores the results are statistically higher (63 – 65%) than in the case of spending not at all (55%) or more than 2 hours per day (60%). The effect of not participating in the household chores is seen to be more drastic in community schools than in institutional schools. Actually, it is seen that in institutional schools, it does not make any difference whether the children work for two hours or less or at all; the effect comes when spending four or more hours in chores. In community schools, the learning results are significantly lower if students are not participating in the chores. Differences are significant ($p < 0.001$) though the effect size is low or moderate ($f = 0.18$ in community schools and $f = 0.09$ in institutional schools). It seems interesting that more than 5% of the students (n = 727) respond that they spend more than 4 hours per day doing household work. In the rural area, this may be obligatory if the family is involved in cattle raising and when the cattle is kept far from home. It is understandable that, in these cases, there is not much energy to decently concentrate on their study or school work.

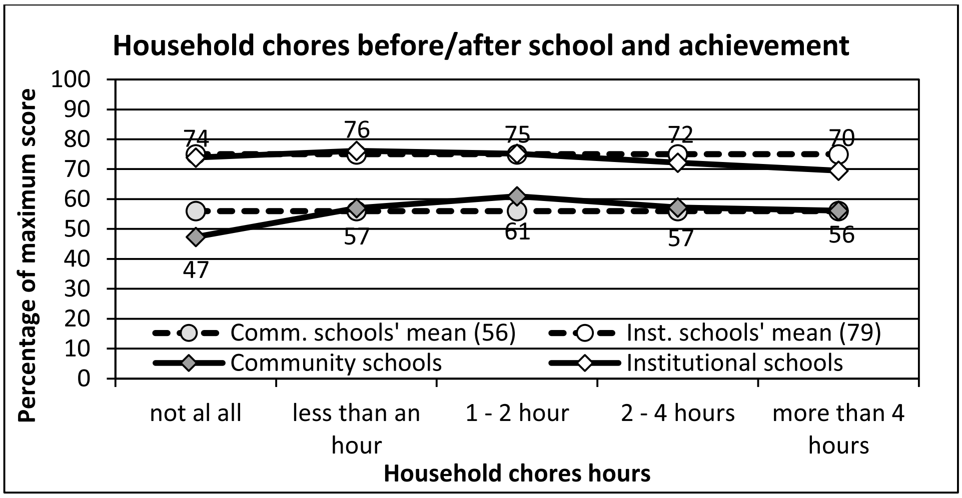

Figure 3.1.26 Relation between household work and achievement

The dataset reveals that either working for a paid job or for more than two hours per day in unpaid household work before or after school reduces the school achievement of the student. However, a decent amount of household work up to two hours per day does not hamper the learning for the students in Mathematics. According to the data set, 33.3% of the students worked for the paid job and 14.9% spent more than 2 hours in the household chores.

Attitude towards the subject and achievement

In the assessment of Mathematics achievement, attitude tells us what the students think about Mathematics and its usefulness in their daily life and future. There is more or less a firm relationship between the attitude of the students and achievement. Though the connection is not always clear, the correlation between achievement and attitude towards the subject as well as self-efficacy in the subject is widely studied (see in mathematics, for example, Metsämuuronen 2012a; 2012b; House & Telese, 2008; Shen & Tam, 2008; Kadijevich, 2006; 2008). Some researchers have noticed remarkable differences in correlation between countries (e.g., House & Telese, 2008; Kadijevich, 2006; 2008; Wilkins, 2004; Shen, 2002; Papanastasious, 2000; 2002; Stevenson, 1998). In some countries, the correlation between attitude and achievement has been found near zero, like in Macedonia (Kadijevich, 2008), in the Philippines (Wilkins, 2004), in Indonesia and in Moldova (Shen, 2002), whereas in some other countries, the correlation has been found as high as 0.60 (e.g., in Korea, Shen, 2002). In NASA 2011, it was noticed that the grade 3 students were not consistent in the attitude test and the reliability of the international test stayed low (see ERO, 2013).

In NASA 2012, technically speaking, the same shortened version of Fennema–Sherman Attitude Scales (Fennema & Sherman, 1976) as is used in several international comparisons like in TIMSS and PISA studies was used. The original scales included nine dimensions, but in these comparisons only three were used with four items on each dimension and two negative items on each of the first two dimensions (see in detail in chapter 2, section 2.4 and 2.5).The names of the original factors were "Liking Mathematics", "Self-Efficacy in Mathematics", and "Experiencing Utility in Mathematics" (compare naming in, e.g., Kadijevich, 2006; 2008). Because of students' inconsistent manner in answering the attitude scale in NASA 2011, only the dimension of "Experiencing Utility in Mathematics" was taken into the measurement instrument for grade 3 students. Reliability of the score of five items is sufficient (α = 0.74). The relation between the attitude (divided into seven group more or less equal number of the students, that is, sextiles[16]) and achievement score is shown in figure 3.1.27.

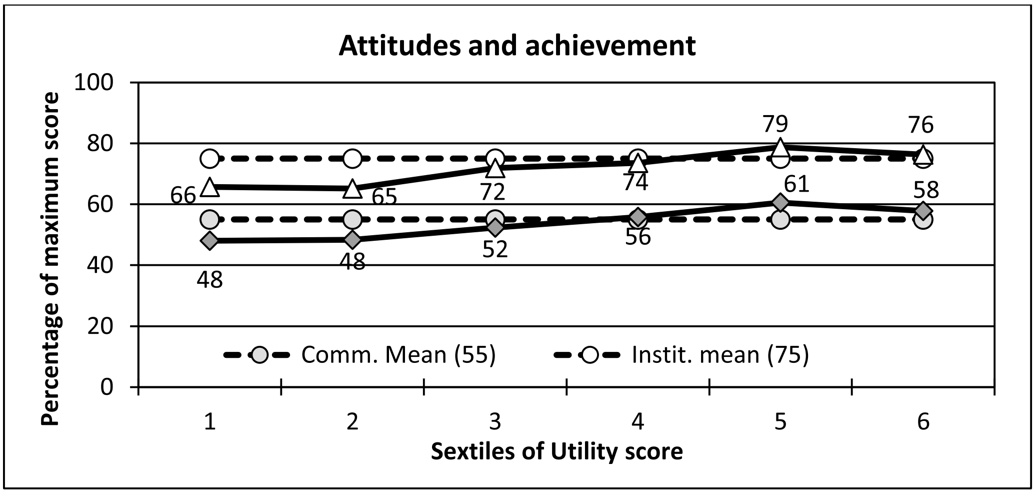

Figure 3.1.27 Relation between attitude and achievement by school type

There is a positive correlation between attitude and Mathematics achievement in the whole dataset (r = 0.18). The relation is moderately high (d = 0.40) – indicating that the difference between the means of the lowest attitude group (50%) and highest one (63%) is remarkable. In the whole dataset, the division of attitude to six groups explains the achievement level somehow 4% (η^2 = 0.042). Connection is found to have been slightly stronger in institutional (r = 0.15) than community schools (r = 0.18). The relation between attitude and achievement is not found fully logical; it seems that, within the highest attitude group, there are many students who may have either fooled in the test or did not understand the questions (see the same kind on phenomenon in the SES analysis

[16] The original score is short (maximum was 15%) and quite many students (36%) gave the maximum score. Hence, it was not possible to form more precise classification such as deciles. Six classes (sextiles) was the most precise alternative with the given dataset.

above). The difference between the lowest and highest attitude group is 13 percent in both the community schools ($f = 0.17$) and institutional schools ($f = 0.23$).

The connection of the sense of utility in Mathematics and achievement is clear though it is not known whether the positive attitude is a consequence of high achievement or the other way round. From statistical point of view, on the basis of simple ANOVA GLM procedure, attitude explains the achievement by 4.2%. Hence, it seems to be difficult to define, in Nepal in grade 3 Mathematics, whether the better achievement is a consequence of more positive attitude or other way round (compare, for example, English in chapter 5 where it is more probable that attitude affects the achievement than other way round).

Data suggests that positive attitude towards the subject correlates with positive achievement in Mathematics. There is no difference between the community and institutional schools in this regard.

Age and student achievement

In Nepalese context, the age of the students attending to grade 3 studies varies widely. Some students have mentioned their age as below nine years and some above 16. All the ages of the students below 10 were encoded as 'up to 9 years', and all students above 14 were encoded as '15 years or above". The descriptive statistics of the mean for each year are presented in tables 3.22(a) and 3.22(b) and depicted in figure 3.1.28.

Table 3.1.22(a) Descriptive statistics of the students' achievement in different age groups

Age	N	Mean	SD	CV
Up to7 years	393	52.1	25.2	48.3
8	1,991	56.4	25.7	45.5
9	4,696	60.7	25.3	41.6
10	5,030	60.7	24.7	40.7
11	2,253	59.3	24.7	41.7
12	1,342	58.6	25.0	42.6
13 years or higher	604	55.0	26.0	47.3
Total	**16309**	**59.4**	**25.1**	**42.3**

Table 3.1.22(b) Student achievement in different age groups by the type of school

Age	Community				Institutional			
	N	Mean	SD	CV	N	Mean	SD	CV
Up to7 years	363	50.9	25.1	49.3	30	66.5	21.9	33.0
8	1,728	53.9	25.6	47.6	263	73.3	18.5	25.2
9	3,269	54.3	25.6	47.1	1,427	75.4	17.1	22.7
10	3,612	54.8	25.0	45.6	1,418	75.7	16.1	21.2
11	1,776	55.6	24.6	44.3	477	73.1	19.6	26.8
12	1,137	56.6	24.9	43.9	205	69.7	22.8	32.8
13years or higher	521	53.2	26.2	49.3	83	66.5	21.8	32.8
Total	**12,406**	**54.6**	**25.3**	**46.2**	**3,903**	**74.5**	**17.8**	**23.8**

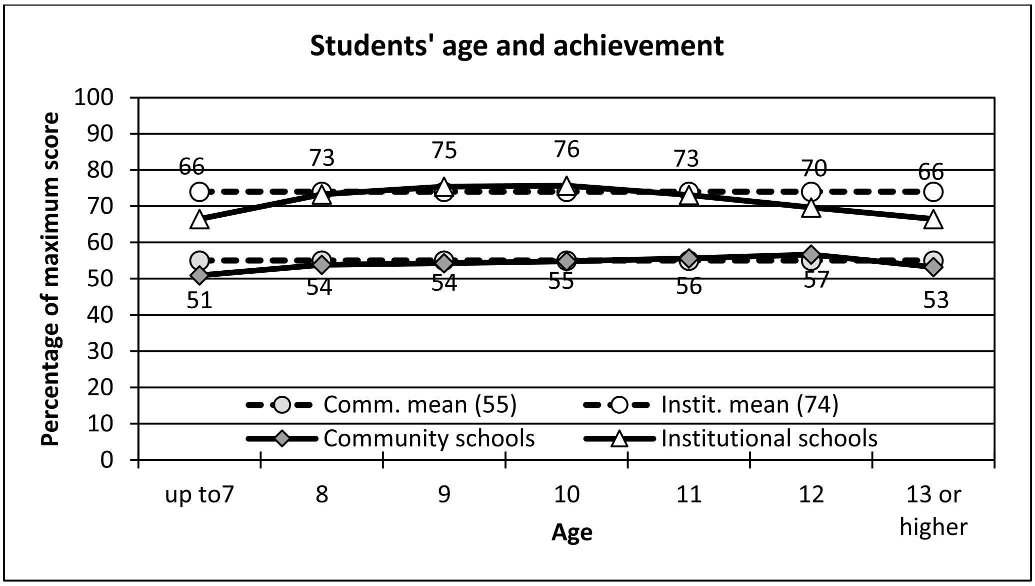

Figure 3.1.28 Association of age with achievement

It is found that in institutional schools, the best achievers are those students who are at the proper age for grade 3 studies (9 to 10 years old), scoring 75–76%. The phenomenon is not that clear in the community schools. However, it is clear that the higher the age is – meaning that the students have either started their schooling much later than they should have, or they have repeated the classes – the weaker the results are. This happens more probably in institutional schools. The achievement level is remarkably lower than the average when the students are at the age of 13 or above (66% in institutional schools and 53% in community schools). Correlation between the age and achievement in the institutional schools is $r = -0.06$ ($p < 0.001$) – indicating low effect size ($d = 0.12$); in the community schools the correlation is near zero ($r = 0.03$, $d = 0.01$). Though their results seem to be weaker, it is good that these "over-aged" students entered school to learn although they should have been provided with remedial instruction earlier by means of extra tuition or additional coaching.

Dataset reveals that the highest performance in mathematics is found with those students studying at their normal age group, that is, at the age of 8 to 10 years. Otherwise achievement lowers down as the age increases. Of the total, 30.1% of the students fell aside 8–10 years.

Support to study and student achievement

The relation between the support provided for studies and achievement was analyzed based on the following question: *"Who helps you when you do not understand what you have read?"* In the question, only one option was selected – in many cases, there might have been several support providers, which cannot be detected now. The descriptive statistics of the supports are given in tables 3.1.23 and 3.2.24.

Table 3.1.23 Descriptive statistics of support to student for study and achievement

Support received	N	Mean	SD	CV
Tuition	1,114	63.0	22.3	35.5
Brother/Sister	6,112	61.4	24.1	39.3
Teacher	1,818	61.2	23.9	39.1
Mother	1,650	61.0	24.5	40.2
No one	318	58.6	25.7	43.8
Father	4,389	57.7	26.2	45.4
Total	**16,814**	**58.6**	**25.5**	**43.5**

Table 3.1.24 Support received by the student and achievement by type of school

Support received	Community				Institutional			
	N	Mean	SD	CV	N	Mean	SD	CV
Brother/Sister	4,730	57.4	24.3	42.3	1,382	75.2	17.8	23.7
Tuition	696	55.8	23.1	41.3	418	75.0	14.7	19.6
No one	275	55.5	25.5	46.0	43	78.7	16.1	20.4
Mother	1,144	55.2	25.1	45.5	506	74.0	17.1	23.1
Teacher	1,180	55.2	24.3	44.1	638	72.4	18.6	25.7
Father	3,554	53.5	26.1	48.8	835	75.5	17.6	23.3
Total	**12,879**	**53.8**	**25.6**	**47.5**	**3,935**	**74.3**	**17.8**	**24.0**

It seems that an external support is, in many cases, necessary for the students to gain better than average marks on the test. However, the reality is found different in the community schools compared to the institutional schools. In the whole dataset, there is about 4 percent difference between those who do not get any kind of support (59%) and those who receive (private) tuition (63%). It is likely that the children with the private tuition also spend more time on the homework which explains higher score. The support received from brothers or sisters as well as mothers or teachers help raise the achievement level above the average (61%). Those who got support from their fathers gained lower than the average – even lower than those having no tuition at all.

The students in community schools are found to have been supported most effectively by their brothers or sisters (57%) or (private) tuition (56%). In the institutional schools, on contrary, the highest results are seen with those who have studied just by themselves (79%). The effect of support providers is, in any case, quite low: effect size is moderate (f = 0.22) in community schools and low (f = 0.11) in institutional schools.

Dataset reveals that the highest performance in mathematics is found with those students studying at their normal age group, that is, at the age of 8 to 10 years. Otherwise the achievement lowers down as the age increases. Of the total, 30.1% students fell aside 8–10 years.

The dataset clearly shows that the support provided by the brother/sister, mother and teacher raises the achievement level more than the support provided by the father. In the whole sample, the highest achieving group is the one who receives private tuition. However, the difference between the highest and lowest performing groups is not notable. It is possible that the group with private tuition also spends more time on their homework, which may explain the higher score.

Availability of textbook and student achievement

The data shows that there are still some students who do not have the proper textbook even by the end academic session. Table 3.1.25 shows the descriptive statistics of availability of the Mathematics textbook and the achievement.

Table 3.1.25 Availability of textbook of Mathematics and the achievement

Availability of Mathematics textbook	N	Mean	SD	CV
Yes	14,601	60.9	24.7	40.5
No	892	53.7	24.5	45.7
Total	**15,493**	**60.4**	**24.7**	**40.9**

Out of 15,493 students who responded to the question, 5.8% (5.0% in community schools and 8.1% in institutional schools) informed that textbook was not available to them. The relation between the accesss to textbook and achievement is significant ($p < 0.001$) though the effect size in the whole dataset is small ($f = 0.07$). The difference in achievement is 7 percent (10 in community schools and 7 in institutional schools).

The dataset shows that 5.8% of the students lack the proper textbook of Mathematics in grade 3. The achievement level of these students is significantly lower (54%) than those who have access to the textbook (61%).

Homework assigned/checked and achievement

Homework is one of the ways to enhance learning; it can be used as drill, exercise, and as an evaluation tool. When homework is systematically checked, most probably it boosts achievement level. Since the related results in NASA 2012 are based on students' reports, it may be possible that in the same classroom some students have given a slightly deviant response than the other students. However, in the wide scope, the results seem to make sense. Statistics related to homework assigned and checked is presented in tables 3.1.26 and 3.1.27 and depicted in figure 3.1.29.

Table 3.1.26 Homework given/checked and the achievement

Status of homework	N	Mean	SD	CV
Given Some day - Checked Every day	832	63.3	22.6	35.7
Given Every day - Checked Some day	1,750	61.4	25.0	40.6
Given Every day - Checked Every day	11,326	60.7	24.8	40.8
Given Some day - Checked Some day	1,016	57.9	23.6	40.7
Not given	250	53.8	24.8	46.0
Given Some day - Not checked	79	53.4	21.3	39.9
Given Every day - Not checked	170	49.6	24.4	49.1
Total	**15,423**	**60.5**	**24.6**	**40.8**

Table 3.1.27 Homework given /checked and achievement by the type of school

Status of homework	Community				Institutional			
	N	Mean	SD	CV	N	Mean	SD	CV
Given Every day - Checked Some day	1,130	53.4	25.1	47.0	620	76.2	16.7	22.0
Given Every day - Checked Every day	9,040	56.9	24.8	43.7	2,286	75.8	17.7	23.3
Given Some day - Checked Every day	458	55.6	24.5	44.0	374	72.8	15.6	21.4
Given Some day - Checked Some day	610	48.8	23.4	48.0	406	71.6	16.0	22.3
Given Every day - Not checked	123	43.8	24.3	55.3	47	64.8	17.1	26.4
Not given	171	49.1	24.7	50.4	79	64.0	21.7	33.9
Given Some day - Not checked	40	43.9	20.3	46.1	39	63.1	17.9	28.4
Total	**11,572**	**55.7**	**24.9**	**44.6**	**3,851**	**74.6**	**17.5**	**23.4**

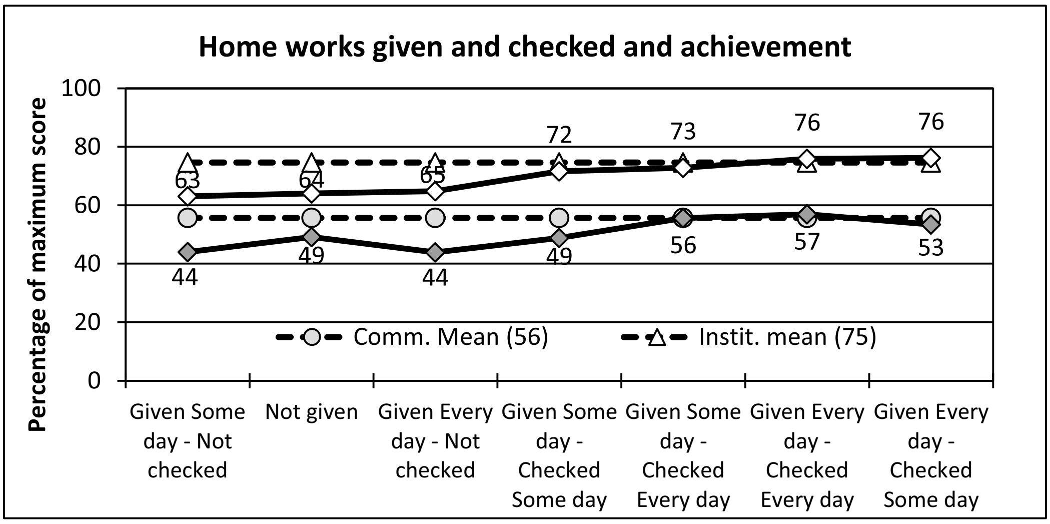

Figure 3.1.29 Relation between homework assigned/checked with achievement

It is evident that, if the students claim that the teachers do not assign them homework or do not check those, the student achievement is notably lower (44–49% in the community schools and 63–65% in institutional schools) compared with the situation that the teacher either assigns or checks the homework regularly (53–57% in the community schools and 73–76% in institutional schools). The differences are statistically significant ($p < 0.001$). However, the groups with no homework assignments or checking are very small (3.2%)

and hence, the effect size is small (f = 0.11 for community schools and f = 0.16 for institutional schools); grouping explains only 1–2% of the variance in the data (η^2 = 0.011 for community schools and η^2= 0.024 for institutional schools).

Dataset suggests that if the teacher assigns and checks homework regularly, the achievement is higher than without checking or assigning of homework. By assigning and checking homework daily, the teacher is seen likely to raise the students' scores up to 13 percent. The dataset shows that 3.2% of the student did not get homework or those were not checked.

Activities in the school and student achievement

The activities of the students and the teacher determine the learning environment of the school. Bullying, for example, is one of the hindering activities of the students in the school that may affect learning. In the student background information questionnaire, several students 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 indicators and students' impressions of school's and teacher's activities are taken as an example of positive indicators.

Negative activities - Bullying

Bullying is one of the problems in the school that worsens the students' learning environment. International studies like TIMSS and PISA give a specific emphasis on identifying such indicators which are seen in their background questionnaires. In the NASA 2012 student questionnaire, five questions indicate the varieties of bullying that are likely to happen in the school. All the questions were stemmed by the phrase *"Which of the following activities happened in your school in the last month?"* The students' responses are presented in tables 3.1.28 and 3.1.29 and depicted in figure 3.1.30. '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 reported that the particular type of bullying happened within the last month. As many as 26% of the students mentioned that something of their own was stolen during the last month, which is an alarming sign for the system.

Table 3.1.28 Frequencies of encountered bullying

Type of Bullying	No (%)	Yes (%)
I was made fun of or called names	71.8	28.2
Something of mine was stolen	71.3	28.7
I was hit or hurt by other student(s)	75.7	24.3
Fellow students kept outside without involving me in activities	75.0	25.0
I was made to do things I didn't want to do by other students	80.5	19.5

Table 3.1.29 Bullying and the achievement by the type of school

Intensity of bullying	Community				Institutional			
	N	Mean	SD	CV	N	Mean	SD	CV
No bullying	5,095	59.2	24.5	41.4	1,991	76.4	17.1	22.4
20% bullying	2,265	57.8	23.8	41.3	977	75.6	15.9	21.0
40% bullying	1,723	54.3	23.5	43.3	491	71.9	16.9	23.5
60% bullying	1,127	52.1	24.9	47.8	251	68.4	19.7	28.9
80% bullying	536	50.8	25.2	49.6	92	68.2	22.4	32.8
100% bullying	815	42.0	25.1	59.9	60	57.0	27.0	47.5
Total	**11,561**	**55.9**	**24.8**	**44.4**	**3,862**	**74.6**	**17.6**	**23.6**

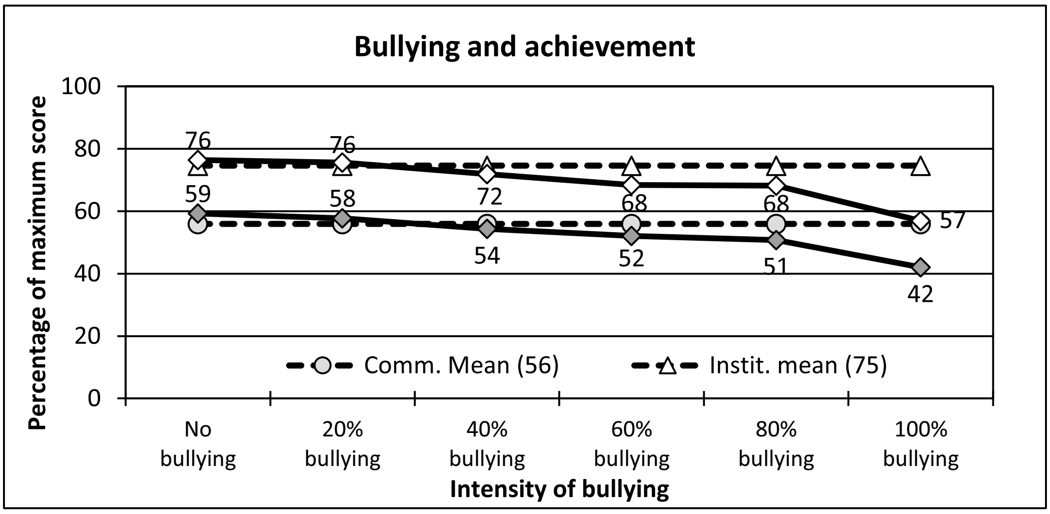

Figure 3.1.30 Relation between bullying and achievement

The sum of all five items forms an indicator of bullying. Figure 3.1.30 shows the achievement of the students in each category of bullying. If only one activity of bullying is reported, it is categorized as 20% bullying, and if all five activities are reported it is categorized as 100% bullying. Knowing that 46% of the students (44% in community schools and 52% in institutional schools) did not encounter any bullying during the last month, one can infer that the remaining 54% did encounter at least one type of bullying. This is a remarkable number of students. As many as 9.7% students (11.7% in community schools and 3.9% in institutional schools) are experiencing a severe kind of bullying (the sum of 80% and 100% bullying). This means, in practice, that more than 83,000[17] grade 3 students in Nepal are encountering physical, psychological, and social bullying every month. The number is too much even though it would not have any effect in learning outcomes. However, it is seen that learning outcomes are notably lower than the average with 19% of the students who have encountered more than two different types of bullying (42–52% in community schools and 57–68% in institutional schools). Students who do not experience bullying and students who encountered extremely bullying of five kinds have

[17] According to the "Primary level total enrollment in all types of schools by district, Flash I_2012–2013", there were 859,593 grade 3 students. 9.7% of these is 83,380 students.

19–23 percent achievement gap; though there are a few students who reported this kind of bullying (n = 875). The difference is statistically significant (p = 0.001) though the effect size is small or medium (f = 0.18 in community schools and f = 0.11 in institutional schools). Though extreme cases of severe bullying are rare, bullying seems to be quite common in schools. This negative phenomenon causes needless harm to young children and has to be rooted out from schools.

Positive activities in school

The activities that can boost the learning achievement of the students are categorized as positive activities. The students were asked about such activities in the school using two sets of questions collected in table 3.1.30. The table shows the responses of the students in all four categories; the responses are in the 4 point rating scale anchored to fully disagree (0) and fully agree (3). Generally speaking, the 3rd grade students express content with the school and student related activities in school. However, remarkably high number of students (8.4%) expressed that they feel that the teacher is not treating them fairly. The same phenomenon was seen also in 2011 datasets with grade 8 students where 11% of the Mathematics students, 2% of Nepali (see ERO, 2013), and 13% of social studies felt – outstanding from the other questions – unfair behavior of teacher.

Table 3.1.30 Students' response towards teacher and school's activities

Teacher and Students activities[1]	**Respondents in %(valid %)**			
	Fully agree	**Partially agree**	**Partially disagree**	**Fully disagree**
q28a: I like come and stay in school	91.6	5.2	1.3	1.9
q27a: Students get along well with most teachers	88.0	8.2	1.5	2.3
q28c: Teacher in the school care about the students	87.3	8.0	2.4	2.3
q27b: Most teachers are interested in student's well-being	86.7	8.6	2.1	2.7
q27d: If I need extra help, I will receive it from my teacher	85.6	9.1	2.2	3.1
q27c: Most of the teachers really listen to what I have to say	84.2	9.7	3.1	3.0
q28b: Students in my school like me	82.3	13.2	2.4	2.2
q27e: Most of my teachers treat me fairly	79.3	12.3	3.2	5.2
Average	**85.6**	**9.3**	**2.3**	**2.8**

1) *The activities are ordered on the basis of percentage in "Fully agree".*

Further analysis is carried out by recoding the variables into two categories (2–3 = 1, that is, agree 1and disagree 0). Furthermore, the sum of eight indicators is converted into the percentage of maximum score to analyze the level of positive activities and its relation to achievement.

DTA finds five attitude groups in the indicator. These boundaries and descriptive statistics are seen in tables 3.1.31 and 3.1.32 and illustrated in figure 3.1.31. 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 positive activities and positive association to achievement is (r = 0.24) statistically significant (p < 0.001) and moderately high (d = 0.55).

Table 3.1.31 Teacher and school related activities and the achievement

Percentage of positive actions	N	Mean	SD	CV
50.0% or lower	1,142	43.3	24.1	55.5
62.5%	635	47.3	23.7	50.1
75.0%	709	52.1	23.9	45.8
87.5%	1,555	58.4	24.1	41.2
100%	11,676	63.2	24.0	38.0
Total	**15,717**	**60.1**	**24.8**	**41.2**

Table 3.1.32 Teacher and school related activities and achievement by school type

Percentage of positive actions	Community				Institutional			
	N	Mean	SD	CV	N	Mean	SD	CV
50.0% or lower	996	41.1	23.6	57.4	146	58.4	21.6	37.0
62.5%	518	44.0	23.2	52.7	117	61.8	20.1	32.6
75.0%	545	47.8	24.0	50.4	164	66.5	16.4	24.6
87.5%	1,110	52.8	24.4	46.2	445	72.4	16.2	22.4
100%	8,681	58.6	24.4	41.7	2,995	76.6	16.9	22.1
Total	**11,850**	**55.5**	**25.0**	**45.0**	**3,867**	**74.5**	**17.7**	**23.8**

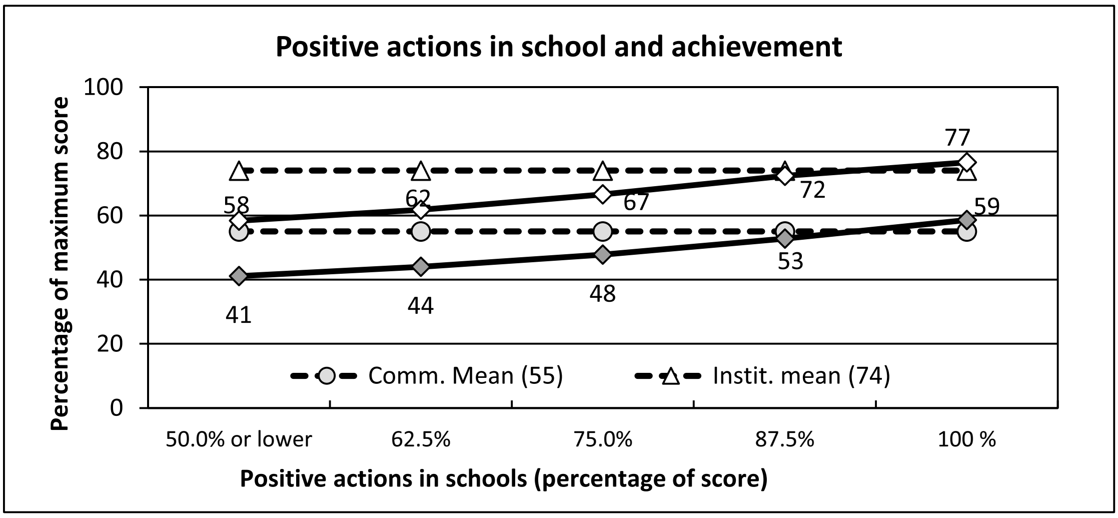

Figure 3.1.31 Relation between positive actions in school and achievement

The data shows that there is a firm positive connection between the students' feeling about the teacher and school related activities and their achievement. The increase in achievement is directly proportional to the increase in the intensity of such activities. After dividing the indicator into five groups on the basis of DTA, the differences between the groups are statistically significant ($p < 0.001$), however, the effect size is moderate (f = 0.24 in community schools and f = 0.27 in institutional schools); the difference between the most positive group and the most negative group is notable (19–22 percent). Most students (73% in community schools and 77% in institutional schools) were content in all items (100% contentment). Only when the students are extremely positive towards school and teachers' behavior, the learning achievement is higher than the average. Students with a negative feeling of five or more of the eight indicators (62.5% of the total) are in great danger of achieving much lower than the average in mathematics.

The dataset reveals that an alarmingly high number of the students (54%) have encountered some kind of bullying in school within the last month, and 9.7% of students are experiencing a severe kind of bullying. This means that more than 83,000 grade 3 students in Nepal are encountering physical, psychological, and social bullying every month. The phenomenon has been affecting the learning outcomes in almost all the groups of the students who felt bullying, so all possible efforts have to be made to root out the phenomenon from schools.

The dataset also suggests that when students feel that the actions of the teachers and the schools are ultimately good, Mathematics results are better than average (59% in community school and 77% in institutional schools). At the other extreme, in feeling ultimately negative of such actions, the results are far below the average (41% in community schools and 58% in institutional schools). According to the datasets, 8.4% of the students feel that their teachers do not treat them fairly.

3.1.4 Synthesis of the Analysis

Several individual student related and geographical factors have been detected which individually explain the difference in achievement between the students. These factors are collected in table 3.1.33. One may note that, except for gender, all the factors showed statistically significant difference between the groups when analysed individually.

Table 3.1.33 Variables handled in the text and their effect in one-way ANOVA

Variable and values[1]	Leverage[2]	Eta squared[3]	Effect size[4]
Ecological zone (1 = Mountain, 2 = Hill, 3 = Tarai, 4 = Valley)	+22.5	0.104	0.34
Development Region (1 = Eastern … 5 = Far-Western, 6 = Valley)	+31.5	0.139	0.40
School location (0 = Rural, 1 = Urban)	+15.0	0.059	0.25
School type (0 = Community, 1 = Institutional)	+20.3	0.117	0.36
Gender (0 = Girls, 1 = Boys)	0.00	0.000	0.00
Cast (1 = Janjati, 2 = Dalit, 3 = Madhesi, 4 = Brahman, 5 = Cheetri)	+14.0	0.032	0.18
Language at home (1 = Nepali, …, 12 = Other)	+22.5	0.032	0.18
Mother's Education (1 = Illiterate, …, 8 = Above MA)	+20.5	0.026	0.16
Father's Education (1 = Illiterate, …, 8 = Above MA)	+20.1	0.038	0.20
Mother's Occupation(1=working abroad,.,8 = working at other home)	+18.5	0.026	0.16
Father's Occupation(1=working abroad,.,8 = working at otherhome)	+15.9	0.064	0.26
Home possessions (sum; max 11)	+27.7	0.103	0.34
Home accessories (sum; max 3)	+17.9	0.062	0.26
SES (sum; max 7)	+29.3	0.129	0.38

Variable and values[1]	Leverage[2]	Eta squared[3]	Effect size[4]
I do jobs at home (1 = not at all, …, 4 = more than 4 hours)	+10.4	0.020	0.14
I work at a paid job (1 = not at all, …, 4 = more than 4 hours)	+8.8	0.021	0.15
Attitude Utility in Mathematics (sum;max 15)	+18.3	0.046	0.22
Age	+8.6	0.006	0.08
Who helps you …? (1 = Father, …, 6 = Teacher)	+5.3	0.053	0.24
Do you have textbook of Math subject (0 = No, 1 = Yes)	+6.8	0.005	0.07
Homework(0=not given,…,6= Given every day, checked every day)	+13.7	0.005	0.07
Bullying (sum; max 5)	+21.0	0.050	0.23
Positive Activities in school (sum; max 8)	+23.7	0.061	0.25

1) *The order of the variables is the same as handled in the Sections above*
2) *Difference between the lowest and highest group-mean 3)On the basis of one-way ANOVA*
4) *Cohen's f*

On the basis of univariate ANOVA, Development region, closely followed by the socio-economic status and school type, seems to be the most effective single factors in affecting the achievement level of the student; effect sizes are *f* = 0.40, *f* = 0.38, and *f* = 0.36 respectively. Some of these variables in figure 3.1.33 may be strongly related to each other and hence they may not add value in explaining why some students are performing much better than others. In what follows, the synthesis of the analysis is done in two ways: Multilevel Modelling and the statistically best factors are collected by using the Regression modelling. For analysis, grouping factors are changed to be so called Dummy variables when needed; for example, Ecological zone is transformed into three variables: variables indicative for Mountain, for Hill, and for Tarai.

Modelling the overall achievement by Multi-level Modelling

The datasets collected from sample schools are always clustered, that is, the students in the school are more alike with each other in comparison with the case that the same number of students would have been sampled totally randomly from the population. Multilevel modelling is used to acquire the correct test values while taking into account the clustering effect of the school. Table 3.1.34 shows the corrected estimates for the variables in table 3.1.33 while modelling the phenomenon in a multivariate manner; by using multivariate ANOVA, the hidden commonalities of the factors are revealed.

When taking into account the clustered structure in the dataset and the conjoint effect of the factors, quite many of the factors do not show main effect. Such variables are living in Hill zone and Central, Western or Far-Western region, school location, gender, students' age, and support provided through a private tuition. These factors could be omitted from the model explaining the differences in achievement between the students.

Table 3.1.34 Individual variables and their effects in Multilevel analysis

Source[1]	df_1	df_2	F	Sig.
Intercept	1	1838.8	354.80	< 0.001
Ecol zone Mountain Dummy (Mountain = 1, other = 0)	1	714.3	4.74	0.030
Ecol zone Hill Dummy (Hill = 1, other = 0)	1	723.1	0.01	0.913
Dev region Central Dummy (Central = 1, other = 0)	1	696.7	0.04	0.848
Dev region Western Dummy (Western = 1, other = 0)	1	693.4	2.57	0.109
Dev region Mid-Western Dummy (Mid-Western = 1, other = 0)	1	706.9	9.14	0.003
Dev region FarWestern Dummy (Far-Western = 1, other = 0)	1	713.4	0.96	0.327
Dev region Valley Dummy (Valley = 1, other = 0)	1	690.2	17.31	< 0.001
School location (0 = Rural, 1 = Urban)	1	660.3	0.24	0.626
School type (0 = Community, 1 = Institutional)	1	679.1	29.13	< 0.001
Gender (0 = girls, 1 = boys)	1	9838.5	2.20	0.138
Caste Brahman& Cheetri Dummy (Brahman&Chhetri =1,other = 0)	1	9965.2	105.56	< 0.001
Caste Janjati Dummy (Janjati = 1, other = 0)	1	9976.3	44.89	< 0.001
Caste Madhesi Dummy (Madhesi = 1, other =0)	1	10073.1	58.20	< 0.001
Caste Dalit Dummy (Dalit = 1, other = 0)	1	9967.7	23.31	< 0.001
Caste Other Dummy (Other = 1, other = 0)	1	9939.0	21.83	< 0.001
Language Dummy (Nepali = 1, other = 0)	1	10393.4	3.49	0.062
Home chores Dummy 1or2h (1 – 2 hours = 1, other = 0)	1	10042.5	8.50	0.004
Paid work Dummy (0 hours = 1, other = 0)	1	10059.1	19.68	< 0.001
Attitude ”Utility in Mathematics” (Sum, max 15)	15	10040.6	4.03	< 0.001
Age Dummy 8to10y (8 – 10 years = 1, other = 0)	1	9941.7	2.08	0.150
Help by Father Dummy (Father = 1, other = 0)	1	10105.3	3.60	0.058
Help by Mother Dummy (Mother = 1, other = 0)	1	10067.2	5.27	0.022
Help by Brother&Sister Dummy (Brother/Sister = 1, other = 0)	1	10091.5	8.12	0.004
Help by Tuition Dummy (Tuition = 1, other = 0)	1	10078.6	1.14	0.285
Help by No One Dummy (No one = 1, other = 0)	1	10041.2	7.65	0.006
Help by Teacher Dummy (teacher = 1, other = 0)	1	10099.0	3.01	0.083
Do you have a text book in Mathematics (Yes = 1, No = 0)	1	9964.9	7.92	0.005
Homeworks Given and Checked Dummy (everyday = 1, other = 0)	1	10100.7	8.98	0.003
Bullying (Sum, max 5)	5	9987.3	14.89	< 0.001
Positive Activities in school (Sum, max 8)	8	9951.5	11.05	< 0.001
SES[2] (Sum, max 6)	6	12423.2	15.27	< 0.001

1) *In Ecological zone and Development region, one of the classes needs to be omitted in the analysis because of singularity reasons. Tarai zone and Eastern region is omitted; they showed no statistical significance in Regression analysis.*
2) *Shortened SES; school type is taken away; this enables estimating the parameters for school type.*

Statistically the best factors by using Regression modelling

Traditional linear regression analysis with stepwise regression is used to explain the total score by the same above mentioned variables (see table 3.1.33). Table 3.1.35 shows the results.

Table 3.1.35 Statistically the best model of linear regression analysis explaining the student achievement (Method: stepwise)

Model	Coefficients				
	Unstandardized Coefficients		Standardized Coefficients		
	B	Std.Error	Beta	T	Sig.
(Constant)	20.73	1.852		11.19	< 0.001
School type 0 = Community,1 = Institutional)	8.27	0.585	0.17	14.13	< 0.001
Positive Activities in school (Sum; max 8)	0.18	0.014	0.12	12.97	< 0.001
Dev region Valley Dummy	9.52	0.616	0.17	15.46	< 0.001
Bullying (Sum; max 5)	-1.94	0.145	-0.12	-13.39	< 0.001
Dev region Mid-Western Dummy	-7.31	0.684	-0.10	-10.69	< 0.001
Caste Brahman& Cheetri Dummy	13.71	1.051	0.28	13.04	< 0.001
Attitude"Utility inMathematics"(Sum;max 15)	0.41	0.067	0.06	6.11	< 0.001
Language in home Dummy (Nepali =1, other = 0)	-3.18	0.491	-0.07	-6.48	< 0.001
Homework Dummy 1or 2h(1–2 hours = 1,other = 0)	2.99	0.449	0.06	6.66	< 0.001
Do you have a text book in Math (Yes = 1, No = 0)	4.98	0.928	0.05	5.37	< 0.001
SES[1] (Sum; max 6)	0.82	0.152	0.06	5.36	< 0.001
Homeworks Given and Checked Dummy	3.63	0.703	0.05	5.17	< 0.001
Paid work Dummy (0 hours = 1, other = 0)	1.90	0.472	0.04	4.03	< 0.001
Ecozone Mountain Dummy(Mountain=1,other = 0)	3.40	0.694	0.04	4.89	< 0.001
Caste Madhesi Dummy (Madhesi = 1, other = 0)	11.15	1.236	0.13	9.02	< 0.001
Caste Janjati Dummy (Janjati = 1, other = 0)	9.43	1.042	0.20	9.05	< 0.001
Caste Dalit Dummy (Dalit = 1, other = 0)	8.29	1.134	0.12	7.31	< 0.001
Caste Other Dummy (Other = 1, other = 0)	7.60	1.283	0.08	5.92	< 0.001
Dev region Western Dummy(Western = 1, other = 0)	1.66	0.574	0.03	2.89	0.004

The model in table 3.1.35 can be interpreted as follows: The average mean of the students is 20.7% of the maximum score – assuming that the student was in the lowest group in all the factors. If the school was an institutional one (value = 1), the students' score was, on average, + 8.3% higher (note the sign of the coefficient). Additionally, if the students was a Brahmin or Chhetri, the additional score was 11.1 percent; if they came from the Valley, they gained + 9.5 percent more; if having the textbooks in their use, the score was 5.0 percent higher than with only two or less evaluations per week. If, on the other hand, the student came from the Mid-Western region, the expected achievement level was 7.1 percent lower than if coming from the other regions. Also, if they face bullying, with each step (of five), the achievement level would drop 1.9 percent; the difference between the lowest and highest group is 5 × 1.94 = 9.7 percent.

3.2 Assessment Results of Mathematics in Grade 5

This section analyses the assessment results of grade 5 Mathematics. It starts with analysing basic results including overall distribution of scores, achivement scores in various content areas of mathematics in general and analyses the effects of various diversity factors from equality point of view. It then analyses the influences of factors explaining the differences in mathematics achievement.

3.2.1 Basic Results of Assessment in Mathematics for Grade 5

As the basic results of assessment in Mathematics, this sub-section analyses the overall distribution of scores, result in the various content areas and various levels of cognitive domains of mathematics, results variations based on item types, and comparison of results with previous assessments and international assessment results.

Overall distribution of score

The mathematics sample was big enough to form a normal distribution (13,714 students). Though there are at least two populations which are, however, quite close to each other, and therefore, the distribution is quite close to the normal shape.

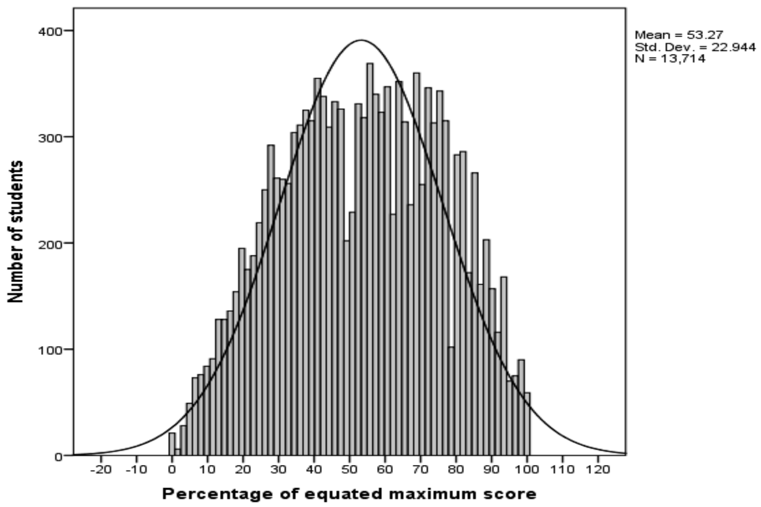

Figure 3.2.1 Distribution of overall results

A closer look to the distributions shows two different, normal, populations in the dataset: students from community schools and students from institutional schools (fig. 3.2.2).

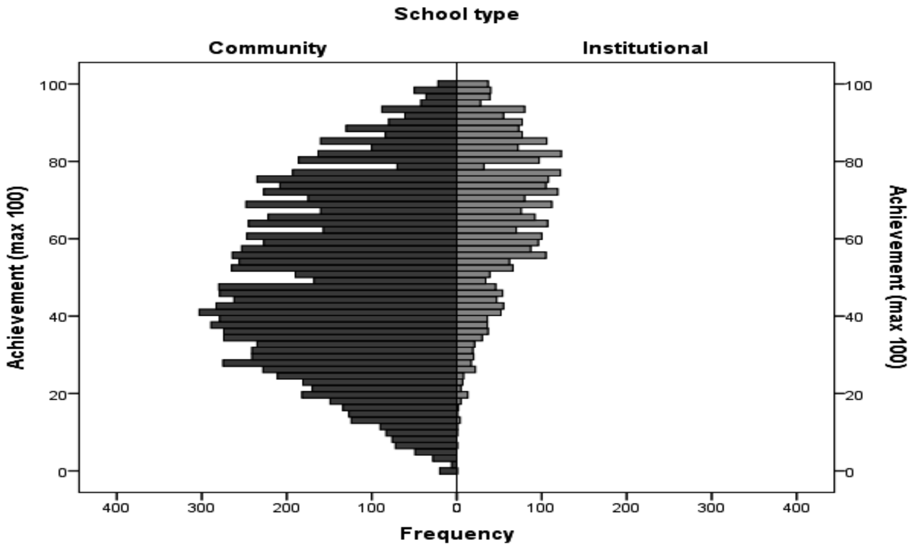

Figure 3.2.2 Distribution of the students' achievement scores by school type

In figure 3.2.2, the left-hand side distribution shows students in community schools whereas the right-hand side is the distribution of institutional school students. The main system is shifted towards the lower performing level because the main population comes from the community schools. The main population in institutional schools performs quite well. There are two normal populations, one for community schools and another for institutional schools. One may note also the long tail of the low-performing students in institutional schools. One also notices that that there are quite a number of students in community schools getting equally high marks as the students from institutional schools gain. On the basis of figure 3.2.2, it is evident that the students in community schools are varying from the low-performer to the highest performer whereas most of the students from institutional schools are performing high or medium.

By analyzing the data further with the scatter plot, and combining the socio-economic status (SES) with the average achievement in the school, figure 3.2.3 shows that two types of schools fall into two groups: most of the institutional schools (triangle) are performing very well and the average SES is very high. The community schools (circle) vary from very high-performing to very low-performing.

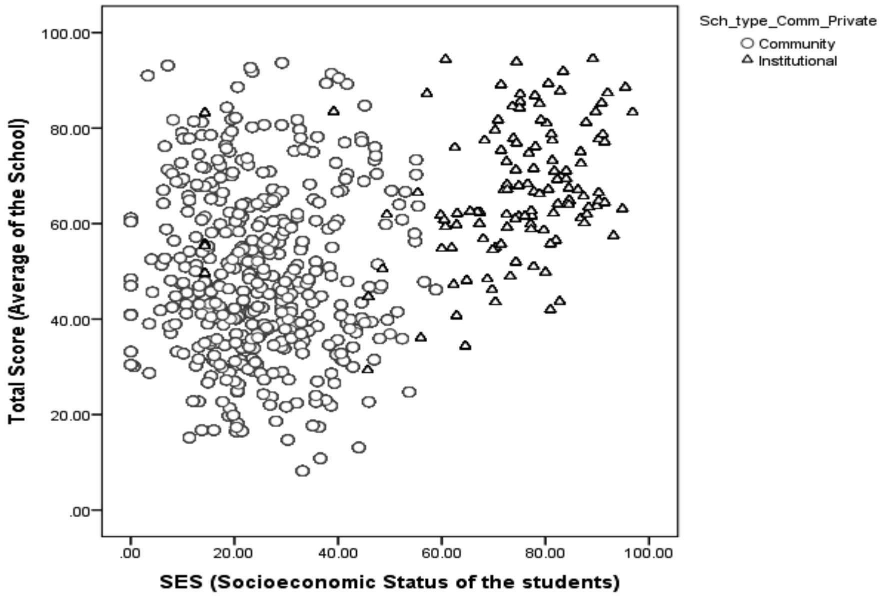

Figure 3.2.3 Achievement and Socio-economic status by types of school

The dataset shows that the 5th graders' Mathematics population is more normally distributed than population for other subjects. However, the population of the students is found divided into two groups: low and high performing students from community schools and high performing students from institutional schools with a long tail of low performing students. The variance in the community schools is remarkable.

Achievement in various content areas

The whole Mathematics test was a combination of four content areas: Arithmetic, Algebra, Geometry and Numeracy. In curriculum, more weightage is given to Arithmetic skills compared to the others. To compare the achievement in all the topics, these sub-scores are converted into a percentage. Table 3.2.1 shows the students' achievement in Mathematics as a whole and the achievement level by each four content areas; figures 3.2.4 and 3.2.5 illustrate the differences.

Table 3.2.1 Achievement score in various content areas

Content area	Mean	SD	Min	Max
Arithmetic	54.5	23.3	0	100
Algebra	49.4	21.3	0	92
Geometry[18]	56.9	23.4	0	100
Numeracy	44.1	31.3	0	100
Math as Total	**53.3**	**22.9**	**0**	**100**

[18] The mean score of geometry seems to be slightly higher than the mean of the original latent Theta value. The differences between the Theta of arithmetic (– 0.60) and geometry (– 0.73) indicates that the equated mean should be notably lower than that of arithmetic. The reason for this is technical: In the equating process the (accurate) Theta values are transformed into (less accurate) scores. In some cases the equated scores are rounded up which causes higher mean in scores than in Theta.

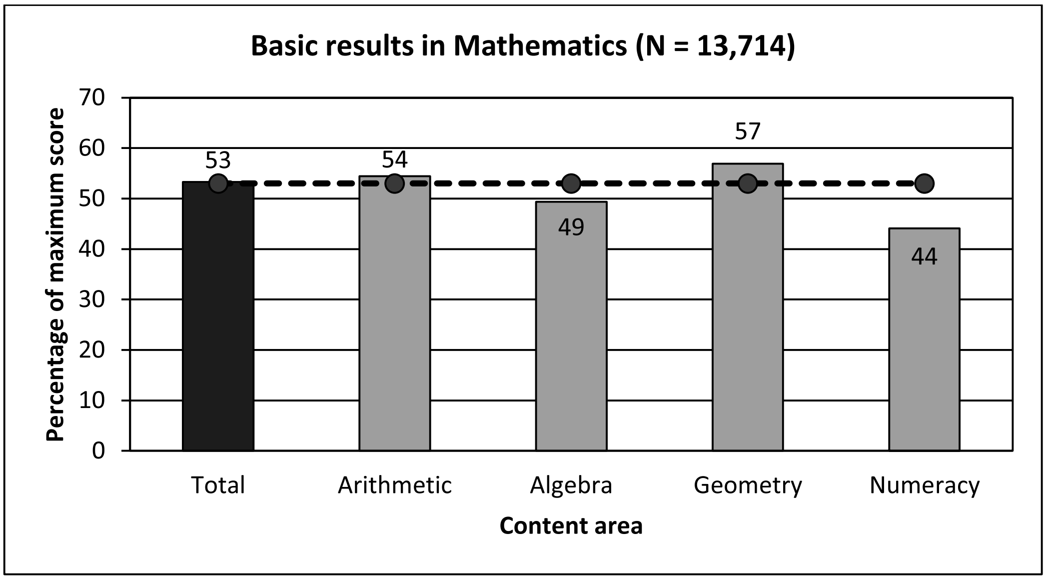

Figure 3.2.4 Basic results in various content areas

The percentage of achievement score explains that the national average of Mathematics is 53%. Of the different content areas, students are somehow weaker than the average in Numeracy (44%) and Algebra (49%). They perform better than average in Geometry (57%), and Arithmetic (54%).

Because of the difference in the average level between community and institutional schools, it is interesting to know whether there is proportional difference in the content areas between the students. Table 3.2.2 and figure 3.2.5 illustrate the differences.

Table 3.2.2 Achievement in various content areas by the school type

Content area	Community schools (N=10,657)			Institutional schools (N=3,057)		
	Mean	**SD**	**CV**	**Mean**	**SD**	**CV**
Arithmetic	50.5	23.1	45.8	68.2	18.4	27.0
Algebra	46.5	21.1	45.3	59.3	18.9	31.8
Geometry	53.8	23.2	43.1	67.7	20.5	30.4
Numeracy	39.1	30.3	77.5	61.5	28.5	46.3
Total[1]	**49.3**	**22.5**	**45.7**	**67.0**	**18.6**	**27.7**

1) *Note that the total score is not the mean of the various content areas because it has been equated independently from them.*

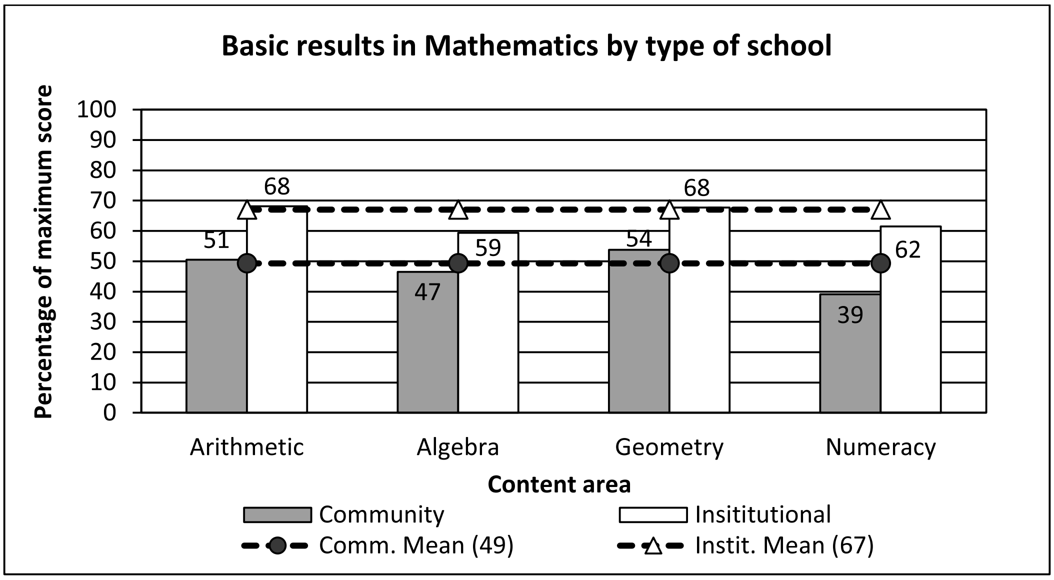

Figure 3.2.5 Basic results in various content areas by the school type

It is also evident that the differences between the content areas are wider in community schools compared to institutional schools. While, in community schools, the difference between Geometry (54%) and Numeracy (39%) is 15 percent, in institutional schools it is only 6 percent. Partly, though less probably, this can be explained by ceiling effect in the institutional schools; the test might have been too easy for the best students in the institutional schools (see figure 3.2.2). Hence, the best students were unable to show how far they could have been able to rise.

The learning achievement of grade 5 in Mathematics is 53%. Among the different content areas, students performed lower than the averave achievement in Numeracy (44%) and Algebra (49%). They preform better than the average in Geometry (57%) and Arithmatic (54%). The differences between the content areas are wider in community schools than in institutional schools.

Achievement in various levels of cognitive domain

Mathematics test as a whole was constructed based on Bloom's taxonomy of cognitive levels (Bloom *et al.,* 1956; Metfesser, Michael, & Kirsner, 1969), that is, *knowledge*, c*omprehension*, *application*, and *higher ability* (reasoning/problem solving). The achievement of the students on various levels of cognitive domain is shown in table 3.2.3 and illustrated in figures 3.2.6.

Table 3.2.3 Achievement level on various levels of cognitive domain

Hierarchical level	Mean	SD	Min	Max
Knowledge	64.6	23.36	0	100
Comprehension	58.8	23.28	0	100
Application	51.8	24.12	0	100
Higher ability	40.1	31.60	0	100

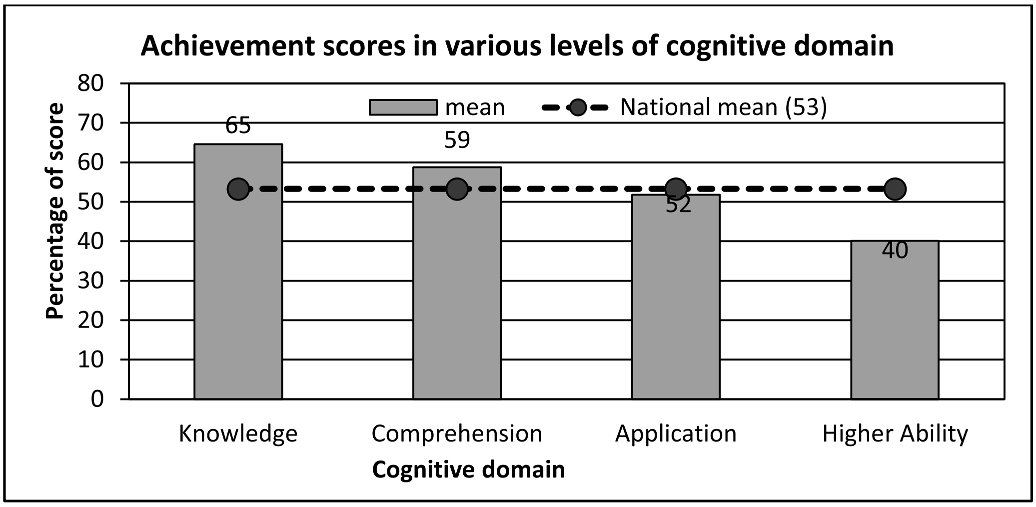

Figure 3.2.6 Achievement at various levels of cognitive domain by school type

Remarkably a high number of students were able to solve only a couple of practical problems, that is, 5 out of 26 marks of the application type of items (12% of the students). Other 40% students gained just one mark out of 5 in the tasks requiring the higher cognitive abilities and 20% of the students did not solve any of these problems.

Because of the difference in the average level between community and institutional schools, it might be interesting to know whether there is proportional difference in various levels of cognitive domain among the students. Table 3.2.4 and figure 3.2.7 illustrate the differences.

Table 3.2.4 Achievement in various levels of cognitive domain by school type

Content area	**Community schools (N= 10,657)**			**Institutional schools (N =3,057)**		
	Mean	**SD**	**CV**	**Mean**	**SD**	**CV**
Knowledge	61.5	23.8	38.8	75.4	17.8	23.7
Comprehension	55.2	23.4	42.4	71.1	17.8	25.1
Application	47.3	23.4	49.4	67.1	20.0	29.7
Higher ability	37.0	30.6	82.6	50.8	32.7	64.4
Total	**49.3**	**22.5**	**45.7**	**67.0**	**18.6**	**27.7**

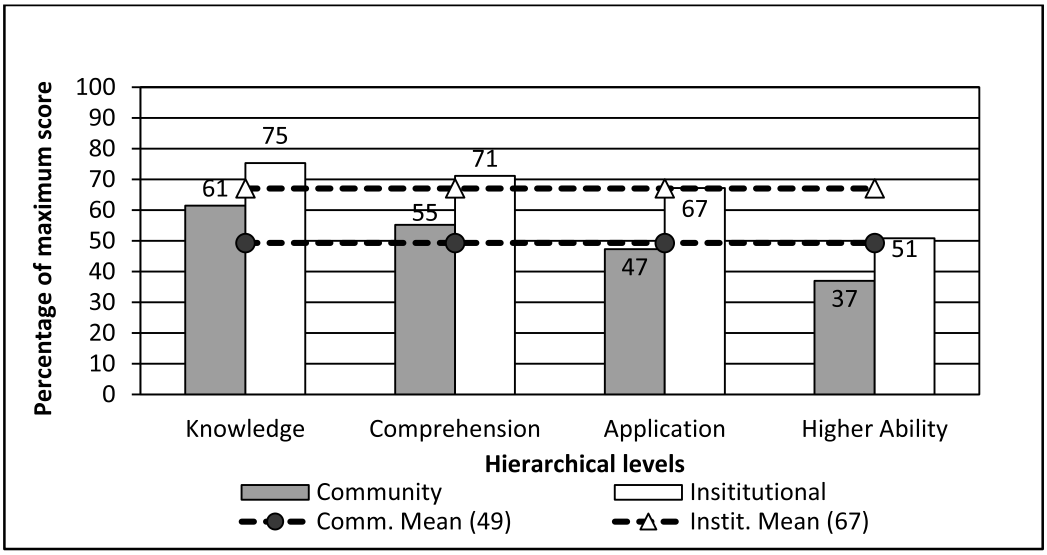

Figure 3.2.7 Achievement in various levels of cognitive domain

The main trend is obvious in the community schools as well as institutional schools – in both cases the students are much better in the recall type of questions than in the tasks requiring higher skills. However, there is another noteworthy tendency. Except for the higher ability type of questions,a trend is seen that the difference between the school types increases systematically with the more complex problems the students need to solve (14 – 16 – 20 for Knowledge, Comprehension, and Application respectively).[19] In both type of schools, the difference between the scores of knowledge type and higher ability type of question is equal (24 percent), though. Statistically, the tendency is seen in the effect size: though the differences are remarkable in any case, the difference is smaller in the area of knowledge ($d = 0.61$) than in application type of items ($d = 0.87$). This means that, for one reason or another, the students in institutional schools are found to be more able to solve complex problems in relation to the simple tasks than their peers in community schools.

One detail may be worth noting regarding table 3.2.5: the exceptionally high value of Coefficient of Variation within the community schools in higher ability (83%). The moderate standard deviation (30.6%) with very low mean (37%) means that, within the community schools, there are also reasonably highly performing students.[20]

The dataset shows that the students' ability to solve complex problems is quite low; only 37% of the maximum scores of tasks requiring higher ability were reached. Students are much better in the recalling type of questions (65%). Remarkable number of students (20%) was not able to solve any of the tasks requiring higher ability. The students in

[19] The similar type of tendency is there when analyzing only the students from the Kathmandu Valley: the differences are 4 – 7 – 9 indicating, first, that the differences are minimal between community and institutional schools, second, the difference between community and institutional schools is narrow compared to the national level.

[20] In the community schools of the Kathmandu Valley, the mean of higher ability items is 53 (CV = 31.7). Hence, the CV is not exceptionally high compared to the CVs of the other cognitive levels in the total sample. Both results, the mean and CV indicate, whichever the school type, that the Valley students perform better and the variation is low.

institutional schools seem to be more able to solve complex problems than their peers in community schools.

Type of item and achievement

There were basically two types of questions 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 was needed to get a 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 were made in most of the hierarchical levels (knowledge, comprehension, application, and higher ability) and a wide range of difficulty levels though the subjectively scored items tended to be more demanding the higher cognitive level. Table 3.2.5 comprises the basic statistics of the item type-wise achievement levels.

Table 3.2.5 Mean score by item type

Types of items	Mean	SD	Minimum	Maximum
Objective	62.2	20.7	0	100
Subjective	44.1	26.4	0	100

It is obvious that the subjective items – usually with more demanding requirements for the correct answer – are solved much lower (49%) than the objective items (62.2%). Most of the objective items were knowledge, comprehension and application type whereas subjective items were application and higher ability type. Though the differences between community and institutional schools are, in any case, wide, the effect size is notably higher when it comes to solve the productive type of items (d = 0.87) compared to objective type of items (d = 0.56).

Dataset indicates that the students are performing well in recognizing the correct answer and in recalling simple facts from the texts, fundamental thinking, the basic interpretation of paragraph, table, and chart, and a few steps of logical thinking. They are much weaker in solving verbal mathematical problems, calculating the answer in more than one step, constructing the geometrical shapes, solving the linear equation of one variable etc. In many cases, the students started to do such problems but the skills were not high enough for the highest marks.

Comparison of achievement with previous datasets

The datasets of previous Mathematics assessment are, however, somehow fragmented and obtained by using various strategies for sampling which makes the comparison difficult.[21] The previous datasets also carry two other challenges that hinder the comparison with the

[21] For example, CERID (1993) was sampled exclusively in three districts in the Kathmandu Valley. BPEP (1994) was also sampled in three districts though thay were not in the Valley: Tanahu, Kapilvastu, and Chitwan. CERID 1998 was sampled in five districts covering all Development regions (one district from each Development region) and Ecological zones. EDSC (1999) was sampled in 20 districts covering all Development regions and Ecological zones. This deviance in samplings makes the comparison somehow difficult.

present dataset. First, the National Assessment of grade 5 students carried out by the Basic and Primary Education Project (BPEP, 1994: 10) shows that the national average of the students was 40%. Later, in 1999, the results showed that the national average of the students was around 27% (EDSC, 1999: 34). The National Assessment in 2012 conducted by the ERO shows that the national average of Mathematics in grade 5 is 53%. These figures are coming from Classical Test Theory and, unfortunately, they are not comparable with each other because of lack of a proper linking procedure. The differences between the scores can easily be explained by the various difficulty levels of the tests. Second, the previous datasets of grade 5 are not available and, hence, any IRT modeling-based procedures for comparison could not be made.

Though the comparison cannot be made in the absolute sense, proportional comparisons can be made, with cautions, on the basis of the previous results. The proportional differences are presented in tables 3.2.6 to 3.2.11.

Table 3.2.6 Comparison between the achievement of girls and boys in 1999 and 2012

Indicator	1999 (EDSC, 1999)		NASA 2012[1]	
	Boys	**Girls**	**Boys**	**Girls**
Mean	29.3	24.5	54.4	52.7
SD	17.3	16.5	22.5	23.1
CV	59.2	67.3	41.3	43.9
N	1,864	1,646	6,570	6,302

1) Some students did not inform their sex category.

Compared to the 1999 dataset, the difference between boys and girls has narrowed somehow radically though boys still outperform girls (less than 5 percent in 1999 and 2 percent in 2012). Though in both years the difference was significant ($p < 0.001$), in 1999, the difference between boys and girls was of medium size (Cohen's $d = 0.28$) but, in 2012, this difference turned to be non-existent (Cohen'sd = 0.07). This is a good sign from equality point of view.

Table 3.2.7 Comparison of achievement by Ecological zones in 1999 and 2012

Indicators	199 (EDSC, 1999)			NASA 2012		
	Mountain	**Hill**	**Tarai**	**Mountain**	**Hill**	**Tarai**
Mean (%)	27.7	24.9	28.5	55.8	49.2	49.1
SD	15.2	15.3	18.3	22.9	22.0	22.8
CV	54.8	61.6	64.3	41.1	44.7	46.3
N	344	1,642	1,291	1,449	6,235	3,603

1) Both datasets exclude the Valley.

Compared with the 1999 dataset, the differences between Ecological zones are found to have widened slightly. The difference between the students in the Mountain and Hill zones has risen from a small effect size (d = 0.18) to medium one (d = 0.30), between the Mountain and Tarai from a non-existent effect size d = 0.05 to a medium one (d = 0.29), and between Hill and Tarai the difference has reduced from a small effect size ($d = 0.21$) to non-existent one ($d = 0.00$). The Coefficient of Variation (CV) shows that the means of students have risen in all the regions more than the variance. This is a good sign from equality point of view.

Table 3.2.8 Comparison of achievement among Development regions in 1998 and 2012

Indicator	1998 (EDSC, 1999)					NASA 2012[1]				
	Eastern	Central	Western	Mid-Western	Far-Western	Eastern	Central	Western	Mid-Western	Far-Western
Mean (%)	31.4	30.1	24.3	21.3	26.8	44.7	50.9	56.8	45.5	50.8
SD	19.0	19.9	13.7	11.8	15.1	23.2	22.2	21.6	19.2	23.2
CV	60.4	66.1	56.5	55.3	56.2	52.0	43.7	38.0	42.2	45.6
N	802	932	1,018	465	293	2,146	3,161	2,134	1,587	2,259

1) *Students from Kathmandu are excluded.*

When comparing the Development regions, two things should be kept in mind: First, in the dataset from 1999 (EDSC, 1999), few districts were taken from each Development region, and second, the Valley was not included from the Central Development region. However, the datasets can be considered comparable and taken the Western region (with the highest score in 2012) as a reference point. It is notable that the students in the Western Development region have raised their position from the second lowest after the Mid-Western region in 1999 to the highest in 2012. In 1999, the students in the Eastern region were the highest achievers but, in 2012, they are placed at the lowest level; the effect size has changed from $d = +0.46$ to $d = –0.44$ – indicating that the difference is in the same (medium) size, but now it is lower instead of higher than that of the Western region. Mid-western region has remained in the low position but the Far-Western region has risen at the same level as the Central Development region. The values of the Coefficient of Variance (CV) have lowered remarkably – indicating that, compared with the standard deviations, the means have risen remarkably. This is a good sign. However, the falling position of the students from the Eastern region is, naturally, not a good sign from the equality of opportunity point of view.

Table 3.2.9 Comparison of achievement by the location of the school in 1998 and 2012

Indicator	1998 (EDSC,1999)		NASA 2012	
	Rural	Urban	Rural	Urban
Mean (%)	27.3	27.1	48.6	58.7
SD	17.9	14.2	22.2	21.5
CV	65.4	52.4	45.7	36.6
N	2,701	809	9,105	1,382

Compared to the 1999 dataset by the location of the school, it is evident that the world has changed radically during the 15 years. While, in the 1999 dataset, the difference between the rural and urban schools was very low ($d = 0.013$) and the difference was not significant, in the 2012 dataset the difference is remarkable ($d = 0.46$), favoring more the urban schools even though the Valley schools are not included in the comparison.[22] The change in the phenomenon is remarkable. It most probably indicates the rise of the institutional schools in the urban areas. The values of the Coefficient of Variance (CV) have lowered remarkably indicating that, compared with the standard deviations, the means have risen remarkably. This is a good sign from equality perspective.

Table 3.2.10 Comparison of achievement by the school type in 1999 and 2012

Indicators	1999[1] (EDSC, 1999)		NASA 2012[1]	
	Community	Institutional	Community	Institutional
Mean (%)	36.3	42.1	64.1	71.2
SD	20.4	16.6	20.3	17.1
CV	56.2	39.3	31.7	24.0
N	**233**	**242**	**971**	**1,456**

1) Comparison includes only students from the Kathmandu Valley.

The results of community and institutional schools are compared on the basis of the 1999 dataset (EDSC, 1999, 36). The 1999 dataset was collected in the Kathmandu Valley. It is seen that the difference between the community and institutional schools in the Valley has been widened somehow, though not radically (Cohen's d = 0.31 in 1999 and d = 0.39 in 2012). Practically speaking, the difference between the school types has stood a medium size within the 14 years. Based on the variance in the same kind of schools, it has been noted that CV has been reduced, especially in the community schools (from 56 to 32%). This indicates the homogeneity in the ability of the students compared to the 1999 result. This is a good sign from equality point of view.Table 3.2.11 summarises the information provided in tables 3.2.6 to 3.2.10.

Table 3.2.11 Summary of comparison of achievement in 2012 to the previous datasets

	Sex	Selected background variables			School type
		Ecological zone	Development region	School location	
Main finding	Boys still over-perform girls mildly but the difference is reduced radically.	Differences have increased moderately. Students in the Mountain zone score higher while in Tarai they score lower.	Students in the Western and Far-Western regions score higher while in the Eastern region, they score remarkably lower.	Remarkable rise in performance within the Urban schools	No remarkable change in the difference between community and institutional schools.

Compared to the 1999 results, the gap between boys and girls has reduced, the students in the Mountain zone and Western and Far-Western regions score remarkably higher in 2012 but in Tarai zone and Eastern region the results are lower. Dataset from 14 years back

[22] If we include the Valley schools to the analysis, the effect size will be d = 0.67.

indicates that the difference in Mathematics achievement between the students in the community and institutional schools in the Kathmandu Valley has not changed.

Comparison of achievement with international standard

The NASA 2012 scores of grade 5 Mathematics were made comparable with the international TIMSS assessment scale of grade 4. Six items were used as linking items from the international item bank. Appropriate items of TIMSS were selected so that they fit with the Nepalese curriculum, and items are familiar to the Nepalese students.[23] 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 2012 is zero, it corresponds to the average level of the international students of grade 4 in TIMSS. Note that the most accurate comparison would be expected if the items were chosen randomly from the bank and the items were taken from grade 5 TIMSS item bank; however, TIMSS assesses only grade 4 students, not grade 5.

Figure 3.2.8 shows the comparison of the students' achievement with the international standard. In the figure, x-axis shows the content areas of mathematics and y-axis shows the ability shown by the students. The middle horizontal line indicates the international average. When the ability is above the average (0.00, the international average line), the bars are going up whereas when the ability would be below the international average, the bars would have been going downward.

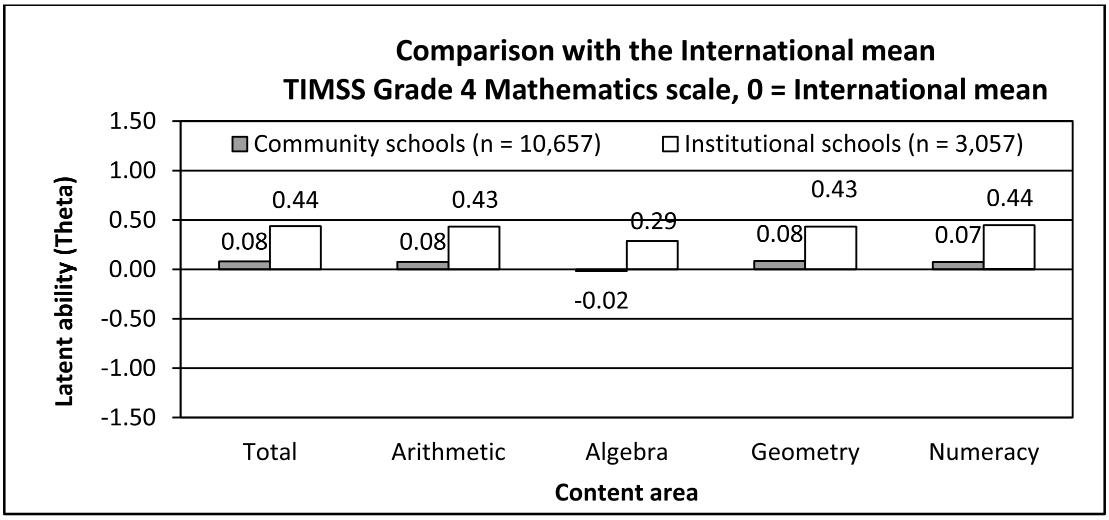

Figure 3.2.8 Student achievement from NASA 2012 grade 5 in international TIMSS scale

Figure 3.2.8 shows that, in community schools, the 5th graders are seen to be at the same level as the 4th graders in the international dataset. In institutional schools, the level is remarkably higher (θ = 0.436). It is seen that the average ability shown by Nepalese students in Algebra is the lowest of the all content areas. The achievement level of an average grade 5 student in the whole dataset (θ = 0.159) is found to be higher compared

---

23 This principle is the same as used in NASA 2011 with grade 8 (ERO, 2013). This causes, most probably, that the results are better than if selected the items randomly from the international item bank (see discussion in NASA 2011 report).

with an average international student of grade 4. It is good to remember three things. First, all the linking items came from the content areas of Numeracy, Arithmetic and Geometry, and hence there actually is no real equating in algebra. Second, the difficulty level of the items was not best suiting to the 5th graders' Mathematics in Nepal (TIMSS items were intended for the 4th graders). Third, the items were selected to be familiar type for the students, and hence the results are better than if selected randomly. Hence, the higher proficiency is expected. From this point of view, the students in the community schools perform unexpectedly low.

While looking at the dataset from the international pont of view, the average mathematical ability of the grade 5 students in Nepal is found about the same as of the grade 4 students in the international dataset. The students in community schools perform unexpectedly low.

3.2.2 Results Based on Diversity Factors

Several diversities can be found in Nepalese society. However, considering their relevance for Nepalese context, NASA 2012 background information questionnaire included geographical/ecological, language, gender/sex, ethnic/caste and economic diversities. Additionally, district-wise, school type-wise, and their location-wise (rural/urban) diversities are also taken into consideration while analysing the results. These factors have been taken as equality factors, since all the children regardless of their sex, language, birth place, or family background should have equal opportunities to reach the same educational goal.

District variations in student achievement

While looking at the district-wise results, one can notice the variations in achievement among the districts from the low to the high performing. It is also good to keep in mind that there may be lower or better performing districts within those not selected in the sample. The district-wise differences are depicted in table 3.2.12 and figure 3.2.9. The table shows the achievement in descending order according to the achievement scores. The mean represents the average achievement percentage of the particular district.

Table 3.2.12 Average achievement score in the sample districts

Districts	N	Mean	SD	CV	Districts	N	Mean	SD	CV
Kathmandu	1,551	71.0	17.6	24.8	Chitwan	590	50.5	21.3	42.2
Bhaktapur	374	68.7	20.96	30.5	Baglung	574	49.3	19.6	39.8
Kaski	722	64.4	21.3	33.0	Saptari	453	47.3	21.3	44.9
Humla	127	60.4	23.4	38.7	Jumla	157	47.1	23.9	50.7
Lalitpur	502	59.8	18.1	30.3	Kailali	739	47.0	21.9	46.6
Solukhumbu	288	58.9	24.3	41.3	Dhankuta	374	45.4	22.1	48.6
Baitadi	597	57.9	23.4	40.4	Salyan	503	45.0	17.2	38.3
Parsa	448	57.8	20.4	35.2	Sindhuli	591	44.6	20.9	46.9
Dolakha	448	56.1	20.8	37.1	Achham	512	44.4	21.9	49.2
Myagdi	271	55.4	16.6	29.9	Rolpa	425	44.4	16.4	36.9
Kapilbastu	549	55.4	23.0	41.5	Mahottari	449	42.8	27.4	64.1
Darchula	411	55.3	23.1	41.8	Bardiya	375	41.5	18.4	44.4
Makwanpur	635	54.1	18.8	34.8	Khotang	508	41.2	22.3	54.2
Manang	18	50.8	16.5	32.4	Udayapur	523	37.4	22.0	58.7
Total						**13,714**	**53.3**	**22.9**	**43.1**

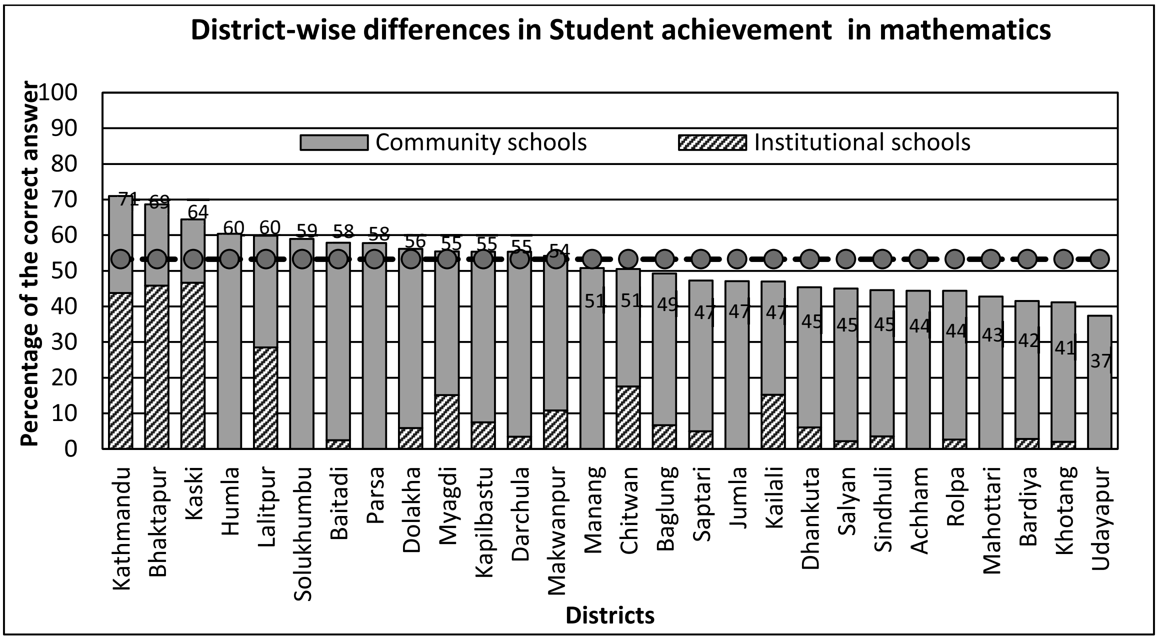

Figure 3.2.9 Average achievement in districts by community and institutional schools

Of the total sample, student performance was very low in Udayapur (37%), Khotang (41%), and Dhankuta (45%) from the Eastern region; Sindhuli (45%) and Mahottari from the Central region; Bardiya (42%), Salyan (45%), and Rolpa (44%) from the Mid-Western, and Achham (44%) from Far-Western region. Out of the five highest performing districts, three are from the Valley: Kathmandu (71%), Bhaktapur (69%) and Lalitpur (60%). Remaining two districts are Kaski (64%) from Western region and Humla (60%) from Mid-Western region. Comparison may be unfair because the 74% of the schools in the Kathmandu Valley are private ones and in Kaski 70% were private, while in the other districts in the sample, on average, only 10% were private ones. From this perspective, interesting districts are those where the number of institutional schools is low but the results are higher than the national average. Some examples of these districts are Humla, Solukhumbhu, Baitadi, and Parsa.

The difference in achievement due to the district is statistically significant ($p < 0.001$). The variation explained in achievement due to the district is $\eta^2 = 0.177$, that is, the district explains 18% of the variation in the data which is a very high percentage. Effect size is $f = 0.46$ indicating that the difference between the lowest performing district (37%) and highest performing district (71%) is remarkably high.

The dataset suggests 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 mathematics. The results are bound to the 28 sample districts; even lower-performing districts could be found if other districts would have been selected. The results are very high in the districts where the proportion of institutional schools with high socio-economic status is high. Interesting districts are those which have demonstrated almost equally high results even with no institutional schools in the sample, which are Humla, Solukhumbhu, Baitadi, and Parsa.

Ecological zone and student achievement

Similar to the district-wise variations, achievement varies from one Ecological zone to another. Among the four zones, the Valley stands on the top with 67.0% and the Tarai lags farthest behind of all with just 59.5% mean score. The variation in achievement by Ecological zones is presented in table 3.2.13.

Table 3.2.13 Achievement in various Ecological zones

Ecological zone	Community schools				Institutional schools			
	N	Mean	SD	CV	N	Mean	SD	CV
Mountain	1,368	55.3	23.0	41.6	81	64.0	20.3	31.7
Hill	5,256	46.2	21.3	46.2	979	65.4	18.1	27.6
Tarai	3,062	47.3	22.7	48.0	541	59.5	20.2	34.0
Valley	971	64.1	20.3	31.7	1,456	71.2	17.1	24.0
Total	**10,657**	**49.3**	**22.5**	**45.7**	**3,057**	**67.0**	**18.6**	**27.7**

The data shows first that, on average, the students from the Valley outperform the students from the other Ecological zones. The difference is wider in community schools (46% to 64%) than in institutional schools (60% to 71%). In community schools, the students from Hill and Tarai zone perform the lowest (46% and 47% respectively) and in institutional schools, the students from Tarai area are found to be performing the lowest (60%). One may note also the low value for the Coefficient of Variation in the Valley in institutional schools indicates the high result with the lower variation compared to other zones and with community school.

The achievement in the zones differs significantly in both the schools types ($p < 0.001$) even if the Valley is excluded from analysis. Tukey's *post hoc* test tells that, in community schools, there is no difference between Hill and Tarai, but the students from Mountain zone differ from the students of both Hill zone ($p < 0.001$) and Tarai zone ($p < 0.001$). In institutional schools, there is no convincing difference between Mountain and Tarai but the students from Hill zone differ from the students of Tarai zone ($p < 0.001$). Ecological zone explains 6% of the variance in both the community- ($\eta^2 = 0.06$) and institutional schools ($\eta^2 = 0.056$).[24] As a comparison, remember that the district explains more than 18% of the variation. The effect size is $f = 0.25$ in community schools and $f = 0.24$ in institutional schools – showing moderate difference between the highest and lowest performing Ecological zones. The effect sizes are small if the Valley is excluded from analysis ($f = 0.14$ and $f = 0.15$ respectively). This means that the real differences are not remarkable between the Ecological zones but the Valley differs radically from the other areas. From the equality point of view, this can be taken as a possible good sign.

[24] If the Valley is excluded from analysis, the values for Eta squared (η^2) would be 0.019 and 0.021 respectively, that is, only 1% and 2% explanation. The impact of the KathmanduValley students in the whole national mean is remarkable.

Dataset reveals a moderate difference between the student performances in four Ecological zones in both community and institutional schools. Students in the Kathmandu Valley outperform the other students. The achievement is the lowest in Hill and Tarai zones compared to Mountain and the Valley zones.

Development region and student achievement

Student achievement varies according to the Development regions which are divided into Eastern, Central, Western, Mid-Western, and Far-Western. Additionally, the Kathmandu Valley is taken as the 6th Development region though administratively it falls under the Central Developmental region. The mean achievements within the Development regions are given in table 3.2.14.

Table 3.2.14 Achievement in various Development regions

Development	**Community schools**				**Institutional schools**			
Region	N	Mean	SD	CV	N	Mean	SD	CV
Eastern	2,041	44.4	23.4	52.6	105	49.9	19.7	39.5
Central	2,715	48.8	22.3	45.7	446	63.3	16.8	26.6
Western	1,387	50.3	20.5	40.8	747	68.7	18.1	26.4
Mid-Western	1,528	45.1	19.3	42.7	59	53.6	14.0	26.2
Far-Western	2,015	50.3	23.4	46.6	244	55.2	20.4	36.9
Valley	971	64.1	20.3	31.7	1,456	71.2	17.1	24.0
Total	**10,657**	**49.3**	**22.5**	**45.7**	**3,057**	**67.0**	**18.6**	**27.7**

The highest performances are found in institutional schools in the Kathmandu Valley (71%) and in Western region (69%). The performance is the lowest in community schools in Eastern (44%) and Mid-Western (45%) regions. The difference between the regions is statistically significant both in community and institutional schools ($p < 0.001$). Tukey's *post hoc* test shows that, in the community schools, the average achievement level in Eastern region is significantly lower than in any other regions ($p < 0.001$) except for Mid-Western region, and in the Valley the achievement is highest of all other regions ($p < 0.001$). Also, in Mid-Western region the average achievement level is lower than in other region except for the Eastern region ($p < 0.001$). There is no difference between Central, Western, and Far-Western regions when it comes to the achievement level in Mathematics. In the institutional schools, Tukey's *post hoc* test shows that the students in the Valley outperform the students in all other regions ($p < 0.001$) and in the Eastern region the students perform lower than the other regions ($p < 0.01$).

Development region explains 5% of the variance in community schools ($\eta^2 = 0.054$) and 10% in institutional schools ($\eta^2 = 0.103$).[25]This is somehow the same proportion as found

[25] If the Valley is taken out of the analysis, the values for Eta squared would be 0.012 and 0.104 respectively, that is, 1% and 10% of the explanation. It is notable that the figure for the community schools is around one fifth of that with the Valley included in the analysis but there is no difference when it comes to the institutional schools. This means that the role of the Kathmandu Valley students in the whole national mean of grade 5 mathematics is remarkable when it comes to the community schools. The means in the institutional schools are not that much dependent on the students in the Valley.

with the Ecological zone. One remembers that the district explains more than 36% of the variation which means that within the Developmental regions there are lower and higher performing districts.The effect size is f = 0.24 in community schools and f = 0.34 in institutional schools – showing moderate or wide difference between the highest and lowest performing regions. The effect sizes are moderate if the Valley is excluded from the analysis (f = 0.11 and f = 0.34 respectively). Compared to the Ecological zones, the differences are wider between the Development regions.

The dataset reveals that there is an inequality across the Development regions for children's opportunities to reach an adequate level in Mathematics. As the widest, the difference is between the Valley and Eastern region (20% variation in community schools and 21% in institutional schools). This wide difference indicates a strict sign of inequality in opportunities in learning Mathematics.

School type and student achievement

All the schools are categorized into community and institutional (private schools). The differences in Mathematics achievement have been handled in the sections above. Here the main differences are presented in table 3.2.15.

Table 3.2.15 Type of school and the average achievement in various content areas

Content area	Community (N=10,657)			Institutional (N=3,057)			Mean difference	Cohen's *d*
	Mean	**SD**	**CV**	**Mean**	**SD**	**CV**		
Arithmetic	50.5	23.1	45.8	68.2	18.4	27.0	17.6	-0.80
Algebra	46.5	21.1	45.3	59.3	18.9	31.8	12.8	-0.62
Geometry	53.8	23.2	43.1	67.7	20.5	30.4	13.9	-0.61
Numeracy	39.1	30.3	77.5	61.5	28.5	46.3	22.4	-0.75
Total	**49.3**	**22.5**	**45.7**	**67.0**	**18.6**	**27.7**	**17.7**	**-0.82**

The achievement levels in the community schools and institutional schools differ from each other remarkably as noted above. The average performance in total score in the private schools is 67.7%, whereas in the community schools it is 53.8%; thus, 18 percent difference is remarkable. The difference is statistically significant (p < 0.001) and the effect size is very high (d = 0.82) – showing that community schools are far below institutional schools. From the effect size viewpoint, the difference is the highest in the content area of Arithmetic (18 percent, d = 0.80) though the absolute difference is the highest in the content area of Numeracy (22 percent, d = 0.75); for some reason, the variance is also highest in Numeracy. Division of the students into the community and institutional schools explains 10% of the student variation in total score (η^2 = 0.103) and 10% (η^2 = 0.099) in Arithmetic. From figure 3.2.5, it is known that the deviation in community schools is remarkable – ranging from 20% to 80%. Contrarily, most private schools in the sample show very high performance. This may be explained partly by much higher socio-economic input into students' life in institutional schools and strict selection of the students.

One may note the comparatively high value of Coefficient of Variance with the content area of Numeracy (77.5 and 46.3%). In both the community and institutional schools, this is due to the fact that the variance was relatively high in comparison to the mean scores in various content areas.

The dataset indicates that, on average, the students in institutional schools outperform the students in community schools. The difference is the widest in the area of Numeracy (22 percent). This deviance may be explained partly by much higher socio-economic input into students' life in institutional schools and strict selection of the students.

School location and student achievement

The variations in achievement has been studied in terms of school location by dividing them into two: rural and urban schools. This information was obtained from the head teacher though some of the head teachers did not inform about the school location. The students' achievement in rural and urban schools is depicted in table 3.2.16.

Table 3.2.16 Student achievement on the basis of location of schools

Location of school	Community				Institutional			
	N	Mean	SD	CV	N	Mean	SD	CV
Rural	8,873	48.6	22.4	46.1	1,067	61.7	18.4	29.9
Urban	907	54.9	23.1	42.1	1,731	70.0	18.0	25.7
Mean difference		6.3				8.3		
Cohen's d		-0.24				-0.55		
Total	**9,780**	**49.3**	**22.5**	**45.6**	**2,798**	**67.0**	**18.6**	**27.8**

The achievement level of the students in the urban community schools (55%) is 6 percent higher than that in the rural community schools (49%). The difference is statistically significant ($p < 0.001$) but the effect size is small ($d = 0.24$). While excluding the community schools in the Valley, the score of the urban community schools lowers to 46%; the difference (3 percent) is still statistically significant ($p < 0.000$), but the effect size is non-existent ($d = 0.07$). The main difference in the community schools is, hence, caused by the high level of the students in the Valley schools. The division into rural and urban schools explains only 0.8% of the student variation in the community schools ($\eta^2 = 0.008$) and, excluding the Kathmandu Valley schools, only 0.1% ($\eta^2 = 0.001$), the latter is a good sign from equality point of view; except for the Valley, there is no difference between the rural and urban community schools.

The achievement level of the students in the urban institutional schools (70%) is 8.3 percent higher than that in the rural institutional schools (62%). The difference is statistically significant ($p < 0.001$) and the effect size is medium ($d = 0.55$). Excluding the institutional schools in the Valley, the difference remains the same (7.5 percent) which still is statistically significant ($p < 0.001$) and the effect size is moderate ($d = 0.40$). In institutional schools, the effect of Kathmandu Valley is not that remarkable as in community schools. The division into rural and urban schools explains 5.5% of the student variation in the institutional schools (η^2 = 0.055), and excluding the Valley schools, it explains 4% (η^2 = 0.037).

The data suggests that the students in the urban community schools have gained 6 percent more than the students in the rural areas. Excluding the Valley schools, the difference is 3 percent; remarkable degree of the difference is caused by the Valley schools. In institutional schools, there is wider difference between the rural and urban areas even if the students from the Kathmandu Valley are excluded.

Language at home and student achievement

In the context of Nepal, the student achievement has been found depended on the language spoken in their homes i.e., the mother tongue of the students. The mother tongue reflects, in many cases, the ethnic background, and hence any difference may be taken as a possible source for inequality in society.

On the basis of the total data, 37% of the 5th graders speak a native (first) language other than Nepali. These “other” languages are quite fragmented; the largest groups in the Mathematics dataset are Tharu (5.3%), Newari (4.5%), and Urdu (4.2%). After dividing the languages into ten groups excluding Nepali, there were still 17.5% of the students classified into the group “others”. Because the languages are very fragmented and the Nepali speakers are the majority of the students, for the purpose of the statistical analysis, all the other languages are first grouped into “non-Nepali” and "Nepali". The results are presented in tables 3.2.17 and 3.2.18.

Table 3.2.17 Student achievement on the basis of home language

Content area	Community				Institutional			
	N	Mean	SD	CV	N	Mean	SD	CV
Nepali	4594	49.6	23.7	47.7	402	61.2	20.0	32.7
Non-Nepali	5649	48.9	21.5	44.0	2530	68.1	18.0	26.4
Mean difference		0.6				7.0		
Cohen's *d*		**0.028**				**-0.38**		

When combining all the minor language groups as “Non-Nepali”, there is not a notable difference between the language groups in the community schools (0.6 percent favoring the students having first language Nepali). The difference is not statistically significant and the effect size is very low (d = 0.028) which indicates no difference. The difference of 7 percent in institutional schools is moderately high ($p < 0.001$, $d = 0.38$).

On the basis of the original categorization of the minor languages, the issue looks quite much interesting: It is evident that in the community schools, the Magar and Tamang students are quite much at higher level in Mathematics than the Nepali speaking students (63% and 62% respectively) compared with Gurung students (36%). On the other hand, in the institutional schools, the students from Sherpa (41%) and Gurung (25%) language groups perform much lower than the average. Tamang students outperform the other language groups (average of 67%). The Nepali speaking students represent the major population, and hence it includes both the lowest and the highest extremes, and their result is on average.

Table 3.2.18 Achievement in the different language groups

Language/Ethnicity[1]	Community				Institutional			
	N	Mean[2]	SD	CV	N	Mean	SD	CV
Magar	57	62.9	22.3	35.4	104	66.7	18.7	28.0
Tamang	319	62.0	20.1	32.4	6	72.4	16.9	23.4
Rai	111	60.7	23.6	38.9	0	-	-	-
Urdu	522	53.3	21.7	40.7	27	65.6	21.8	33.2
Maithili/Awadhi	20	50.2	16.0	32.0	11	51.0	13.2	26.0
Nepali	5649	48.9	21.5	44.0	2,530	68.1	18.0	26.4
Tharu	666	48.1	25.0	52.0	33	54.2	21.7	40.0
Newari	554	46.7	20.6	44.1	34	57.8	13.7	23.8
Sherpa	30	46.3	16.4	35.5	2	52.5	11.6	22.1
Limbu	27	41.2	16.8	40.8	1	29.5	-	-
Gurung	36	36.4	22.8	62.8	1	88.5	-	-
Not Specified	121	28.8	21.4	74.4	3	60.1	21.6	36.0
Others	2131	48.6	23.9	49.1	180	59.6	20.7	34.7
Total	**10243**	**49.2**	**22.5**	**45.7**	**2,932**	**67.2**	**18.4**	**27.4**

1) *Those language groups in which number of the students was less than 10 are omitted.*
2) *The language groups are sorted on the basis of the mean in the community schools.*
3) *SD = Standard Deviation* 4) *CV = Coefficient of Variation*

The differences between the students in the highest and lowest performing language groups are statistically significant ($p < 0.001$) and notable; the effect sizes are moderately high ($f = 0.17$ in community schools and $f = 0.16$ in institutional schools). The division into smaller language groups explains about 3% of the variation in the data ($\eta^2 = 0.029$ in community schools and $\eta^2 = 0.027$ in institutional schools). Though the differences are wide between the extreme groups, the number of students is quite small in some of the language groups and, hence, the moderate effect size. When analyzing only the minority languages and hence, excluding the Nepali speakers and the group "Other", the effect size is moderate ($f = 0.29$) in the whole dataset – indicating really remarkable difference between the highest performing minority group (Tamang, 65%) and the lowest performing group (Gurung, 38%).

The dataset shows that there is an educational inequality within the language groups in possibilities of learning Mathematics. In the whole dataset, the students from Magar (65%) and Tamang (62%) language groups perform very high in mathematics while the students from Gurung (38%) and Limbu (41%) language groups perform very low. The differences between the language groups are remarkable.

Ethnicity/Caste and student achievement

While looking at the achievements in terms of the ethnicity/caste, overall, the Dalit students are found to be the lowest performers in the sample. The ethnicity/caste-wise result in the whole dataset is illustrated in figure 3.2.10.

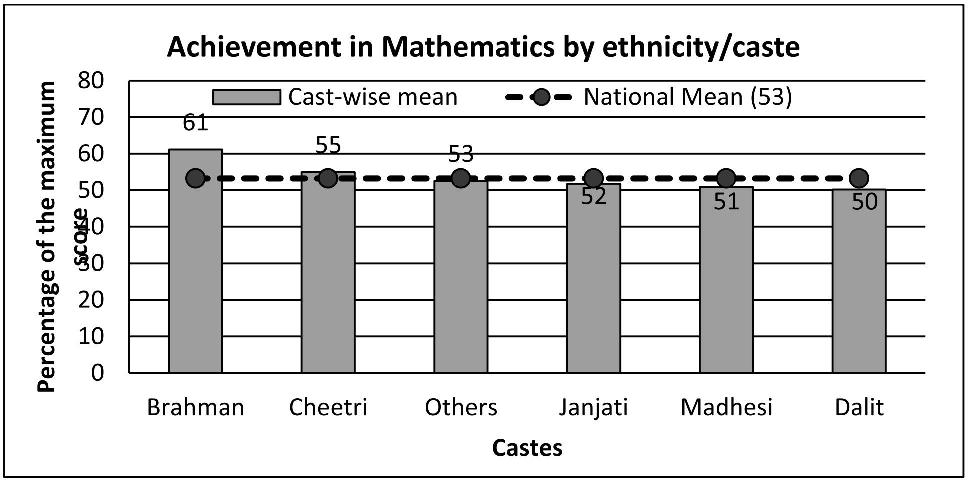

Figure 3.2.10 Achievement by ethnicity/caste

The students from the Brahman and Chhetri communities outperform the other groups and the Dalit students are the lowest performers in overall. However, the main result does not reflect variations among the strata. Gurung students, for example, who are from the Janjati background, are the highest achievers in some regions and lowest in others. Hence, variation is wide among the ethnicities/castes and across the strata.The results concerning the ethnicities/castes and achievements by the school type are depicted in table 3.2.19.

Table 3.2.19 Achievement of different ethnic/caste groups

Caste/ethnicity	Community				Institutional			
	N	Mean	SD	CV	N	Mean	SD	CV
Brahman	890	53.9	23.8	44.1	678	70.6	16.5	23.3
Chhetri	2,380	51.5	20.6	40.0	727	66.1	18.1	27.3
Madhesi	682	48.6	24.2	49.7	120	63.8	20.7	32.5
Janjati	3,033	48.6	21.6	44.5	634	66.7	18.2	27.2
Dalit	1,400	49.2	22.3	45.3	138	60.5	21.2	35.0
Others	1,580	46.7	24.2	51.9	608	67.6	18.9	27.9
Total	**10,243**	**49.2**	**22.5**	**45.7**	**2,932**	**67.2**	**18.4**	**27.4**

Based on the data, the students from "Other" castes are performing the lowest (47%) in mathematics, followed by Janjati (49%), Madhesi (49%), and Dalit students (49%) in community schools. Overall, achievement of Dalits is better than Janjati in community schools. Though, Dalit students perform the lowest in institutional schools. The overall difference between the groups is statistically significant ($p < 0.001$) though the effect size is small ($f = 0.13$) in the community schools; dividing the students according to their ethnic/caste background explains less than 2% of the student variation ($\eta^2 = 0.017$). In the institutional schools, the effect size is also small (f = 0.14); the division of students according to their ethnic/caste background explains 1.8% of the student variation (η^2 = 0.018). From equality point of view, this is a good sign though there is still a lot to do to reduce the gap between the castes.

Though. Dalits have historically been deprieved from education, their participation in education has been improved. Table 3.2.20 shows the detail figure of Dalit students.

Table 3.2.20 Dalit students' achievement in different Ecological/Development regions

School Type	Ecological zone	Development region					
		Eastern	Central	Western	Mid-Western	Far-Western	Total
Community	Mountain	53.2	50.8	39.3	62.1	46.0	52.5
	Hill	39.3	45.0	51.8	38.4	52.4	47.5
	Tarai	52.7	48.9	57.9	36.8	46.5	49.5
	Total	52.7	48.9	57.9	36.8	46.5	49.5
Instutitional	**Ecological zone**	**Eastern**	**Central**	**Western**	**Mid-Western**	**Far-Western**	**Total**
	Mountain					60.7	60.7
	Hill	48.4	67.5	67.1	52.0	52.5	64.8
	Tarai	72.1	50.8	89.3	46.4	44.9	49.0
	Total	**56.3**	**60.1**	**68.0**	**49.6**	**46.2**	**60.5**

A positive sign from the equality viewpoint is that the Dalit students perform higher than the national mean of community school students (49.2%) in the Eastern Mountain (53%), Tarai (53%), Central Mountain (51%), Western Tarai (58%), Mid-Western Mountain (62%), Far-Western Hill (52%), and the Kathmandu Valley (62%). However, the results are found much lower than average in the Eastern Hill (39%), Western Mountain (39%), and Mid-Western Hill (38%) and Tarai (37%). Generally speaking, few Dalit students in institutional schools (n = 138) perform always lower than the average (67%). Especially, the results of the Dalit students are much lower than the average in institutional schools in the Eastern Hill (48%), Central Tarai (51%), Mid-Western Tarai (46%), Hill (52), and Far-Western Tarai (45%).

The dataset informs that there are statistically significant, though, not necessarily remarkable differences between the ethnicities/castes in Mathematics. Dalit (50%) and Madhesi students (51%) as well as Janjati (52%) are performing somehow lower than Brahmin and Chhetri students. Dalit students perform lower especially in the Mid-Western's Tarai, Hill and Far-Western's Tarai area.

Gender and student achievement

Efforts have been put globally into reducing the difference between boys' and girls' learning achievement. This matter is handled somehow more extensively, with importance, than the previous sections of equality. Basic results are depicted in figure 3.2.11.

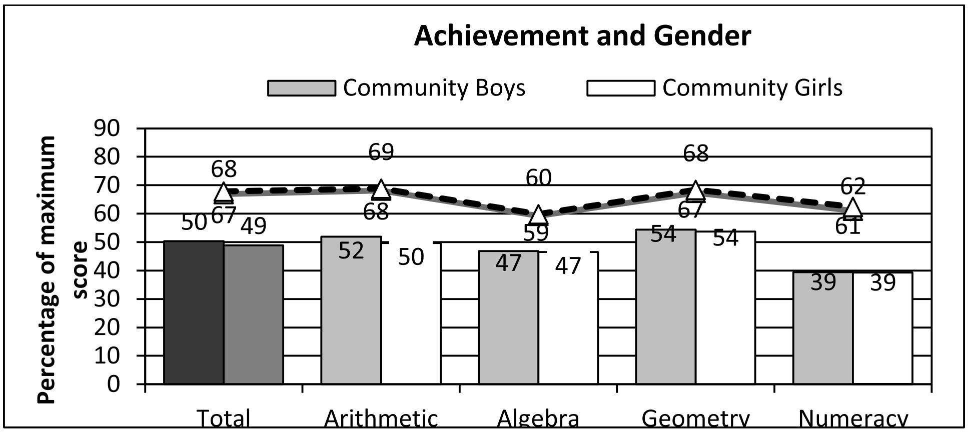

Figure 3.2.11 Comparison of achievement between boys and girls in various content areas

In community schools, there are no statistically significant differences between boys and girls in any of the content areas of Mathematics (p = n.s.) except for Arithmetic ($p < 0.001$) and total score ($p = 0.001$). The differences in Arithmetic and total score are very small, though the boys are found to be performing slightly better than girls. In institutional schools, there are no statistically significant differences between them. The effect sizes are very small in both the community ($d < 0.10$) and institutional schools ($d < 0.06$), which is a good sign from equality point of view.

Gender and Ecological zone

In community schools, boys are found slightly outperforming girls, except in the Valley where girls are slightly better (figure 3.2.11). At highest, the differences are less than 3 percent. In institutional schools, the girls perform better in Mountain zone (6 percent), whereas best performers are the boys in Tarai zone (7 percent). When it comes to the Ecological zones, the differences between boys and girls are small. However, they are noticeable in both types of schools.

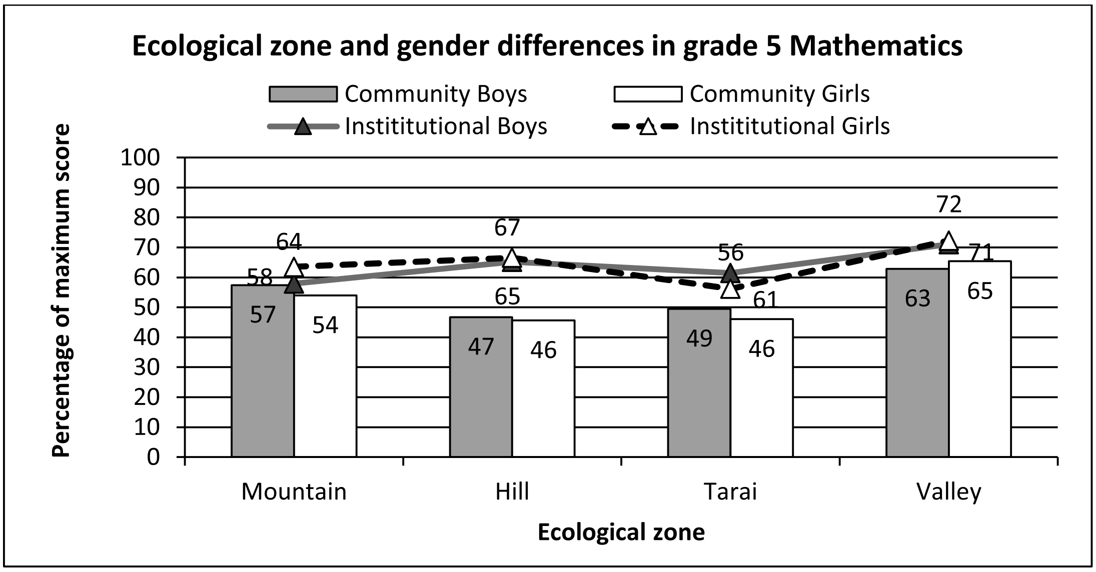

Figure 3.2.12 Ecological zone and gender differences in achivement by school type

Gender and Development region

There are some notable differences between the Development regions when it comes to boys' and girls' equal opportunities to reach the same educational goals (figure 3.2.12).In the community schools, differences are not wide – usually less than 3 percent favoring boys except in Valley where the girls slightly outperform the boys. However, in the institutional schools of Eastern region, the boys show as high as 10 percent higher performance than the girls.

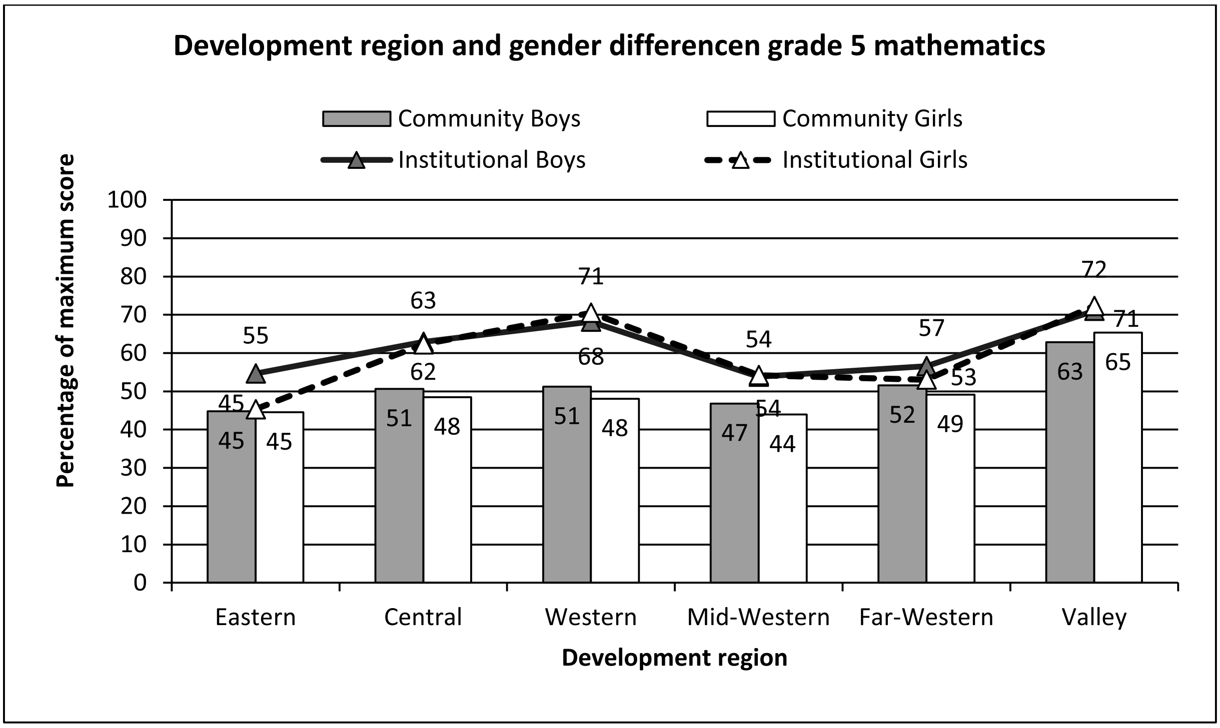

Figure 3.2.13 Development region and gender differences in achievement

The dataset shows that the differences between boys and girls in Mathematics proficiency are very small and in many cases non-existent.Though the differences in the institutional schools are statistically significant in total score and in Arithmetic, the effect sizes are very small – indicating that the differences are not at all remarkable. From the equality point of view, this is a positive indication. However, there are wide differences between the boys and girls in the Eastern region and among the Madhesi students.

3.2.3 Selected Explanatory Factors and Achievement

The simplistic model in chapter 2, section 2.4 represents several possible factors, which explain the differences in student achievement. Many of the factors have already been handled above: geographical factors, such as districts, Ecological zone, and Development region, as well as school-related factors, such as school type and school location. Also some individual related factors were handled, such as home language, ethnicity/caste and sex/gender. In this section, some other factors are also taken into consideration; they are socio-economic status (SES) of the students' families, work for paid job before and after school, students' attitude towards Mathematics as a school subject, age of the student, and support provided to studies are mainly the family and individual related factors. As a sample of deepening school and teacher-related factors, also the availability of textbooks, homework given by the teacher, students' future plans, and the selected activities in the school are handled. Many other factors could also be selected since the background questionnaires are rich.

Two things are worth mentioning. First, some of the questions were too demanding for the grade 5 students to be answered meaningfully by themselves. These kinds of questions included, for example, information of the parents' education. For example, the number of parents of being "just literate" (42%) looked very high. Second, maybe because of the first point, there were quite many missing values in the background questionnaires. For example, in the question of mothers' education, 7% of the students did not answer. These facts evidently have an effect to the analysis and therefore the readers need to be critically aware of the crudeness of the information when it comes to the grade 5 students. In most cases, however, the results looked credible and comparable with grade 8 datasets.

Parents' education, occupation, home possession and student achievement

There are many variables indicating the socio-economic status. In this analysis, these were categorized into parents' education, parents' occupation, home possessions, home accessories, and whether the student attends a private school or not. Finally, the SES is estimated on the basis of seven indicators related to the economic, educational, and occupational background of the family. In this section, parents' education is further elaborated, so that the illiteracy of the parents is analyzed in relation to achievement in mathematics.

Several SES related variables were analyzed by using a data mining tool of SPSS, DTA. The method is very effective in finding the cut-offs of the predicting variable, such as mothers'education, and classifying the factor into several groups, which differ statistically in the most significant way from each other in relation to student achievement. Some examples of this are handled with parents' educational background and its relation with students' achievement in mathematics.

Parents' education

In the background questionnaire, parents' education is divided into eight categories: 1) illiterate, 2) literate, 3) grade10, 4) SLC, 5) IA, 6) BA, 7) MA, and 8) above MA. The question was asked to the students and hence there may be some impurities embedded in the data; the number of (just) literate mothers in the dataset is found too high (see figure 3.2.13). However, with the huge dataset the results look credible.

DTA classifies mothers' education into four groups with statistically significant differences in students' achievement levels (figure 3.2.14): illiterate (students' average is 52%), just literate (53%), grade 10 passed mothers (59%), and SLC passed or higher (68%). The difference between each group is statistically significant ($p < 0.001$). In practical words, the results show that when the mother has passed at least SLC, she can give, on average, + 16 percent advance for her child in the national test compared with illiterate mother. Figure 3.2.14 shows that if the mother was BA passed or higher, the advance was + 18 percent over the illiterate mother. Mothers' education explains 10% of the student variation ($\eta^2 = 0.035$) which indicates a moderate effect size ($f = 0.19$). Obviously, the result suggests that the children from the highly educated mothers are mainly found in the private schools.

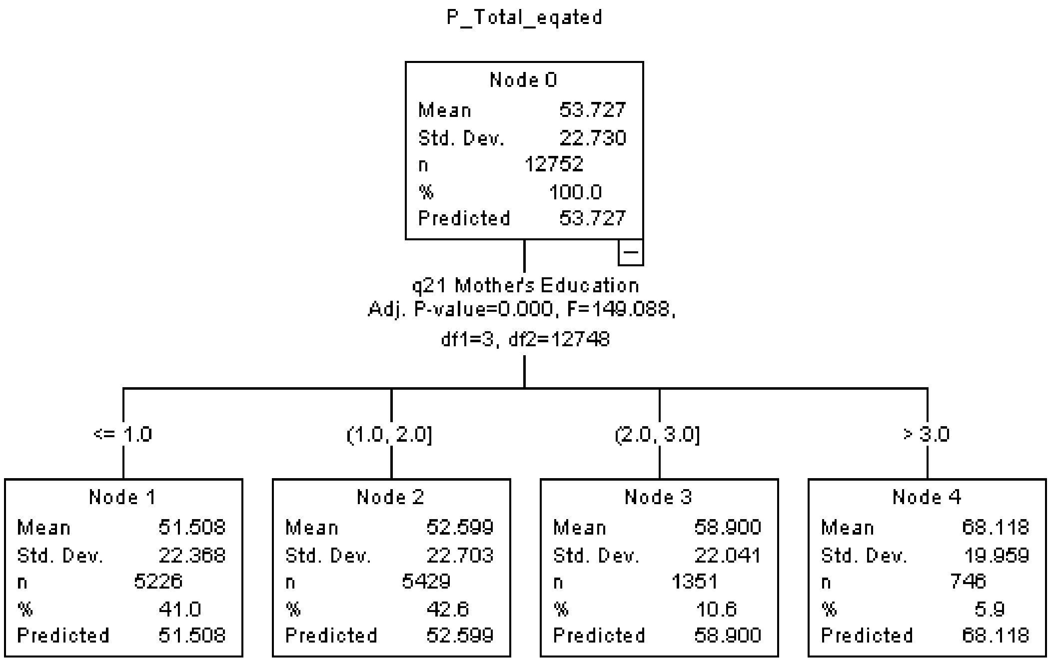

Figure 3.2.14 DTA of mothers' education and students' achievement

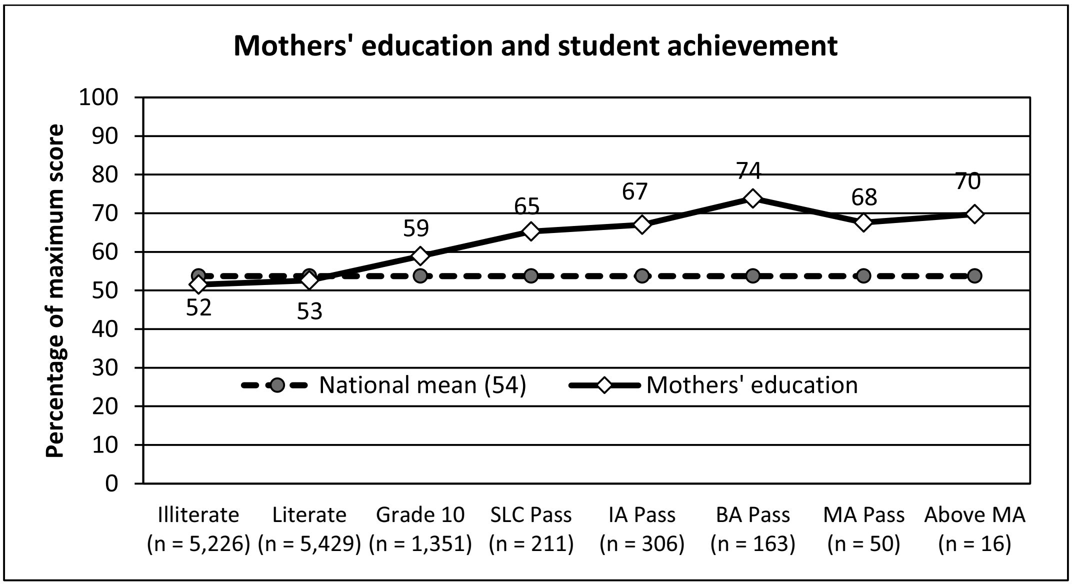

Figure 3.2.15 Mothers' education and students' achievement

In parallel, DTA groups fathers' education into four categories: illiterate (49%), just literate (51%), grade 10 (58%), and SLC or higher (65%) (Figure 3.2.16). The difference between each group is statistically significant ($p < 0.001$). Practically, the results show that when the father has passed SLC, he can contribute on average, + 16 percent advance for his child in the national test compared with illiterate father. Figure 3.2.16

shows that if the father was MA passed, the advance was + 22 percent points over the illiterate father. Obviously, the high average means that the children from the highly educated fathers (as well as of mothers) are mainly found in private schools. Father's education explains 11% of the student variation ($\eta^2 = 0.110$) which indicates a high effect size ($f = 0.35$).

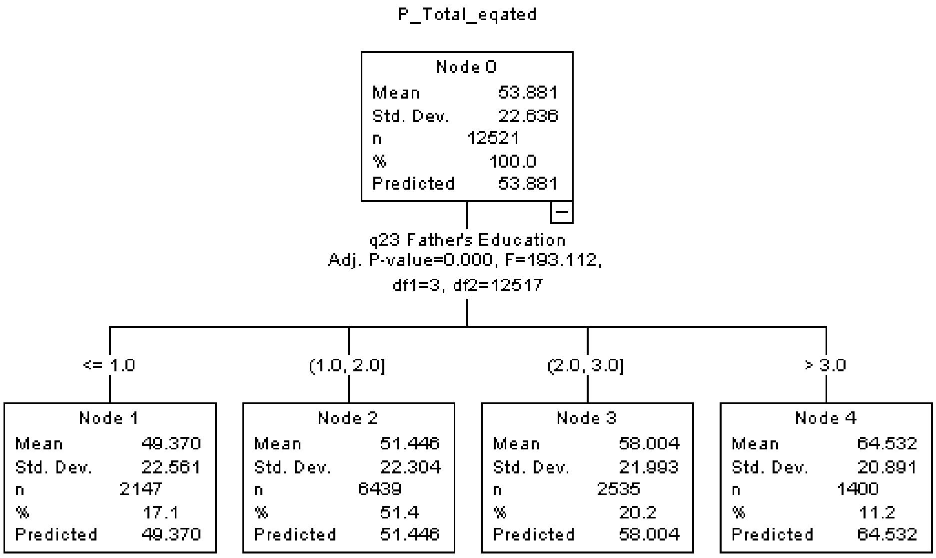

Figure 3.2.16 DTA of fathers' education and students' achievement

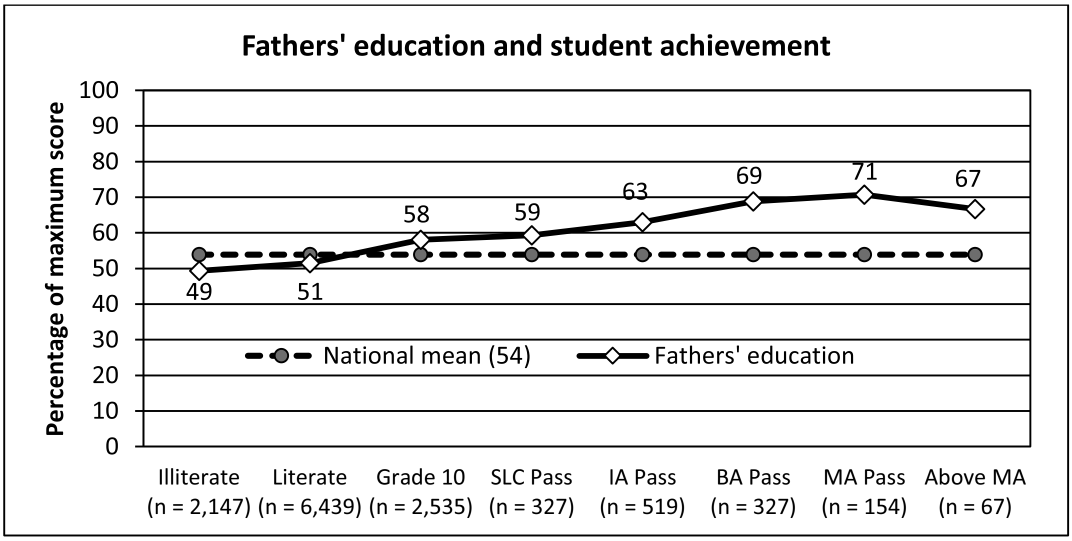

Figure 3.2.17 Fathers' education and students' achievement

It is noteworthy that mother's education provides slightly higher leverage to children's achievement in Mathematics than that of fathers'. While fathers' literacy (up to grade 10) raises the child's achievement by 10 percent, mother's literacy (up to grade 10) is found to have raised it by 14. In parallel, while fathers' high education (BA or higher) contributes the child's achievement by 20 percent, mothers' high education raises it by 21. In both cases, the effect size is moderate ($f = 0.19$ to 0.22) – showing that the difference between the highest and lowest group is remarkable.

After combining mothers'and fathers'education, the poorest prediction in DTA for the children's future achievement in Mathematics comes when the father is (just) literate but the literacy of the mother is not known (42%). Results are quite low also for the groups where the father is illiterate regardless the mothers' education (49%) or when the father is literate regardless mother's education (51%) and when the father is (just) literate but the mother is illiterate (52%). The highest results are found for those groups whose father and mother both have passed (at least) the grade SLC (70%) or when the father has passed SLC and mother is (at least) literate or passed grade 10 (62%). It is evident that the educational capacity provided by the parents can be utilized by the students which implies: higher the parents' education, the better results the children achieve.

In what follows with the final SES variable, the cut-off for parental education was set to "grade 10", that is, when being passed the grade 10 (or higher), the indicator for mother's (and father's) education for SES was set to 1, and the lower education than grade 10 pass was given the value 0.

Parents' occupation

The occupation of parents was categorized into eight groups: 1) working abroad 2) farming and working at home 3) only working at home 4) teaching 5) services 6) business 7) working for daily wages, and 8) working at others' home. Those occupations are recoded as given in the figure 3.2.17. The result related to mothers' occupation is seen in figures 3.2.18 and 3.2.19 and to fathers' occupation in figures 3.2.20 and 3.2.21.

While comparing the students' means by DTA, the achievement is the lowest when the mothers' occupational background is agriculture (51%). It is statistically significantly lower than the mother working only at home and working for other's home (55%), in comparison to the mothers involved in business or daily wages (62%), nothing to say when mother are teachers or service holders (65%). Mothers' occupation explains 8.5 percent of the student variation (η^2 = 0.028) – which indicates a moderate effect size (f = 0.17).

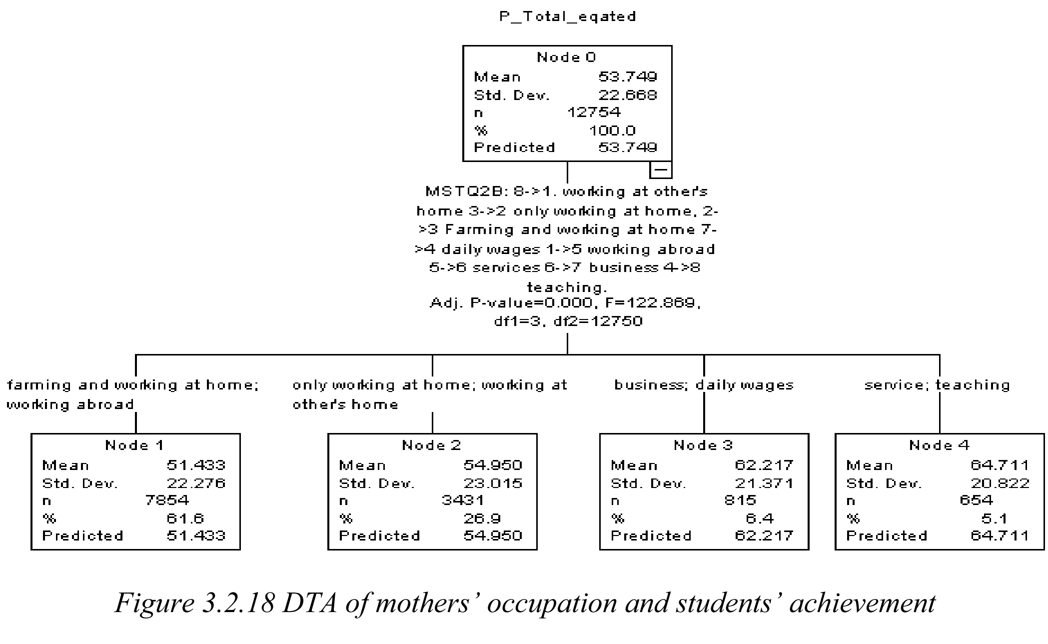

Figure 3.2.18 DTA of mothers' occupation and students' achievement

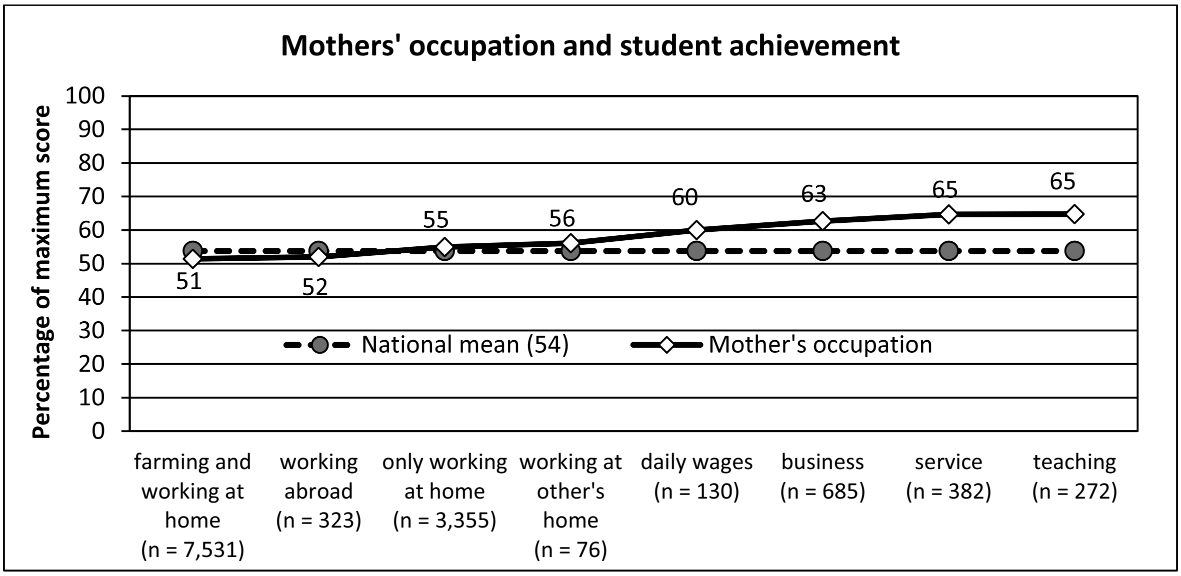

Figure 3.2.19 Mothers' occupation and students' achievement

Students whose mothers' work for agriculture and at home or working abroad have the poorest achievement (with 51% and 52% achievement respectively), whereas the achievement is highest when the parents having either service or teaching profession (65%). It is also found that students' achievement is also equal or more than the national average if mothers are involved in earning jobs like daily wages and business. Students are seen to be at risk of achieving poor when their mother is involved in household work or work in others' home.

With regard to the fathers' occupation, on the basis of DTA (figure 3.2.19), those students' achievement in Mathematics is the lowest whose father is only working at home (47%), followed by farming and working at home and working at others' home (50%), and working abroad (53%). The highest achieving students are from those groups whose parents have a regular income source: such as daily wages (56%), business (60%), service (63%), and teaching (63%). The difference between the lowest and highest group is 16 percent which is a wide gap. Fathers' occupation explains 6 percent of the student variation ($\eta^2 = 0.058$) which indicates a medium effect size ($f = 0.25$). Relationship between fathers' occupation and student achievement is shown in figure 3.2.20.

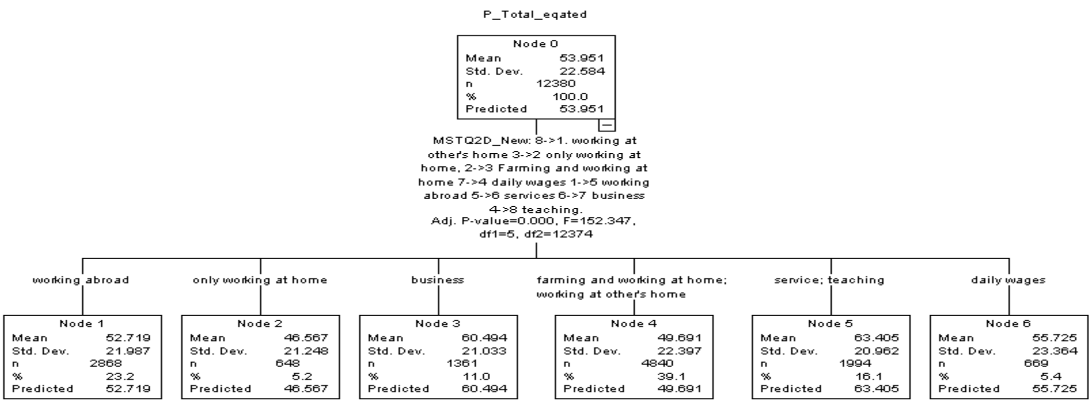

Figure 3.2.20 DTA of fathers' occupation and students' achievement

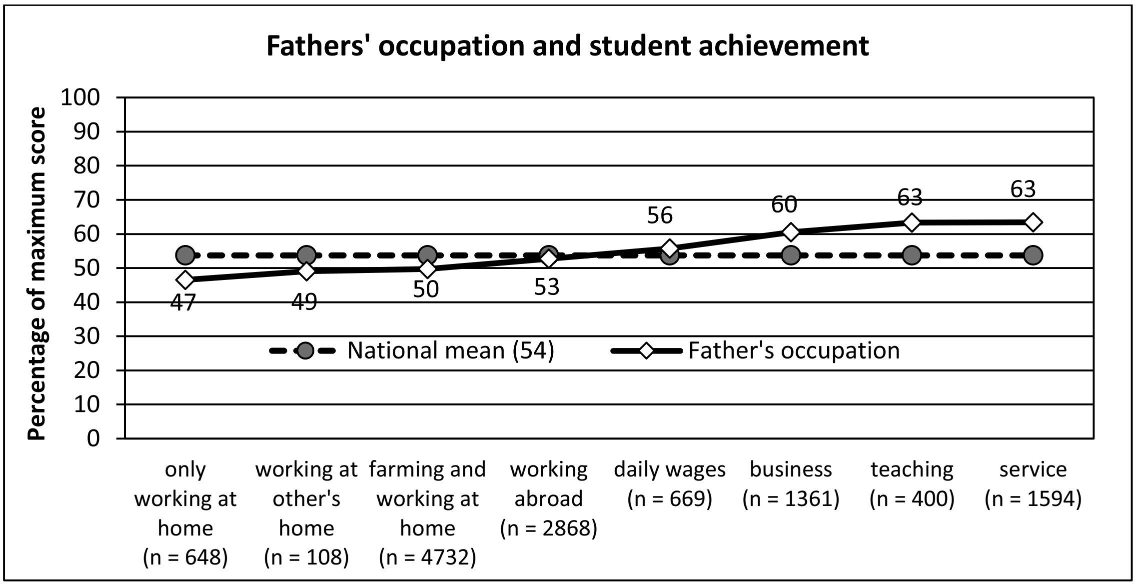

Figure 3.2.21 Fathers' occupation and students' achievement

Students achieve below average when their father is involved in household chores and farming. Combining the mothers' and fathers' occupation, DTA shows that the lowest achievement is found among the students from the families where both parents work only for home (49, n = 437), which means that the father works in agriculture regardless the mother's occupation (50, n = 4,732) or where father works abroad regardless the mother's occupation (53, n = 2,868). The highest achieving students come from the families where the father comes from the services regardless mothers education (63, n = 1,594), both the father and mother are teachers, or the father is a teacher and the mother is in business or services (63, n = 176). It is worth noting that teaching, service, and business occupations indicate permanent salary in the family. For the later use as a SES-indicator, the cut-off for parents' occupation was made so that being in agriculture gives 0 and all others 1.

Home possessions and accessories

Facilities and resources available in home may have some effects on the achievement. There were two kinds of home possessions defined in the background information questionnaire for the students. There were 11 questions in the student background questionnaire related to home possessions. Each was scored 1 if the student had an access to possession. Adding these items up, the maximum score was 11 when the student responded they have access to all the possessions and the lower the score, the fewer possessions they have at home. Figure 3.2.21 shows the connection of home possessions and achievement level: Except for the highest category, the achievement level of the students' raises logically the more there is access to home possessions. Pearson correlation between the achievement level and the factor (r = 0.12) is statistically significant (p < 0.001) but indicates small effect size (d = 0.25).

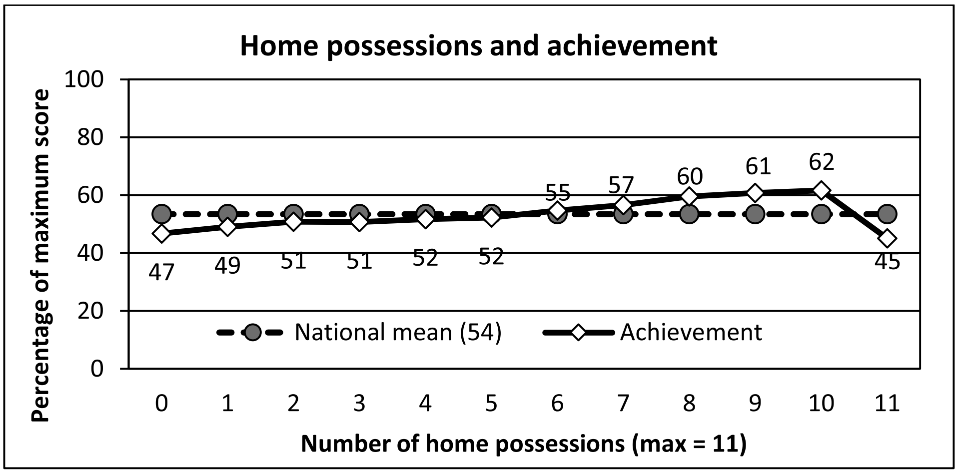

Figure 3.2.22 Relation between home possessions and achievement

For the later use in SES, the cut-off for the factors was set on 6 possessions: if 6–10 items as mentioned in the background questionnaire were met, the student was given 1 otherwise 0.

The same pattern – the more possession, the better results – can be seen also with home accessories, as seen in figure 3.2.23. The question in the background questionnaire was set differently compared with home possessions; with the accessories it was asked "*How many of the following accessories do you have in your family?*" with the options 0 – 3 (or more). For the indictor, the availability of the home accessories is dichotomized in the same way as the home possessions.

Table 3.2.21 Dichotomizing the indicators for home accessories

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

After dichotomizing the items individually by using meaningful cut-offs was found with ANOVA and DTA (see, table 3.2.21), all three indicators were summed up.[26] The maximum score was 3 – indicating that the students' possessed at home a set number of all of the accessories.[27]

[26] There was also fourth item in the questionnaire – the number of radios in home. However, this item behaved pathologically in the analysis: the more there were radios in home the less achievement. Hence, it was not taken as an indicator for SES.

[27] The analysis is bound to the fact that values were given by the students – they are, in many cases, credible.

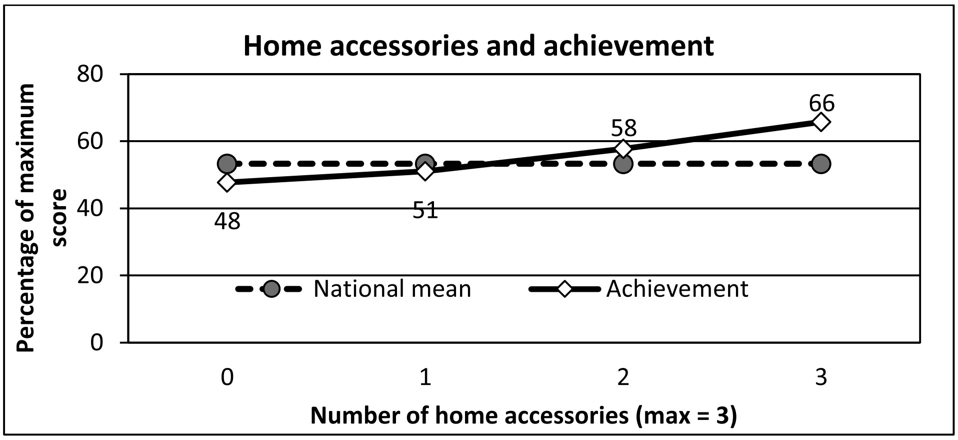

Figure 3.2.23 Relation between home accessories and achievement

Figure 3.2.23 clarifies that how increase in the number of home possessions or accessories increases the students' achievement. Accordingly, achievement has increased from 48% (if none of them are available) to 66% (if all three of them are available). Availability of all the stated facilities indicates the higher SES of the family. Correlation between the number of home accessories and achievement is r = 0.26 (p < 0.001) which is certainly positive – indicating a moderate or high effect size (d = 0.62).

The dataset suggests that parents' educational level predicts the children's future achievement level in mathematics. Especially negative impact for the achievement level is the situation where either the father or mother or both are illiterate or just literate. Among the samples, 41.0% students had an illiterate mother and 17.1% had an illiterate father.

The dataset is also evident that either economic or intellectual capacity or the both at home helps children to increase their Mathematics achievement. If the father or mother or both are associated with an agricultural or related occupation and working abroad, the students' achievement in Mathematics is significantly lower than with other occupational groups. Among the samples, 59.0% mothers and 38.2% fathers worked in agriculture or only at home.

The dataset shows that when the children have very few home possessions – zero of the 11 – the achievement level is remarkably lower than the national average (47%). With nine to ten possessions, the average score is very high (61–62%) in comparison with the national average. The same is true of home accessories: When none or only one accessory indicator out of three is met, the results are lower than average (48–51%) and when there are two or more are met, the results are remarkably higher (58 to 66%). Of the total 3.7% students did not have any of the home possessions and 36.4% had no accessories.

SES and achievement

The socio-economic status of the family was formed on the basis of seven indicators which were all first dichotomized. The variables (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 up (as SES) and changed into the percentage (P-SES). The P-SES represents the percentage of SES of the student's family;

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 are positive) and 0 refers to the lowest possible SES (that is, all the seven indicators of SES are negative). The analysis of the P-SES by using Univariate GLM (that is, the Regression modeling) shows the strong relation between SES and achievement. Figure 3.2.24 presents the relationship between SES of the students and the achievement.

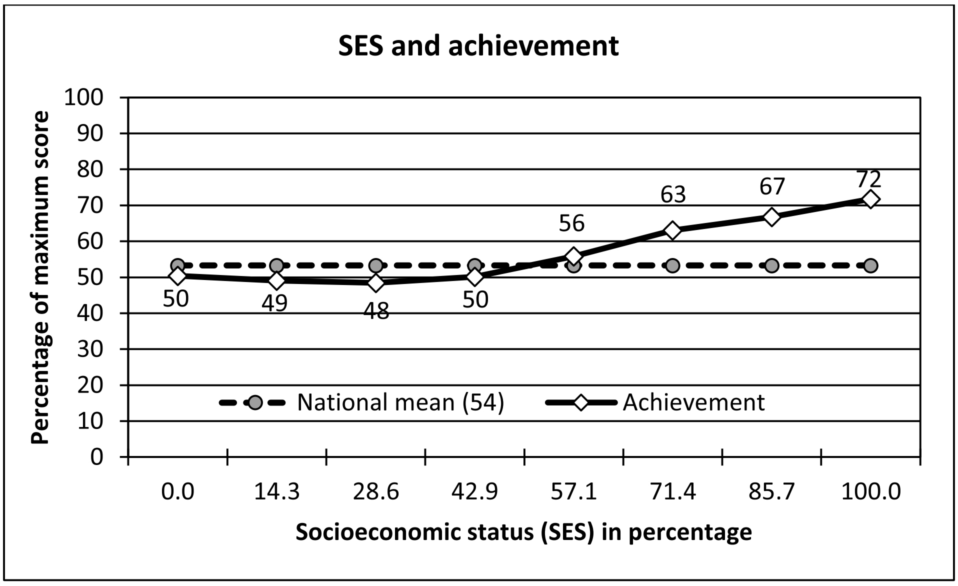

Figure 3.2.24 Relation between SES and achievement

Figure 3.2.24 shows a strict positive relationship between SES and the Mathematics achievement; the higher the SES, higher is the achievement. Pearson correlation between the variables is $r = 0.27$ which is a high value ($p < 0.001$) and indicates very high effect size ($d = 0.62$). The difference in achievement between the lowest SES groups (50%) and the highest one (72%) is remarkable. SES explains about 32% of the student variation ($\eta^2 = 0.32$) which is a very high percent, but it is seen to be at the same level as was found in English dataset ($\eta^2 = 0.311$, see chapter 5) and in Nepali language in the grade 8 dataset (0.28, see ERO, 2013, 152).

It is worth noting that SES as a variable is more school related than student related factor. The correlation of SES and achievement is $r = 0.27$ in the student dataset but $r = 0.36$ in the school-wise dataset. It is also worth noting that even though the SES is controlled in the student-wise dataset[28], there is still statistically significant difference between community and institutional schools ($p = 0.001$). However, the effect size is reduced from $f = 0.34$ to $f = 25$, that is, from high to moderate.

From the sociological point of view, it is interesting to know which of the individual indicators of SES are not met in those families where the children perform the lowest. Figure 3.2.25 illustrates the fact that in the families with meeting less than four SES

[28] Because attending private school is imbedded in the SES, the school type does not explain the achievement in ANCOVA when controlling the SES. For the ANCOVA, another SES – without the school type – was created.

indicators, the challenge lies mainly in three factors marked in figure 3.2.25 with red/dark circle, triangle, and square: both mother's and father's education is low and the child does not attend the private school. The last is difficult to change in community schools but the low adult education (and especially mother's low educational level) would be possible to tackle with an educational policy.

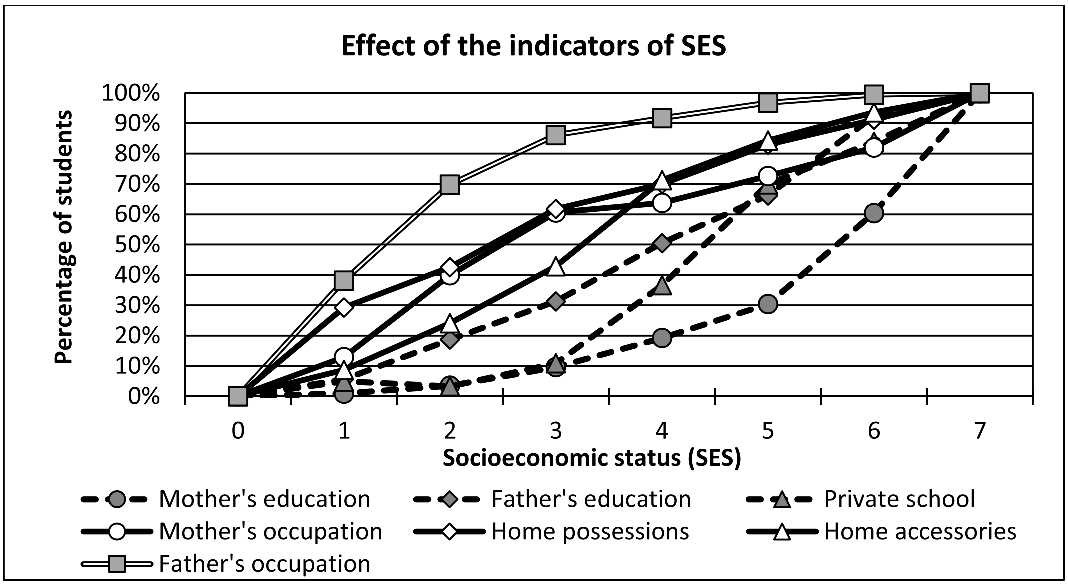

Figure 3.2.25 Effect of individual SES indicators in achievement

The dataset strongly suggests that the socio-economic status plays an important role in the Mathematics achievement in Nepal. The difference between the lowest and highest SES groups is remarkable (23 percent). This means that if the problems of parents' low educational level is solved or the SES of the lowest performing students is raised into a decent level, the results in these groups will raise remarkably. Especially challenging is the situation in the families where the father or both parents are illiterate or they both work for agriculture or at home. Out of the total sample, 16.0% of the students are at the lowest level of SES.

Working beyond school hours and achievement

Several questions were set in the student background questionnaire on the students' activities outside the school. Two of them are briefly handled here: Working before and after the school for a paid job and participating in household work/chores. The values of the variables are divided into five categories: 0 (no time at all), 1 (less than 1 hour per day), 2 (1–2 hours per day), 3 (2–4 hours per day), and 4 (more than 4 hours per day).

The DTA indicates that, in case the working after and before school, the cut-off is on whether the students work for a paid job or not. The DTA shows that when the children have no paid work at all, the results are above the national average in both community and institutional schools. If the students work for a paid job – even less than one hour, the results are statistically significantly lower than the average. The ANOVA shows that the relationship is strict ($p < 0.001$) though mildly ($f = 0.18$) negative when the students need to be engaged in paid work before and after school. It is notable, though, that most of the grade 5 children do not need to be engaged in paid work. Working outside school indicates that the family is poor and the extra income is needed for survival. It is obvious that when the student needs to work for more than 4 hours per day there is no time or energy to handle school homework. In the institutional schools, the difference between the children working over 4 hours per day (60.5, n = 40) is notably lower than those with no need to work at all for paid job (69, n = 2463) (see figure 3.2.26 institutional schools).

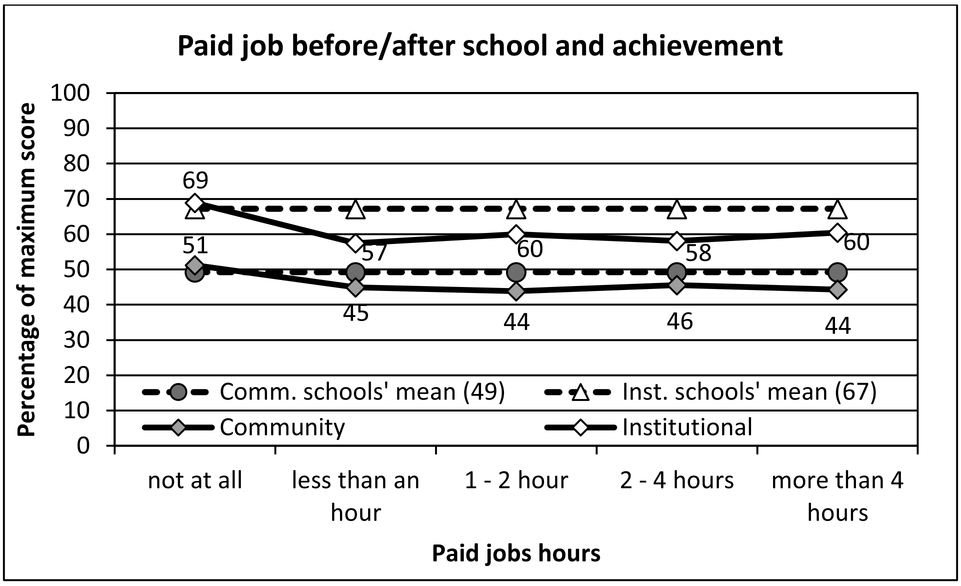

Figure 3.2.26 Relationship between achievements and paid job beyond school time

In the case of unpaid participation in the household work, it is usual – and a supported – practice in families to involve children in household chores, which is a part of the socializing process of children. The DTA shows that when the child spends some time (less than two hour) for the household chores, the results are statistically higher (55–58%) than the children not spending at all (47%) or more than 4 hours per day (48.5%). The effect of not participating in the household chores looks to be more drastic in the community schools than in the institutional schools. Actually, it is found that in institutional schools, it does not make any difference whether the children work for two hours or less or not at all; but the effect is seen as in spending four or more hours in chores. In community schools, the results are significantly lower for those not participating in the chores. Differences are significant ($p < 0.001$) though the effect size is small or moderate ($f = 0.19$ in community schools and $f = 0.12$ in institutional schools). It is somehow interesting that more than 10% of the students (n = 921) inform us that they spend more than 4 hours per day doing household work. In the rural area, this may be obligatory where farmers are involved in cattle raising and when the cattle are far from the home. It would be understandable that, in these cases, there is not much energy to decently concentrate on their school work.

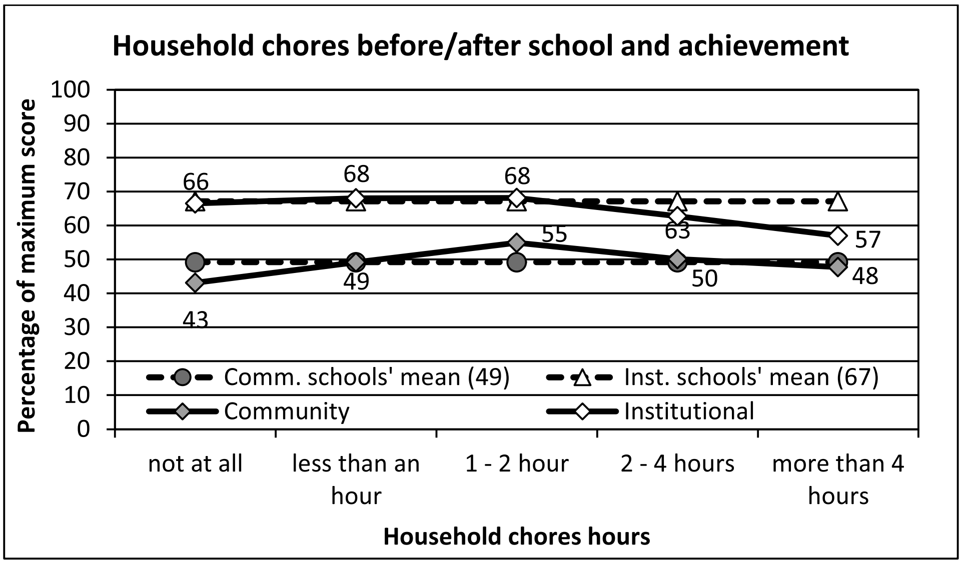

Figure 3.2.27 Relation between household work and achievement

The dataset reveals that either working for a paid job or for four hours per day in an unpaid household work outside school effectively reduces the school achievement of the student. However, a decent amount of household work up to two hours per day supports the students' learning in mathematics. Of the total sample population, 27.1% students worked for the paid job and 18.1% spent more than 2 hours for the household chores.

Attitude toward the subject sand achievement

In the assessment of the Mathematics achievement, attitude tells us what the students think about Mathematics and its usefulness in their daily life and future. There is a more or less firm relationship between the attitude of the students and achievement. Though the relation is not always clear, the correlation between achievement and attitude towards the subject as well as self-efficacy in the subject is widely studied (see in Mathematics, for example, Metsämuuronen 2012a; 2012b; House & Telese, 2008; Shen & Tam, 2008; Kadijevich, 2006; 2008). Some researchers have noticed remarkable differences in correlation between countries (e.g., House & Telese, 2008; Kadijevich, 2006; 2008; Wilkins, 2004; Shen, 2002; Papanastasious, 2000; 2002; Stevenson, 1998).In some countries, the correlation between attitude and achievement is found near zero, like in Macedonia (Kadijevich, 2008),in the Philippines (Wilkins, 2004), in Indonesia (Shen, 2002) or in Moldova (Shen, 2002), whereas, in some other countries, the correlation can be as high as 0.60 (e.g., in Korea, Shen, 2002). In NASA 2011, it was noticed that grade 8 students were not consistent in the attitude test and the reliability of the international test stayed low (see ERO, 2013 and table 2.11).

In NASA 2012, the same shortened version of Fennema–Sherman Attitude Scales (Fennema & Sherman, 1976) as used in several international comparisons like in TIMSS and PISA studies was used. The original scales included nine dimensions but in these international comparisons only three are used with four items on each dimension and two negative items on each of the first two dimensions. The names of the original factors were "Liking Mathematics", "Self-Efficacy in Mathematics", and "Experiencing utility in

Mathematics" (compare naming in, e.g., Kadijevich, 2006; 2008). Because of students' inconsistent manner in responding to the attitude scale in NASA 2011, only the dimension of "Experiencing utility in Mathematics" was taken into the measurement instrument of grade 5 students. Reliability of the score of five items is sufficient ($\alpha = 0.65$). The relation between the attitude (divided into six groups with about an equal number of students, that is, sextiles[29]) and achievement score is shown in figure 3.2.28.

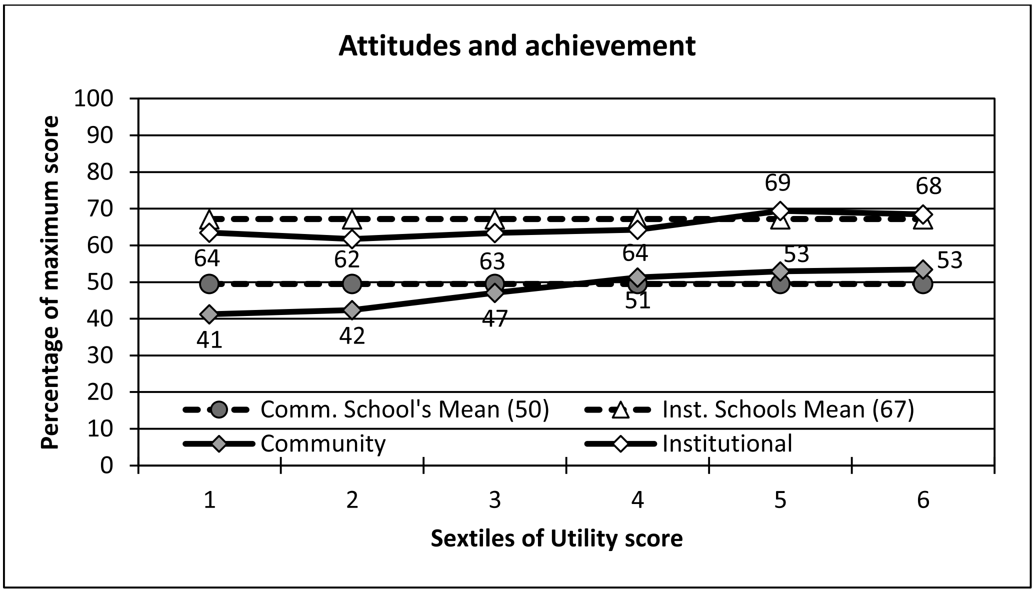

Figure 3.2.28 Relation between attitude and achievement by school type

There is a clear positive correlation between attitude and Mathematics achievement in the whole dataset ($r = 0.23$). The connection is moderately high ($d = 0.53$) indicating that the difference between the means of the lowest attitude group (43) and highest one (58) is remarkable. Connection is higher in the community schools ($r = 0.19$) than in the institutional schools ($r = 0.10$). The connection of the attitudes and achievement is not found to be fully logical; it is likely that within the highest attitude group, there are many students who may have either did not give the answers or did not understand the questions (see the same kind on phenomenon in the SES analysis above). The difference between the lowest and highest attitude group is 12 percent in the community schools ($f = 0.22$) and 8 percent in the institutional schools ($f = 0.14$).

The dataset indicates that positive attitude towards the subject correlates with positive achievement in Mathematics. The better achievement is more probable a consequence of more positive attitude.

Age and student achievement

In the Nepalese context, the age of the students attending grade 5 studies varies widely. Some students have mentioned their age below nine years and some above 16. All the

[29] The original score is short (maximum was 15 points) and quite many students (36%) gave the maximum score. Hence it was not possible to form more precise classification such as deciles. Six classes (sextiles) was the most precise alternative with the given dataset.

ages of the students below 10 were encoded as 'up to 9 years', and all students above 14 were encoded as '15 years or above". The descriptive statistics of the mean for each age group are given in tables 3.2.22 (a) and 3.2.22 (b) and depicted in figure 3.2.29.

Table 3.2.22 (a) Students' achievement by different age groups

Age	N	Mean	SD	CV
Up to 9 years	385	48.8	22.0	45.0
10 years	1,885	53.1	23.5	44.2
11 years	3,573	55.8	23.1	41.4
12 years	4,140	53.4	22.8	42.7
13 years	1,972	52.0	22.2	42.6
14 years	738	50.4	22.7	45.0
15 years or above	479	46.0	22.2	48.4
Total	**13,172**	**53.2**	**22.9**	**43.1**

Table 3.2.22 (b) Student achievement in different age groups by the type of school

Age	Community				Institutional			
	N	Mean	SD	CV	N	Mean	SD	CV
Up to 9 years	361	47.9	21.7	45.3	24	62.8	21.7	34.6
10 years	1571	50.3	23.4	46.6	314	66.9	18.1	27.0
11 years	2458	49.7	22.7	45.6	1115	69.0	17.8	25.8
12 years	3029	48.6	22.3	45.8	1111	66.5	18.7	28.2
13 years	1677	49.8	22.0	44.1	295	65.0	18.5	28.5
14 years	686	49.3	22.5	45.7	52	63.9	20.4	32.0
15 years or above	458	45.5	22.1	48.6	21	56.8	23.5	41.4
Total	**10240**	**49.2**	**22.5**	**45.7**	**3,608**	**78.9**	**13.5**	**17.1**

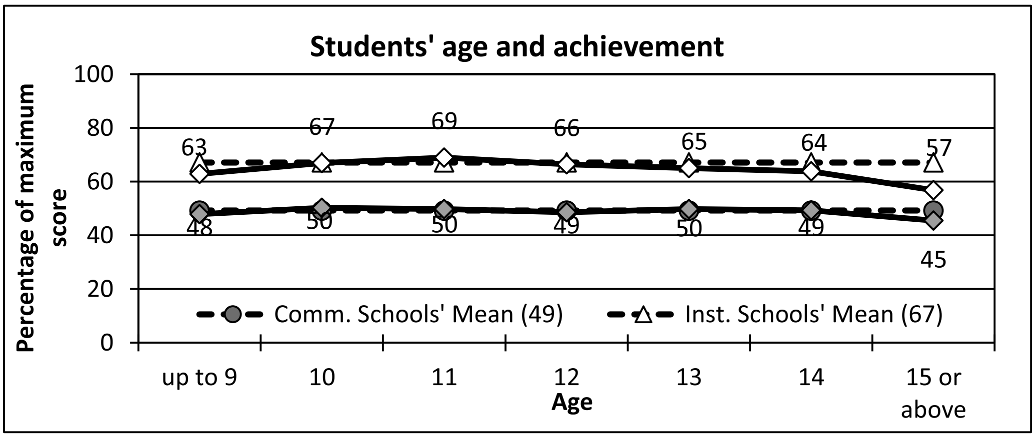

Figure 3.2.29 Relation between age and achievement

It is evident that the best achievers are those students who are at the proper age year for grade 5 studies (10 to 12 years old), scoring 49–50% in community schools and 66–68% in institutional schools. The higher the age is – meaning that the students have either started much later than they should have, or they have repeated the grades – the weaker the results are. The achievement level is notably lower than the average when the students are at the age of 15 years or above (45 in community schools and 57 in institutional schools). Correlation between the age and achievement in institutional schools is $r = -0.06$ ($p = 0.001$) indicating a small effect size ($d = 0.14$); in community schools the correlation is non-existent ($r = -0.02$, $d = 0.045$); though the correlation varies from zero to1 ($p = 0.021$ in community schools and $p < 0.001$ in institutional schools) the connection is not notable. The reason for the zero correlation is the curvilinear shape of the phenomenon.

The dataset shows that the highest performance in Mathematics is found with the students studying at their proper age years – that is, at the age of 10 to 12 years. Otherwise the achievement decreases as the age increases. Of the sample, 27.1% students fell aside 10–12 years.

Support provided for study and student achievement

The relation between the support received for studies and achievement was analyzed based on the following question: "*Who helps you when you do not understand what you have read?"*. In the question, only one option was selected – in many cases, there might be several helpers, which cannot be detected now. The descriptive statistics of the support received are given in Tables 3.2.23 and 3.2.24.

Table 3.2.23 Support received by the student and achievement level

Support received	N	Mean	SD	CV
Tution	1,067	58.3	22.0	0.38
Mother	763	55.9	23.4	0.42
Brother/Sister	5,645	54.9	22.1	0.40
No one	341	54.3	22.5	0.41
Father	2,540	52.2	23.7	0.45
Teacher	2,366	50.3	22.7	0.45
Total	**12,722**	**53.8**	**22.7**	**0.42**

Table 3.2.24 Support provided to the student and achievement by type of school

Support received	Community				Institutional			
	N	Mean	SD	CV	N	Mean	SD	CV
Tuition	662	53.8	23.0	42.7	405	65.6	18.0	27.4
Brother/Sister	4,419	51.3	21.7	42.3	1,226	67.7	18.2	27.0
No one	284	50.6	22.0	43.5	57	72.7	14.6	20.0
Father	2,048	48.1	23.0	47.8	492	69.2	18.2	26.3
Mother	487	48.1	23.1	48.0	276	69.6	16.7	24.0
Teacher	1,907	47.1	22.2	47.1	459	63.6	19.8	31.2
Total	**9,807**	**49.8**	**22.3**	**44.8**	**2,915**	**67.3**	**18.4**	**27.3**

It seems that an external support is, in many cases, necessary for the students to gain better than the average marks on the test. However, the reality is found different in the community schools compared to the institutional schools. In the whole dataset, there is about 4 percent difference between those who do not get some kind of support (54%) and those who receive (private) tuition (58%). It is likely that the children receiving private tuition also spend more time on the homework which may explain the high score. The lowest scores are found among the students who received support from their teacher (50%); this does not necessarily mean that the teachers are in effective in providing support to the students – it may reflect also the low educational capacity in home. Those who were provided support by their father or teacher gained notably lower than the average – even lower than those not receiving tuition at all.

It shows that tuition support for students in community schools is found more contributory to raise the achievement (54%) than the support received from siblings– their brothers or sisters (51%). In institutional schools, on contrary, the highest results are found among those who have studied just by themselves (73%) or the mother or father provides support to them (70% and 69% respectively). In both cases it is found that when the teacher is the main person to provide support for study, the result also goes lower than the average. The effect of the support provided is, in any case, very low: effect size is $f = 0.10$ in community schools and $f = 0.11$ in institutional schools indicating that the difference is low or not notable.

The dataset indicates that the support provided by the mother and by brother and sister raises the achievement level more than the support provided by the father or teacher. In the whole sample, the highest achieving group is the one who receives private tuition. However, the difference between the highest and lowest performing groups is not notable. It is likely that the group receiving private tuition also spends more time on their homework, which may explain the higher score in community school students.

Availability of textbook and student achievement

The data shows that there are still some students who do not receive the textbook even by the end of academic session. Table 3.2.25 shows the descriptive statistics of availability of the mathematics textbook and the achievement.

Table 3.2.25 Availability of textbook of Mathematics and the achievement

Availability of Mathematics textbook	N	Mean	SD	CV
Yes	12,192	54.2	22.6	41.7
No	487	44.7	22.2	49.5
Total	**12,679**	**53.8**	**22.7**	**42.1**

Out of 12,679 students who responded to the question, 3.8% (3.69% in community schools and 4.34% in institutional schools) did not have a textbook available at school even up to the end of academic year. The relation between the availability of textbook and achievement is significant ($p < 0.001$) though the effect size in the whole dataset is small ($f = 0.08$). The difference in achievement is 9.4 percent (10.3 in community schools and 9.2 in institutional schools).

According to the data set, 3.8% of the grade 5 students reported that they lack Mathematics textbook. The achievement level of these students is significantly lower than those who have got the textbook.

Homework assigned/checked and achievement

Use of homework in the form of either drill, exercise, or as an evaluation tool is one of the ways to enhance teaching and learning. When homework is assigned and checked systematically, it probably boosts achievement levels. Based on the results of students' reports, it is seen that homework assignment with its checks has not been regular activity for teachers; though, within the same classroom some students have given a slightly deviant response than the other students from the same class. Some of the students reported that homework is not assigned, those cases are considered as homework not assigned; the number of students of this kind was very few. However, in the wide scope, the results seem to make sense. Statistics related to homework assigned and checked is presented in tables 3.2.26 and 3.2.27 and depicted in figure 3.2.30.

Table 3.2.26 Homework assigned and checked and the achievement

Status of homework	N	Mean	SD	CV
Given Everyday– Checked Everyday	8,794	55	22.8	41.2
Given Someday – Checked Everyday	913	55	22.7	41.0
Given Everyday – Checked Someday	1,820	50	22.0	43.6
Given Someday– Checked Some day	817	48	22.2	46.2
Given Everyday–Not checked	129	43	21.2	48.8
Not given	259	43	22.6	53.1
Given Someday–Not checked	66	39	19.4	50.1
Total	**12,798**	**54**	**22.8**	**42.5**

Table 3.2.27 Homework assigned/checked and achievement by the type of school

Status of homework	Community				Institutional			
	N	Mean	SD	CV	N	Mean	SD	CV
Given Everyday–Checked Everyday	7,084	52	22.5	43.3	1,710	69	18.4	26.8
Given Someday–Checked Everyday	516	45	20.1	45.2	397	69	17.6	25.4
Given Everyday–Checked Someday	1,323	45	21.0	46.7	497	65	17.6	27.1
Given Someday–Checked Someday	574	42	21.0	49.7	243	61	18.8	30.6
Given Everyday– Not checked	113	40	19.5	48.5	16	66	19.5	29.8
Not given	235	40	21.5	53.2	24	63	23.3	37.3
Given Someday–Not checked	56	35	17.4	49.6	10	60	17.7	29.7
Total	**9901**	**50**	**22.4**	**45.1**	**2,897**	**67**	**18.4**	**27.3**

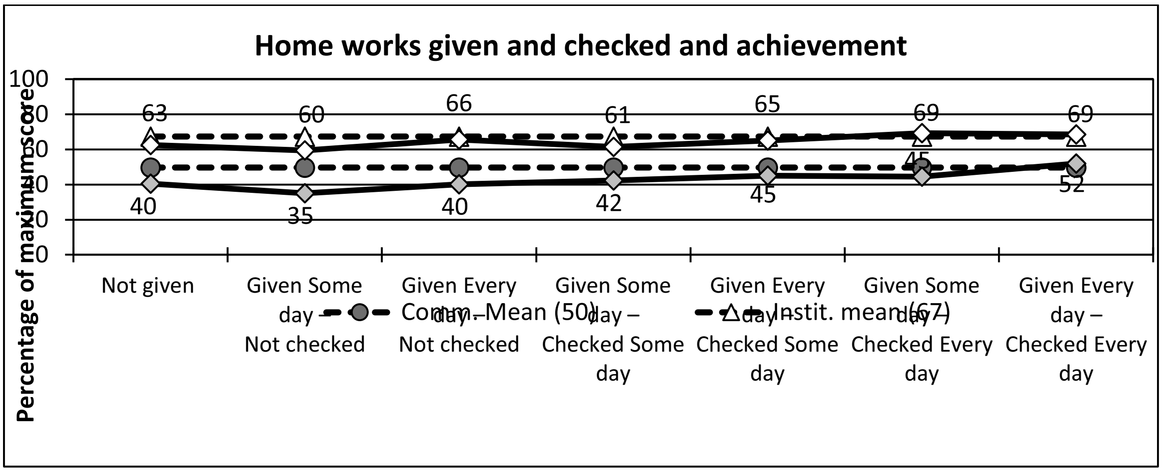

Figure 3.2.30 Relation between homework given/checked and achievement

It is evident that, if students respond that the teachers do not assign them homework or do not check those, students' achievement is notably lower (35–40% in community schools and 60-66% in institutional schools) compared to the situation when the teacher assigns homeworks and checks them (52% in community schools and 69% in institutional schools). The differences are statistically significant ($p < 0.001$). However, those groups with no homework assigned or checked are very small and hence, the effect size is small ($f = 0.18$ for community schools and $f = 0.13$ for institutional schools); grouping explains only 2–3% of the variance in the data ($\eta^2 = 0.031$ for community schools and $\eta^2 = 0.017$ for institutional schools).

The dataset indicates that if the teacher assigns and checks homework regularly, the achievement is higher than with no homework assignment and its checking. By assigning and checking homework regularly, the teacher can contribute to raise the scores up to 11 percent. In total of the whole sample, 3.5% students did not get homework or those were not checked.

Future aspiration of the student and achievement

The future plan or aspiration of the students can encourage in studies – or, in some cases, when knowing that the future plan does not require long studies, the motivation for hard work in school may be declined. The students' future plan was asked in eight categories. Those were (1) farming, (2) business, (3) teaching, (4) government officer, (5) going abroad, (6) engineer, (7) doctor, and (8) other. Future plan of the students is found to be connected strictly with the student achievement, which can be seen in tables 3.2.28 and 3.2.29 and figure 3.2.31.

Table 3.2.28 Students' future aspirations and achievement

Future aspiration	N	Mean	SD	CV
Farmer	479	39.6	23.4	59.0
Business	510	45.2	23.0	50.9
Teacher	3,914	49.0	22.1	45.1
Government officer	1,030	55.3	22.9	41.5
Going aboard	741	56.1	21.7	38.6
Engineer	2,276	56.0	21.2	37.8
Doctor	3,456	58.3	22.5	38.5
Other	488	62.5	20.9	33.5
Total	**12,894**	**53.7**	**22.7**	**42.3**

Table 3.2.29 Students' future aspiration and achievement by the type of school

Future Aspiration	Community				Institutional			
	N	Mean	SD	CV	N	Mean	SD	CV
Farmer	472	39.4	23.4	59.3	7	55.7	22.2	39.8
Business	460	43.2	22.2	51.4	50	63.7	22.3	35.1
Teacher	3,621	47.8	21.7	45.5	293	64.2	20.8	32.3
Going aboard	480	51.2	22.2	43.3	261	65.2	17.4	26.7
Engineer	1,650	52.2	21.1	40.4	626	66.2	17.9	27.1
Doctor	2,208	52.4	22.5	42.9	1,248	68.7	18.3	26.6
Government Officer	871	53.0	23.1	43.5	159	67.8	17.6	26.0
Other	207	55.1	22.8	41.4	281	68.0	17.5	25.8
Total	**9,969**	**49.7**	**22.3**	**44.9**	**2,925**	**67.2**	**18.4**	**27.4**

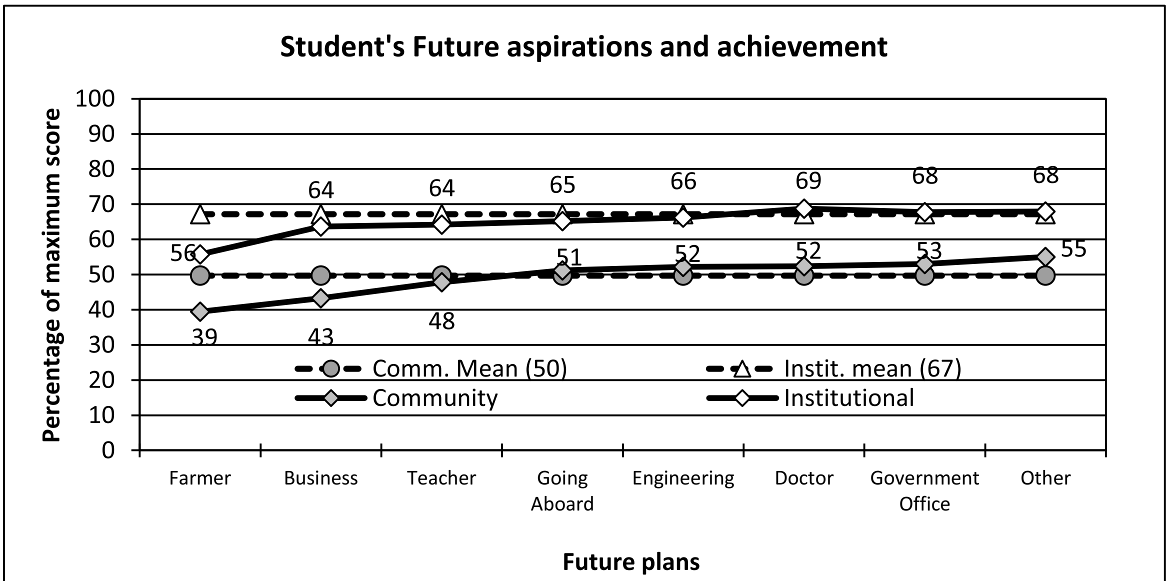

Figure 3.2.31 Relation between students' future aspirations and achievement

Some of the professions, like Engineers, Medical Doctors, and Teachers, are valued in the society – most probably because of more or less guaranteed economic prospects. This can be seen in the fact that even the weakest students in the dataset (scoring less than 20% score) are aspiring to these professions; of these low-level students (n = 1019), 10% are aspiring to be Engineer, 18% to be doctor, and 38% to be teacher. On the basis of their achievement level, this aspiration, most probably, will turn to reality to a very few of them. Another reality is that, according to the fathers' occupation, 38% of the students come from the agricultural background, but in the whole dataset, only 3.7% of the students are aspiring to involve in agricultural occupation – that is, farming. When the student knows that s/he will be continuing the family occupation that is agriculture, the learning achievement is remarkably low (39% in community school and 56% in institutional schools).

Because the professions such as Engineer, Doctor and Teacher are highly demanding, the competition for the study places will also be tough. Hence, the higher the goals the higher should be the achievement level in order to make the dream for the future occupation to be true. From this point of view, the students' future plan looks logical while comparing it to the mean achievement level. Students who are aspiring to be Engineer (56%) or Doctor (58%) really score remarkably higher than those aspiring to be Farmer (39%) or a Business Person (43%). The future plan explains 5% of the achievement level (η^2 = 0.052); the effect size is moderate indicating that the difference between the lowest and highest group is remarkable.

Interestingly, mostly desired profession, as reported, are also teachers and government officers. It might be the case that students seem to think that there is no need to be very good in Mathematics in order to be a teacher (mean is below the average), but to be a government officer the case is found opposite. It is more or less surprising that those students from the community schools who are aspiring to be the government officer have performed higher in Mathematics than those aspiring to be a doctor or an engineer. This may indicate a realistic view of children in their possibilities for the future career.

The dataset shows a relation between the aspiration of student and their achievement. As the student aspires to professional career other than farmer, business or teacher, their achievement is higher than the average. The number of students who aspire to be a government officer or an engineer or a doctor is remarkably high.

Activities in the school and student achievement

The activities of the students and the teacher determine the learning environment of the school. Bullying, for example, is one of the hindering activities for the students in the school that may affect learning. In the student background information questionnaire, several students 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 indicators. Students' impressions on schools'and teachers' activities are taken as positive indicators.

Negative activities: Bullying

Bullying is one of the problems in school that has negative consequences on the learning environment. International studies like TIMSS and PISA give a specific emphasis to study such phenomena that is seen in their background questionnaires. In the NASA 2012 student questionnaire, five questions indicate the varieties of bullying that are likely to happen in the school. All the questions were stemmed by the phrase "*Which of the following activities happened in your school in the last month?*" The students' responses are presented in tables 3.2.30 and 3.2.31 and depicted in figure 3.2.32. 'No (%)' indicates the percentage of the students' response that no such activity happened in the school and 'Yes (%)' indicates the percentage of the students who reported the particular type of bullying they experienced within a month. The fact is that 26% student ed that, during the last month of this test, something of their own was stolen which is an alarming sign in the system.

Table 3.2.30 Frequencies of encountered bullying

Type of Bullying	No (%)	Yes (%)
I was made fun of or called names	71.7	28.3
Something of mine was stolen	74.2	25.8
I was hit or hurt by other student(s)	79.0	21.0
Fellow students kept outside without involving me in activities	77.9	22.1
I was made to do things I didn't want to do by other students	84.7	15.3

Table 3.2.31 Bullying and the achievement by the type of school

Intensity of bullying	Community				Institutional			
	N	Mean	SD	CV	N	Mean	SD	CV
No bullying	4,491	53.1	22.1	41.6	1,250	67.2	18.2	27.1
20% bullying	2,293	50.2	21.9	43.7	834	67.7	17.9	26.4
40% bullying	1,568	47.7	21.4	45.0	462	67.7	18.9	27.9
60% bullying	907	44.0	22.2	50.6	279	65.3	19.1	29.3
80% bullying	313	40.9	20.4	49.9	81	65.6	19.6	29.9
100% bullying	340	36.6	22.4	61.3	11	63.5	28.5	44.8
Total	9,912	49.8	22.3	44.8	2,917	67.2	18.4	27.4

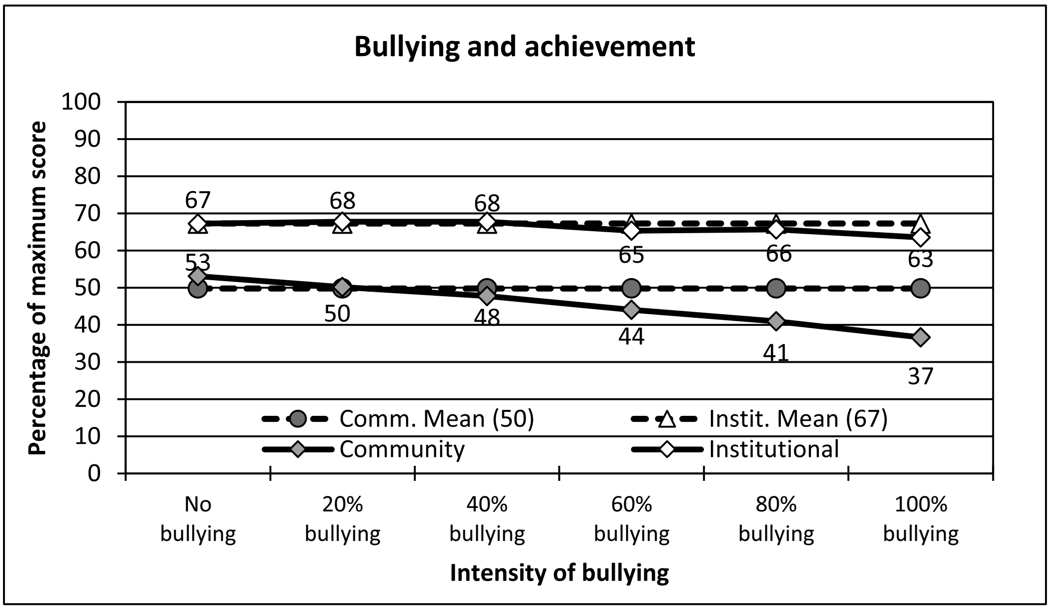

Figure 3.2.32 Relation between bullying and achievement

The sum of all five items form an indicator of bullying. Figure 3.2.32 shows the achievement of the students who encountered each category of bullying. If only one activity of bullying is reported, it is categorized as 20% bullying; and if all five activities are reported it is categorized as 100% bullying. When knowing that 45% of students did not encounter any bullying during the last month, one can infer that the remaining 56% did encounter at least one type of bullying which is a remarkable number of the students. As many as 5.8% students – 6.6% in community schools and 3.2% in institutional schools – are experiencing a severe kind of bullying (the sum of 80% and 100% bullying). This means, in practice, that more than 42,000[30] grade 5 students in Nepal are encountering physical, psychological, and social bullying every month. The number is too much even though it would not have any effect in learning outcomes. However, it is seen that learning outcomes are notably lower than the average with those 15% students who have encountered more than two different types of bullying (36–44% in community schools and 63–65% in institutional schools). The achievement gap between the students who do not experience bullying and those who encountered extreme bullying of four or five kinds is at 19 percent; though there are a few number of students who reported this kind of bullying (n = 351). The difference is statistically significant (p = 0.001) though the effect size is small or medium (f = 0.19) in community schools and very small (f = 0.04) in institutional schools. Though cases of severe bullying are rare, it is found to be quite common in schools. This negative phenomenon causes needless harm to young children and has to be rooted out from schools.

The dataset reveals that an alarmingly high number of the students (55%) have encountered some kind of bullying in school within the a month and 5.8% of students have been experiencing a severe kind of bullying. This means that more than 42,000 grade 5 students in Nepal have been encountering physical, psychological, and social bullying every month. The phenomenon has been affecting the learning outcomes in almost all the groups of the students who experienced bullying, so, all possible efforts are required to root out the phenomenon from the schools.

Positive activities in school

The activities that can boost the learning achievement of the students are categorized as positive activities. Such positive activities at school were asked to the students in two sets of questions described in table 3.2.32. The table shows the responses of the students in four categories; the responses are in the 4 point rating scale anchored to fully disagree (0) and fully agree (3). Generally speaking, the grade 5 students express content with the school and student related activities in school. However, remarkably high number of students (12.0%) reported that they felt that the teacher was not treating them fairly. The same phenomenon was seen also in 2011 datasets with grade 8 students: 11% students in Mathematics, 12% in Nepali and 13% in Social Studies (see 3.1, 3.2 and 3.3; ERO, 2013) felt unfair behavior of teacher.

---

[30] According to the “Primary level total enrollment in all types of schools by district, Flash I_2012–2013”, there were 731,573 grade 5 students. 5.8% of these is 42,431 students.

Table 3.2.32 Students' response on teacher and school-related activities in schools

Teacher and Students activities[1]	Respondents in %(valid percentage)			
	Fully agree	Partially agree	Partially disagree	Fully disagree
q28a: I like coming and staying in school	91.9	4.8	1.1	2.1
q28b: Students in my school like me	91.9	4.8	1.1	2.1
q27a: Students get along well with most teachers	86.9	9.4	1.5	2.2
q28c: Teacher in the school care about the students	86.9	8.8	1.8	2.5
q27b: Most teachers are interested in student's well-being	86.7	8.5	1.9	2.9
q27d: If I need extra help, I will receive it from my teacher	85.7	8.9	2.5	2.9
q27c: Most of the teachers really listen to what I have to say	83.3	11.2	2.4	3.1
q27e: Most of my teachers treat me fairly	73.3	14.7	4.3	7.7
Average	**85.8**	**8.9**	**2.1**	**3.2**

1) *The activities are ordered on the basis of percentage in "Fully agree".*

It has been further analyzed by recoding the variables into two categories (2–3 = 1, that is, agree and 0–1 = 0, that is, disagree). Furthermore, the sum of eight indicators is converted into the percentage to analyze the level of positive activities and its relation to achievement.

Because almost 70% of the students were ultimately positive, it is difficult to divide the score other than into quartiles, that is, four attitude groups. These boundaries and descriptive statistics are seen in tables 3.2.33 and 3.2.34 and illustrated in figure 3.2.33. 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 positive activities and achievement is positive ($r = 0.22$), statistically significant ($p < 0.001$) and moderately high ($d = 0.50$).

Table 3.2.33 Teacher and school-related activities and the achievement

Percentage of positive actions	N	Mean	SD	CV
50% or lower	1,351	44.2	23.5	53.1
62.5 - 75.0%	1,069	42.5	20.9	49.3
87.5%	1,812	51.5	21.7	42.2
100%	9,482	56.1	22.6	40.2
Total[1]	**13,714**	**53.3**	**22.9**	**43.1**

1) *Total excludes the cases without giving their opinion (missing n = 857).*

Table 3.2.34 Teacher and school-related activities and the achievement by school type

Percentage of positive actions	Community				Institutional			
	N	Mean	SD	CV	N	Mean	SD	CV
50% or lower	1,188	41.8	22.7	54.3	163	61.9	21.8	35.2
62.5 - 75.0%	953	40.7	20.6	50.6	116	56.7	18.2	32.0
87.5%	1,274	45.6	20.5	44.8	538	65.3	18.0	27.6
100%	7,242	52.3	22.4	42.9	2,240	68.4	18.3	26.7
Total[1]	**10,657**	**49.3**	**22.5**	**45.7**	**3,057**	**67.0**	**18.6**	**27.7**

1) *Total excludes the cases without giving their opinion (missing n = 857).*

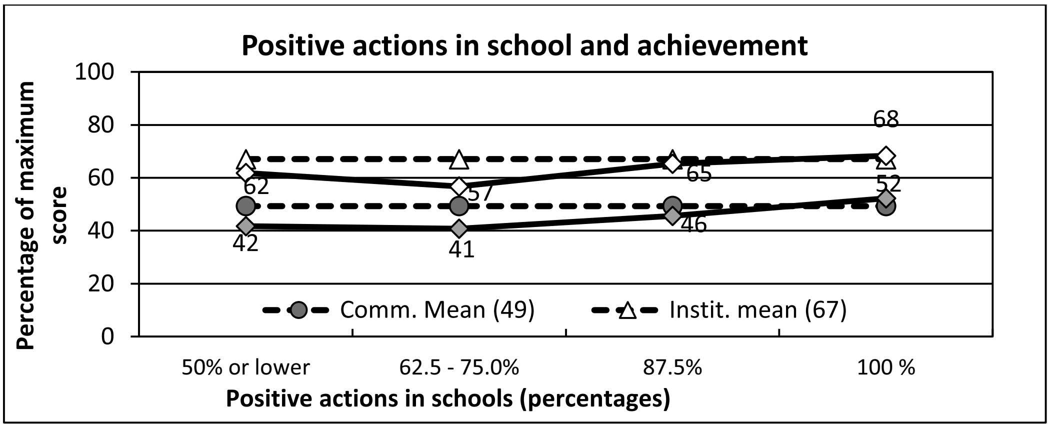

Figure 3.2.33 Relation between positive actions in school and achievement

The dataset suggests that when the students feel that the actions of the teachers and the schools are ultimately good, the Mathematics results are better than average (52% in community school and 685 in institutional schools). Students with a negative feeling in five or more of the eight indicators (at most 75% of the total) are in great danger of achieving much lower than the average in Mathematics. In the whole sample, 12% students feel that their teachers do not treat them fairly.

The data shows that there is a positive relation between the students' feeling of the teacher and school related activities and the achievement.The increase in achievement is directly proportional to the increase in the intensity of such activities. After dividing the indicator into four groups, the differences between the groups are statistically significant (p < 0.001), however, the effect size is moderate (f = 0.21 in community schools and f = 0.15 in institutional schools); the difference between the most positive group and the most negative group is notable (11–13 percent). Only when the students are extremely positive towards school and teachers' behavior, the learning achievement is higher than the average. Students with a negative feeling in five or more of the eight indicators (at most 75% of the total) are in great danger of achieving much lower than the average in Mathematics. Attitude explains that achievement is better in English than in Mathematics; however, dataset shows that when the students have 100% positive attitude, they can achieve much better than other.

3.2.4. Synthesis of Analysis of Results

Several individual student and geographical related factors have been discussed above, which individually explain the difference in achievement between the students. These factors are compiled in table 3.2.35. One can note that, except the gender, all the factors explained statistically significant difference between the groups when analysed individually.

Table 3.2.35 Individual variables handled in the text and their effect in one-way ANOVA

Variable and values[1]	**Leverage[2]**	**Eta squared[3]**	**Effect size[4]**
Ecological zone (1 = Mountain, 2 = Hill, 3 =Tarai,4 = Valley)	+19.2	0.100	0.33
Development Region (1= Eastern, 5 = Far-Western,6 = Valley)	+23.6	0.119	0.37
School location (0 = Rural,1 = Urban)	+14.8	0.059	0.27
School type (0 = Community,1 = Institutional)	+17.7	0.103	0.34
Gender (0 = Girls,1 = Boys)	+1.73	0.001	0.03
Ethnicity (1 = Janjati, 2 = Dalit, 3=Madhesi, 4 = Brahman, 5 = Cheetri, others)	+10.9	0.028	0.17
Language at home (1 = Nepali, …, 12 = Other)	+27.5	0.033	0.18
Mother's education (1 = Illiterate, …, 8 = Above MA)	+18.1	0.035	0.19
Father's education (1 = Illiterate, …, 8 = Above MA)	+21.3	0.048	0.22
Mother's occupation(1= working abroad,…, 8 = working at other home)	+13.4	0.028	0.17
Father's occupation(1= working abroad,…, 8 = working at other home)	+16.8	0.058	0.24
Home possessions(sum; max 11)	+14.9	0.031	0.17
Home accessories(sum; max 3)	+18.9	0.076	0.28
SES (sum; max 7)	+23.5	0.100	0.33
I do jobs at home (1= not at all, …, 4 = more than 4 hours)	+10.8	0.032	0.18
I work at a paid job (1 = not at all, …, 4 = more than 4 hours)	+10.1	0.032	0.18
Attitude "Utility in Mathematics" (sum; max 15)	+18.6	0.062	0.25
Age	+9.8	0.009	0.09
Who helps you …? (1 = Father, …, 6 =Teacher)	+8.0	0.010	0.10
Do you have textbook of Math subject (0 = No, 1 = Yes)	+11.7	0.006	0.07
Homework (0 = not given, …, 6 = Given everyday, checked everyday)	+11.7	0.006	0.07
Bullying (sum; max 5)	-18.7	0.050	0.23
Positive activities in school (sum; max 8)	+22.6	0.028	0.17

1) *The order of the variables is the same as handled in the Sections above; 2)Difference between the lowest and highest group-mean; 3) On the basis of one-way ANOVA; 4)Cohen's f*

On the basis of Univariate ANOVA, Development region, followed closely by the school type, socio-economic status, and Ecological zone, are seen to have been the most effective single factors in affecting the achievement level of the student as effect sizes are $f = 0.37$, $f = 0.34$, $f = 0.33$, and $f = 0.33$ respectively. Some of these variables in table 3.2.35 are found to be strongly related to each other and hence not adding value to explain why some students are performing much better than others. In what follows, the synthesis of the analysis is done in two ways: Multilevel Modelling and collection of statistically best factors by using the Regression modelling. For the analysis, grouping factors are changed to be, so called, Dummy variables when needed; for example, Ecological zone is transformed into three variables: variables indicative for Mountain, Hill and Tarai.

Modelling the overall achievement by Multilevel Modelling

The datasets collected from schools are always clustered, that is, the students within the school are more alike with each other in comparison with the case that the same number of students would have been randomly sampled totally from the population.

Table 3.2.36 Individual variables and their effects in multilevel analysis

Source[1]	df_1	df_2	F	Sig.
Intercept	1	718.4	201.22	<0.001
Ecol zone Mountain Dummy (Mountain =1, other = 0)	1	490.3	9.57	0.002
Ecol zone Hill Dummy (Hill =1, other = 0)	1	500.0	0.23	0.635
Dev region Central Dummy (Central =1, other = 0)	1	489.7	0.33	0.569
Dev region Western Dummy (Western =1, other =0)	1	484.7	2.63	0.106
Dev region Mid-Western Dummy (Mid-Western =1, other = 0)	1	482.3	7.06	0.008
Dev region Far-Western Dummy (Far-Western =1, other =0)	1	491.3	1.86	0.173
Dev region Valley Dummy (Valley = 1, other = 0)	1	482.8	26.62	< 0.001
School type (0 = Community, 1= Institutional)	1	497.9	8.50	0.004
School location (0 = Rural, 1 = Urban)	1	479.6	2.26	0.133
Gender (0 =Girls,1 = Boys)	1	9723.9	35.10	< 0.001
Caste Brahman & Chhetri Dummy(Brahman & Cheetri =1,other = 0)	1	9878.0	33.74	< 0.001
Caste Janjati Dummy (Janjati = 1, other = 0)	1	9989.4	1.18	0.276
Caste Madhesi Dummy (Madhesi = 1, other = 0)	1	10013.6	0.92	0.338
Caste Dalit Dummy (Dalit = 1, other = 0)	1	9871.2	0.06	0.810
Language Dummy (Nepali = 1, other = 0)	1	10083.8	0.74	0.390
Home chores Dummy 1or 2h (1 – 2 hours = 1, other = 0)	1	9841.6	36.09	< 0.001
Paid work Dummy (0 hours = 1, other = 0)	1	9869.2	47.02	< 0.001
Attitude "Utility in Mathematics" (Sum; max 15)	15	9799.7	7.34	< 0.001
Age Dummy 10 to12y (10 – 12 years = 1, other = 0)	1	9782.8	10.20	0.001
Help by Father Dummy (Father = 1, other = 0)	1	9840.7	2.49	0.115
Help by Mother Dummy (Mother = 1, other = 0)	1	9805.8	4.19	0.041
Help by Brother &Sister Dummy (Brother/Sister = 1, other = 0)	1	9831.6	1.37	0.242
Help by Tuition Dummy (Tuition = 1, other = 0)	1	9848.1	0.21	0.645
Help by Teacher Dummy (Teacher = 1, other = 0)	1	9875.8	5.58	0.018
Do you have a textbook of Mathematics (Yes = 1, No = 0)	1	9747.6	36.52	< 0.001
Homeworks Given and Checked Dummy (everyday =1, other = 0)	1	9844.8	17.18	< 0.001
Future plan Farmer Dummy (Farmer = 1, other = 0)	1	9775.0	16.73	< 0.001
Bullying (Sum; max 5)	5	9807.2	12.52	< 0.001
Positive Activities in school (Sum; max 8)	8	9754.9	4.99	< 0.001
SES[2] (Sum; max 6)	6	9778.2	8.15	< 0.001

1) *Ecogical zone, Developmental Region, Caste, and Helper of the homework, one of the classes needs to be omitted in the analysis because of singularity reasons. Hill zone, Mid-Western region, Others ethnicity/caste, and No one helps are omitted; these showed no statistical significance in Regression analysis. 2)Shortened SES; school type is taken away; this enables estimating the parameters for school type.*

Multilevel modeling is used to acquire the correct test values while taking into account the clustering effect of the school. Table 3.1.36 shows the corrected estimates for the variables in table 3.2.35 while modeling the phenomenon in a multivariate manner; by using the multivariate ANOVA, the hidden commonalities of the factors are revealed.

When taking into account the clustered structure in the dataset and the conjoint effect of the factors, quite many of the factors do not show main effect. Such variables are school location and language of the student (when dividing it to Nepali – Other).

Statistically the best factors by using regression modelling

Traditional linear regression analysis with stepwise regression is used to explain the total score by the same variables as are above (see table 3.2.35). Table 3.2.37 shows the results.

Table 3.2.37 Statistically the best model of linear regression analysis explaining the average of student achievement (Using stepwise method)

Model	Coefficients				
	Unstandardized Coefficients		Standardized Coefficients		Sig.
	B	Std.Error	Beta	T	
(Constant)	16.15	1.897		8.51	< 0.001
School type (0 = Community, 1 = Institutional)	6.11	0.634	0.12	9.63	< 0.001
Dev region Valley Dummy (Valley = 1, other = 0)	12.62	0.685	0.22	18.42	< 0.001
Positive Activities in school (Sum; max 8)	0.12	0.014	0.08	8.06	< 0.001
Paid work Dummy (0 hours = 1, other = 0)	5.06	0.452	0.10	11.21	< 0.001
Bullying (Sum; max 5)	-1.61	0.157	-0.09	-10.24	< 0.001
Attitude "Utility in Mathematics" (Sum; max 15)	0.57	0.063	0.09	9.07	< 0.001
Ecol zone Mountain Dummy (Mountain = 1, other =0)	5.34	0.652	0.08	8.19	< 0.001
Home chores Dummy 1or 2h(1–2 hours = 1, other = 0)	3.15	0.404	0.07	7.81	< 0.001
Dev region Mid-Western Dummy (Mid Western =1, other = 0)	-1.54	0.681	-0.02	-2.26	0.024
Homework Given& Checked Dummy(everyday = 1, other = 0)	4.32	0.677	0.06	6.38	< 0.001
School location (0 = Rural, 1 = Urban)	3.30	0.571	0.06	5.79	< 0.001
Do you have a textbook of Mathematics(Yes =1,No = 0)	5.71	1.048	0.05	5.45	< 0.001
Future plan Farmer Dummy (Farmer = 1, other = 0)	-6.20	1.112	-0.05	-5.57	< 0.001
Language Dummy (Nepali = 1, other = 0)	-3.64	0.452	-0.08	-8.04	< 0.001
Ethnicity Brahman & Chhetri Dummy (Br.& Ch.= 1, other =0)	3.13	0.43	0.07	7.27	< 0.001
Dev region Western Dummy (Western = 1, other = 0)	5.04	0.629	0.09	8.02	< 0.001
Gender (0 = Girls, 1 = Boys)	-1.92	0.388	-0.04	-4.96	< 0.001
SES[1] (Sum; max 6)	0.60	0.148	0.05	4.04	< 0.001
Help by Teacher Dummy (Teacher = 1, other = 0)	-2.13	0.497	-0.04	-4.28	< 0.001
Dev region Central Dummy (Central = 1, other = 0)	2.44	0.551	0.05	4.43	< 0.001
Age Dummy 10 to 12y (10 – 12 years = 1, other = 0)	1.28	0.458	0.03	2.80	0.005

The model in table 3.2.37 can be interpreted as follows: The average mean of the students is 16.15% – assuming that the student was in the lowest group in all the factors. If the student was from the Kathmandu Valley, the score is, on average, + 12.62 percent higher (note the sign of the coefficient). Additionally, if the students came from an institutional school, the additional score was 6.11, if s/he did not work on a paid capacity, the additional score was 5.06 percent, and so on. On the other hand, if they face bullying, with each step (of five) the achievement level would drop 1.61 percent; the difference between the lowest and highest group is $5 \times 1.61 = 8.05$ percent.

3.3 Summary of Findings

The main findings of NASA 2012 in Mathematics for the students of grades 3 and 5 have been summarized under three sub-headings: Basic results, Equality indicators, and Selected explanatory factors. They are presented as follows:

3.3.1 Basic results

- The grade 3 result in mathematics is not distributed normally. However, it is distributed more normally in grade 5. There are three distinctive student populations: low and high performing students from community schools and high performing students from institutional schools.
- The students in institutional schools perform well and the students in community schools form two kinds groups of schools: high-performing schools and low-performing schools. The variation among the community schools with regard to achievement score is remarkable, ranging from almost 0 to 100% in both grades.
- Learning achievements are the weakest in the content areas of Algebra and Numeracy and the highest in Arithmetic and Geometry. The achievement in Algebra is remarkably lower (40%) compared to that in Arithmetic (61%) and Geometry (60%). The differences between the content areas are similar in community schools and in institutional schools. Likewise, the average achievement of grade 5 Mathematics is 54%. The learning outcomes are weaker in the content areas of Numeracy (44%) and Algebra (49%); higher in Geometry (57%) and Arithmetic (54%). The differences between the content areas are wider in community schools than in institutional schools.
- In grade 3, students' ability to solve practical problems is quite low (53%). Students are better in recalling type of questions (59%) and comprehension (64%). Similarly, in grade 5, only 37% of the maximum scores of tasks requiring higher ability were obtained. Students are much better in the recalling type of questions (65%), but 20% students were not able to solve any of the tasks requiring higher ability. In both grades, students in institutional schools are found to be more able to solve practical and complex problems than their peers in community schools.
- In grade 3, students are found good in recognizing the correct answer and in recalling the learnt facts (62%), whereas they are weaker in productive type of items (48%). In many cases, the students did not even start to answer the open ended questions and, hence, the lower score. Similarly, in grade 5, students are performing well in recognizing the correct answer and in recalling simple facts from the texts, fundamental thinking, the basic interpretation of paragraph, table, and chart, and a few steps of logical thinking. They are much weaker in solving verbal mathematical problems, calculating the answer in more than one steps, constructing the geometrical shapes, solving the linear equation of one variable etc. In many cases, the students made an

attempt to solve such problems but the skills were not high enough for solving the problems to obtain highest marks.

- Compared to the 1995 results, the number of high performing students in grade 3 has increased in 2012. The gap between boys and girls is in the same level. The students in the Mountain zone and Far-Western region score higher, and the students from Gurung/Magar/Tamang communities score remarkably higher in 2012. In grade 5, students in Mountain zone and Western and Far-Western regions score remarkably higher but in the Tarai zone and Eastern region the results are lower. Compared to the 1999 results, the gap between boys and girls has been reduced. Dataset from 14 years back indicates that the difference in Mathematics achievement between the students in community and institutional schools in the Kathmandu Valley has not changed significantly.
- The average Mathematics proficiency in Nepal is lower than the international average in comparison to TIMSS standard. The students in the institutional schools are at the middle or even higher achievement level than the international average in grade 3. However, in grade 5, the average mathematical ability is somehow the same as of the grade 4 students in the international dataset. The students in the community schools perform unexpectedly low.

3.3.2 Diversity factors and equality

- There is a wide difference between the districts when it comes to the equal opportunities of children to reach the pre-set goals in Mathematics.The results are very high in the districts where the proportion of institutional schools with high socio-economic status is high. Interestingly, there are also districts which can produce almost equally high results with no institutional schools in the sample.
- From the dataset of grade 3 Mathematics, except for Parsa district (73%), the outperforming four districts are from the central region and specifically the Valley area: Kathmandu (79%), Bhaktapur (77%), and Lalitpur (72%). Quite high result was obtained also in Kaski (70%), Humla (70%), and Solukhumbu (70%). Out of the ten lowest performing districts, five had no institutional schools in the sample. In grade 5 Mathematics, students performance was very low in Udayapur (37%), Khotang (41%), and Dhankuta (45%) from the Eastern region; Sindhuli (45%) and Mahottari from the Central region; Bardiya (42%), Salyan (45%) and Rolpa (44%) from the Mid-Western; and Achham (44%) from Far-Western region. Out of the five highest performing districts, three are from Valley: Kathmandu (71%), Bhaktapur (69%) and Lalitpur (60%). Remaining two districts are Kaski (64%) from Western region and Humla (60%) from Mid-Western region.
- There is a moderate difference in the students' performance across four Ecological zones within both community and institutional schools. Students in the Kathmandu Valley outperform the other in both grades. Mountain is also following the Kathmandu valley in grade 5 achievement score.
- There is vast inequality across the Development regions regarding the children’s opportunities to reach an adequate level of achievement in Mathematics. Especially the

wide difference between community schools in the Valley and in the rest of the country (27 percent in grade 3 and 20 percent in grade 5), is an alarming sign of inequality of opportunities in learning Mathematics. There are also wide differences across the regions within institutional schools; the difference in student performance in institutional schools between the Valley and Central region is the highest (17 percent).

- On average, the students in institutional schools outperform the students in community schools. The difference is highest in the areas of Algebra and Numeracy in both grades. This variance can be explained partly by much higher socio-economic input into students' life and strict selection of the students in institutional schools.
- The students in the urban community schools have gained 9 percent more than the rural areas. Excluding schools from the Kathmandu Valley, the difference is practically zero. In the case of institutional schools, there is no difference between the rural and urban areas. From educational equality point of view, this is a good sign. However, in grade 5, students in the urban community schools score 6 percent more than the students in the rural areas. Excluding schools from the Valley, the difference is 3 percent. Remarkable difference is caused by schools from the Kathmandu Valley. In the institutional schools also, there is wider difference between the rural and urban areas even though the Kathmandu Valley students are excluded.
- There is educational inequality within the language groups in possibilities of learning Mathematics. In grade 3 community schools, the students from Magar (79%) and Tamang (73%) backgrounds perform very high in Mathematics while the students from Tharu (50%) and Gurung (51%) background perform much lower. In grade 5, Magar (65%) and Tamang (62%) communities perform very high in Mathematics while the students from Gurung (38%) and Limbu (41%) communities perform very low. The differences between the language groups are remarkable in both grades.
- There are statistically significant, though, not necessarily remarkable differences between the ethnicities/castes in mathematics. In grade 3, Dalit (53%) and Madhesi students (54%) as well as "Other" ethnicities/castes (50%) are performing significantly lower than Brahmin, Chhetri and Janjati. Dalit students perform lower especially in the Central Mountain area (35%) in institutional schools and in the Eastern (45%) and Mid-Western Tarai (39%), and Mid-Western Hill (43%) in community schools. In grade 5, Dalit (50%) and Madhesi (51%) as well as Janjati students (52%) are performing somehow lower than Brahmin and Chhetri students. Dalit students perform lower especially in the Mid-Western, Far Western Tarai and Hill areas.
- The differences between boys and girls in Mathematics proficiency are not so high – and in many cases non-existent. In grade 3, the boys of institutional schools from Mountain zone are seen to be outperforming girls with 13 percent. Also, in the "Other" ethnic groups, the girls are seen to be outperforming the boys by 7 percent. Otherwise the differences are very small. In grade 5, difference in achievement score of institutional school students are statistically significant in total score and Arithmetic, the effect sizes are very small indicating that the differences are not at all remarkable. The girls tend to slightly out-perform boys in institutional schools, and boys are slightly

outperforming girls in the community schools. Dalit girls are much better than boys in institutional schools. From equality point of view, this is a positive sign.

3.3.3 Selected explanatory factors

- Parents' educational level predicts the children's future achievement level in Mathematics. Especially the achievement level of students is low when the father or mother or both of them are illiterate or just literate in comparison to the parents having higher educational qualifications.
- Either economic or educational status or both at home helps children to increase their Mathematics proficiency. If the father or mother or both of them are involved in agricultural or related occupation, or if they are working abroad, the students' achievement in Mathematics is significantly lower than those with the other occupational groups.
- When children have very few home possessions – none of the 11 – the achievement level is remarkably lower than the national average (41% in grade 3 and 47% in grade 5). In grade 3, with 4–10 possessions, the average score is much higher (> 63%) and, in any case, higher than the national average. The same is true of home accessories: When none of the accessory indicators out of three is met, the results are remarkably lower (54%) than when all three were met (72%). In grade 5, with nine to ten possessions, the average score is very high (61–62%) compared to the national average. The same is true of home accessories: when none or only one accessory indicator out of three is met, the results are lower than average (48–51%), and when two or more are met, the results are remarkably higher (58–66%).
- Socio-economic status plays a significant role in Mathematics achievement in Nepal. The difference between the lowest and highest SES groups is remarkable (30% points in grade 3 and 23% in grade 5). This means that if the SES of the lowest performing students is raised into a decent level, that is, in practice, if the problem in parents' low educational level is addressed, the results in these groups will also raise remarkably. Especially challenging is the situation in the families where the father or both parents are illiterate or just literate and they work in agriculture or abroad.
- Either working for a paid job or for more than two hours per day in unpaid household work outside school reduces the school achievement of the student. However, a decent amount of household work up to two hours per day does not hampers the students' learning in Mathematics.
- Positive attitude towards the subject correlates positively with the achievement in mathematics. The better achievement is more probably a consequence of more positive attitude rather than other way round. There is no difference between the community and institutional schools in this matter.
- The highest performance in Mathematics is found among the students studying at their proper age, that is, at the age of 8 to 10 years in grade 3 and 10 to 12 years in grade 5. Otherwise the achievement gets lower as the age increases.
- The support provided by the brother, sister, mother and teacher raises the achievement level more than the support provided by the father or teacher. In the whole sample, the

highest achieving group is the one who receives private tuition. However, the difference between the highest and lowest performing groups is not notable. It is likely that the group receiving private tuition also spends more time on their homework, which explains the higher score.

- It is evident that 5.8% of the students lack the proper textbook of Mathematics in grade 3 and 3.8% students in grade 5. The achievement level of these students is significantly lower (54%) than those who have access to the textbook (61%) in grade 3. Similarly, achievement level of these students who do not have access to the textbook is significantly lower than those who have textbook in grade 5.
- If the teacher assigns and checks the homework regularly, the achievement is higher than with no homework assignment. Assigning and checking of homework regularly can raise the scores up to 13 percent in grade 3 and 11 percent in grade 5.
- There is a connection between the aspiration of student and their achievement. As the students aspire to join a professional career other than farmer, business or teacher, their achievement is higher than the average. The number of students aspiring to be government officer or engineer or doctor is remarkably high.
- In grade three, a large number of students (54%) have encountered some kind of bullying in school within the last month where 9.7% students were experiencing a severe kind of bullying. In grade five, 55% students have encountered some kind of bullying in school within the last month, and 5.8% of students were experiencing a severe kind of bullying. This means that more than 83,000 students in grade 3 and 42,000 in grade 5 in Nepal are encountering physical, psychological and/or social bullying every month. The phenomenon is found to have been affecting the learning outcomes in almost all the groups of the students who felt bullying, so all possible efforts have to be put to root out the phenomenon from the schools.
- When the students feel that the actions of the teachers and the schools are ultimately good, the Mathematics results are better than average (59% in community school and 77% in institutional schools in grade 3, 52% in community school and 68% in institutional schools in grade 5). At the other extreme, in feeling ultimately negative of such actions, the results are far below the average (41% in community schools and 58% in institutional schools) in grade 3 database. In the grade 5 database, students with a negative feeling of five or more of the eight indicators (at most 75% of the total) are in the verge of achieving much lower than the average in Mathematics.

Chapter 4: Assessment Results in Nepali

Nepali as a school subject assessed systematically and frequently in the National Assessments of student achievement (NASA) in Nepal. In the assessment of 2011 (see ERO, 2013), the grade 8 students were assessed – now the grade 3 and 5 students are chosen for the assessment. The frequent assessment is motivated by the fact that the value for Reading and Writing skills are ranked high in the modern society. The Reading skills or the "Reading proficiency" skills are utmost important in the societies where an increasing number of information is given in a written form, or in tables, graphs and plots, as well as in strict numbers. The reading and writing skills are demanding to be able to acquire information of the surrounding world and to adequately communicate in the everyday life. We are expecting that the modern citizens should be able to handle such information to survive in the information flow. Hence the value of commanding Reading and Writing in the modern world is high. In Nepal, though for many social groups, Nepali language is not the first language, it is used in widely as lingua franca by a large population in everyday life and as only the official language.

Nepali Language proficiency at grade 3 and 5 has been assessed more or less systematically and infrequently in the national assessments of student achievement (NASA) in Nepal. The results of the previous national assessments (see BPEP, 1995; EDSC, 1997, 2001; BPEP,1998; CERID, 1998; EDSC, 1999; CERSOD, 2001; EDSC, 2008; Fulbright, 2008) are not fully comparable with each other because of the missing linking procedure between the tests so the proficiency levels are also not comparable in the absolute sense (for example, percentages of correct answers). However, the *proportional* differences between the groups and content areas are compared in what follows.

The Nepali achievement results in Nepal are linked to the international Progress in International Reading Literacy Study (PIRLS) data. One published, two-paged, text ("Antarctica: Land of Ice", see Foy& Kennedy, 2006, 2–3) with four items and their item parameters (Foy & Kennedy, 2006, 5–7; Martin, Mullis, Kennedy, 2006; 277, 284) were borrowed from the international bank in order to compare the results in Nepal with an international standard (see in detail Section 2.4). The text is originally aimed for grade 4 students. The released text is in English and, hence, for the Nepali testing, it was translated into Nepali by a language expert. Based on pre-test result and recommendation of subject committee only one reading paragraph ("Antarctica: Land of Ice", see Foy & Kennedy, 2006, 2–3) and four objective type questions were used in grade 3 students. The same text and the same questions were used for Nepali and English achievement tests of grade 5. By using the Item Response Theory (IRT) modeling, the PIRLS dataset and the NASA 2012 Nepali dataset were linked together to give unique new information on the Nepalese students' proficiency level in Nepali. Another type of international comparison was done on the basis of Common European Framework in Reference for Language (CEFR) testing. CEFR classification was used to assess the criterion-based proficiency in Nepali language.

Assessment results in Nepali are based on the achievement test conducted among 19,501 students of grade 3 in 849 sample schools, and among 13,971 students of grade 5 in 569 sample schools selected from 28 sample districts. The schools represented all Ecological zones and Development regions, rural and urban areas as well as community and institutional schools. Basic results of assessment as well as disaggregated results based on various strata and diversity are included in the analysis. Besides, the extent of influence caused by a number of related factors in student achievement has been scrutinized. Section 4 of this chapter presents the results of grade 3 and section 4.2 presents results of grade 5. The summary and findings of the assessments of both the grades are consolidated and presented together in a separate section 4.3.

4.1 Assessment Results in Nepali for Grade 3

This section analyses the assessment results of Nepali subject at grade 3. It starts with the analysis of basic results including overall distribution of scores. Then it presents the results in the different content areas of Nepali subject in general and goes to the analysis of the effects of different diversity factors from equality point of view. It also analyses the influences of factors explaining the differences in the achievement in Nepali subject.

4.1.1 Basic Results in Nepali for Grade 3

As the basic results of assessment in Nepali, this sub-section analyses the overall distribution of scores, results in various content areas and various levels of cognitive domains of Nepali subject, result variations in item types, and comparison of results with previous assessments as well as with international assessment results.

Distribution of overall results

The sample size in Nepali subject is big enough to form the Normal distribution (over 19,000 students). However, figure 4.1.1 shows that the total score is not normally distributed.

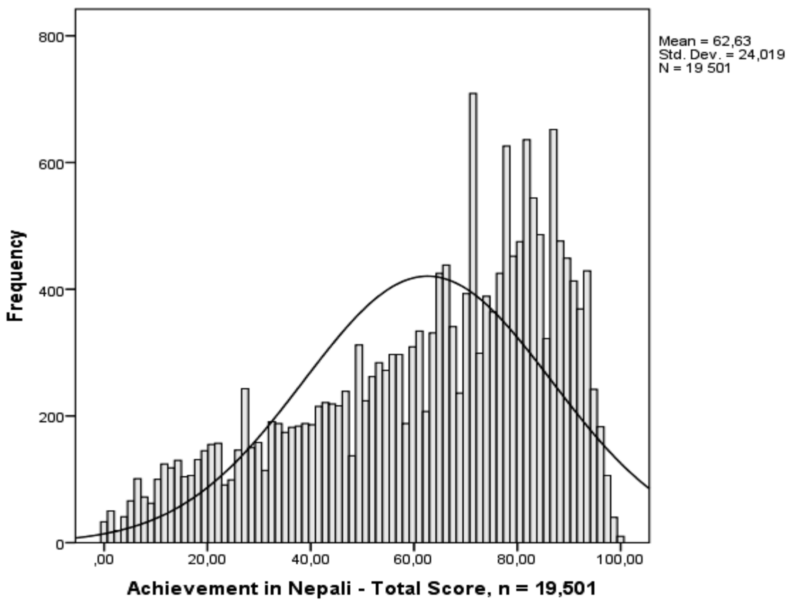

Figure 4.1.1 Final equated scores of Nepali subject

There are two different, clearly distinctive, normal populations in the dataset: students from the community schools and students from the institutional schools (fig. 4.1.2).

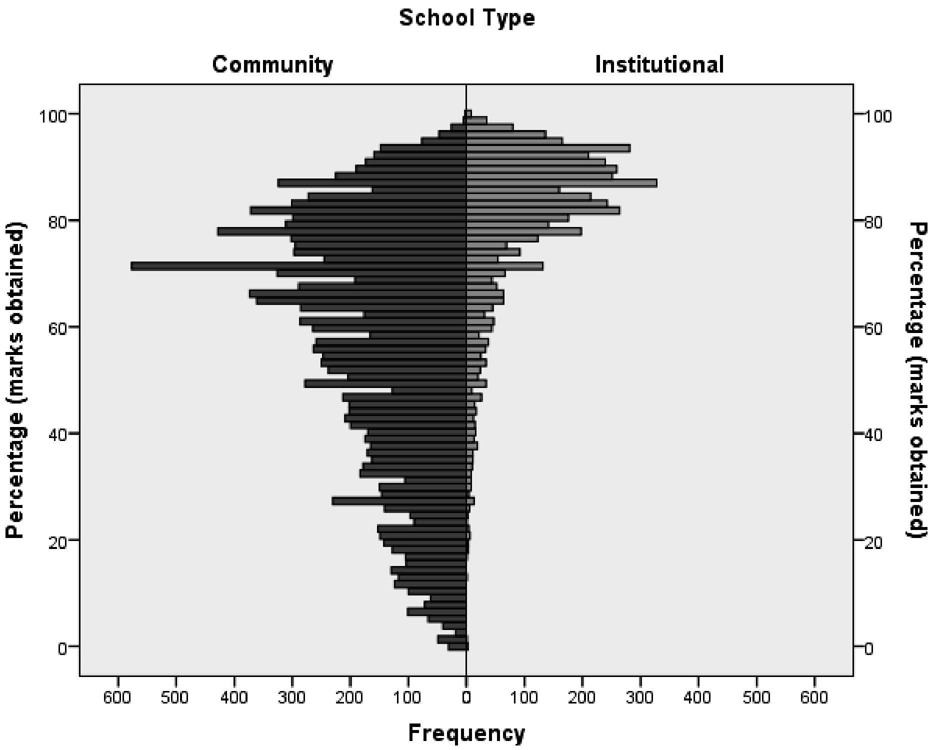

Figure 4.1.2 Distribution of the students' means by school type

In figure 4.1.2, the distribution on the left hand side shows that of the community school students and the distribution on the right hand side that of the institutional school students. The main system is tilted towards the lower performing level because the main population comes from the community schools, whereas institutional schools performed better compared to community schools. One notices two distinct population of students in community schools at the same time, in one extreme, there exists a large number of low performing students and on the other there is also a large number of high scoring population getting equally high score as in institutional schools. Figure 4.1.2 also shows that the students in community schools are varying from the low performer to the highest performer, whereas most of the students from institutional schools are performing high or medium.

The schools – not only the students – are clearly divided into two categories: the high performing and the low performing schools. In figure, 4.1.2, both categories are slightly skewed: community schools are skewed to include more low-performing and institutional schools are skewed to include more high-performing students.

By analysing the distribution further with the scatter plot, and combining the socio-economic status (SES) with the average achievement in the school, figure 4.1.3 shows the schools into two groups based on students' achievement. Most of the institutional schools (triangle) are performing well and the average SES is very high, whereas community schools (circle) vary from very high performing to very low-performing schools.

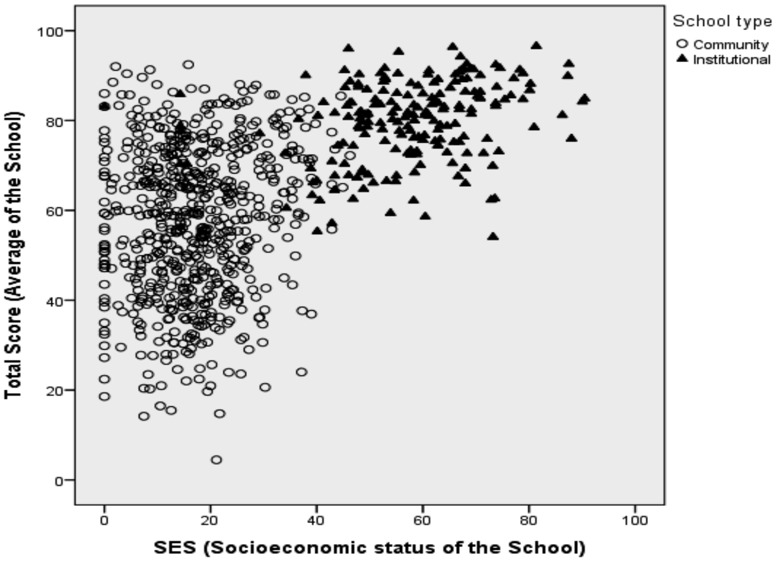

Figure 4.1.3 Achievement and socio-economic status and type of schools

The dataset shows that the grade 3 population in Nepali is not distributed normally. There are three distinctive student populations: low and high performing students from community schools and high performing students from institutional schools. The variation between the community schools is remarkable.

Achievement in various content areas

Nepali test was a combination of four content areas: 1) *Reading*, 2) *Writing*, 3) *Grammar*, and 4) *Vocabulary*. The maximum marks of Reading and Writing were proportional to the weightage given in the curriculum. Grammar and Vocabulary are taught as "functional skills". To compare the achievement in all the topics, these sub-scores are converted into percentage. Figure 4.1.4 shows the students' achievement in Nepali as a whole and the achievement level in four content areas. It should be noted that the difference between maximum and minimum scores is 100% as the scores range from 0 to 100%.

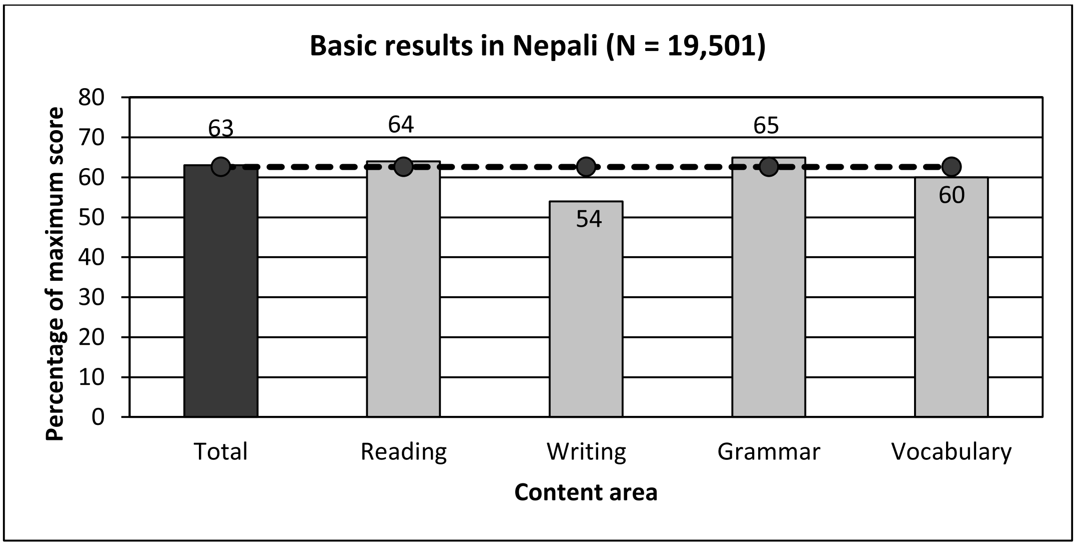

Figure 4.1.4 Basic results in various content areas

The percentage of achievement score shows that the national average of Nepali is 63%. Of the different content areas, students are somehow weaker than the average in Writing (54%) and Vocabulary (60%). They perform better than average in Reading (64%) and Grammar (65%).

Because of the difference in the average level between the community and institutional schools, it is interesting to know whether there is proportional difference in the content areas between the students. Table 4.1.1 and figure 4.1.5 further illustrate the differences.

Table 4.1.1 Achievement in various content areas by school type

Content area	Community schools (N =14,712)			Institutional schools (N = 4,789)		
	Mean	SD	CV	Mean	SD	CV
Reading	59.3	25.2	42.4	79.6	16.8	21.1
Writing	47.3	25.8	54.6	75.0	18.0	24.1
Grammar	59.5	28.7	48.2	81.1	17.5	21.6
Vocabulary	55.1	34.5	62.7	75.8	25.8	34.1
Total[1]	57.0	23.8	41.7	79.9	14.8	18.5

1)Note that the total score is not the mean of the content areas because it has been equated independently from the content areas.

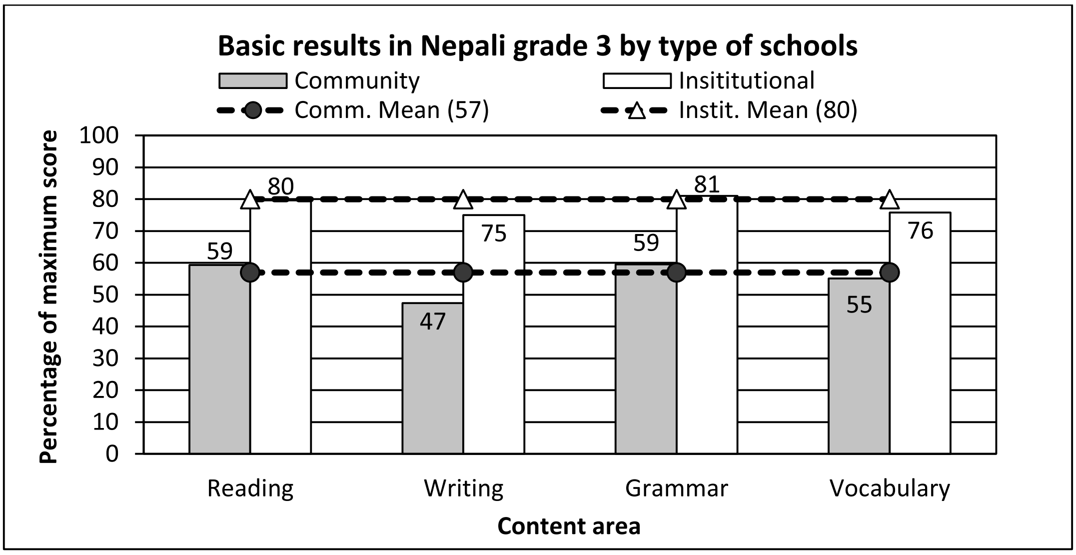

Figure 4.1.5 Basic results in various content areas in Nepali by school type

The differences between the content areas are wider in community schools (47-59%) compared to institutional schools (75-81%). While the difference between Writing (47%) and Grammar (59%) is 12 percent in community schools, it is only 6 percent in institutional schools. Partly this can be explained by ceiling effect in the institutional schools; the test might have been too easy for the students in the institutional schools. Hence, the best students were not able to show how high they could score.

The dataset informs that learning outcomes are the weakest in the content areas of Writing and Vocabulary, and the highest in Grammar and Reading. The differences between the content areas are wider in the community schools than in the institutional schools. This can be caused by the ceiling effect.

Achievement in various levels of Cognitive Domain

The Nepali test as a whole was constructed based on Bloom's taxonomy of hierarchical cognitive levels (Bloom *et al.,* 1956; Metfesser, Michael & Kirsner, 1969), that is, *knowledge*, *comprehension*, *application* and *higher ability (reasoning/problem solving).* The achievement of the students on the hierarchical levels in total sample is shown in table 4.1.2 and further illustrated in figure 4.1.6.

Table 4.1.2 Achievement in various levels of cognitive domain

Levels of cognitive domain	Mean	SD	Min	Max
Knowledge	72.1	27.3	0	100
Comprehension	65.3	24.0	0	100
Application	56.2	29.7	0	100
Higher Ability	37.4	22.2	0	100

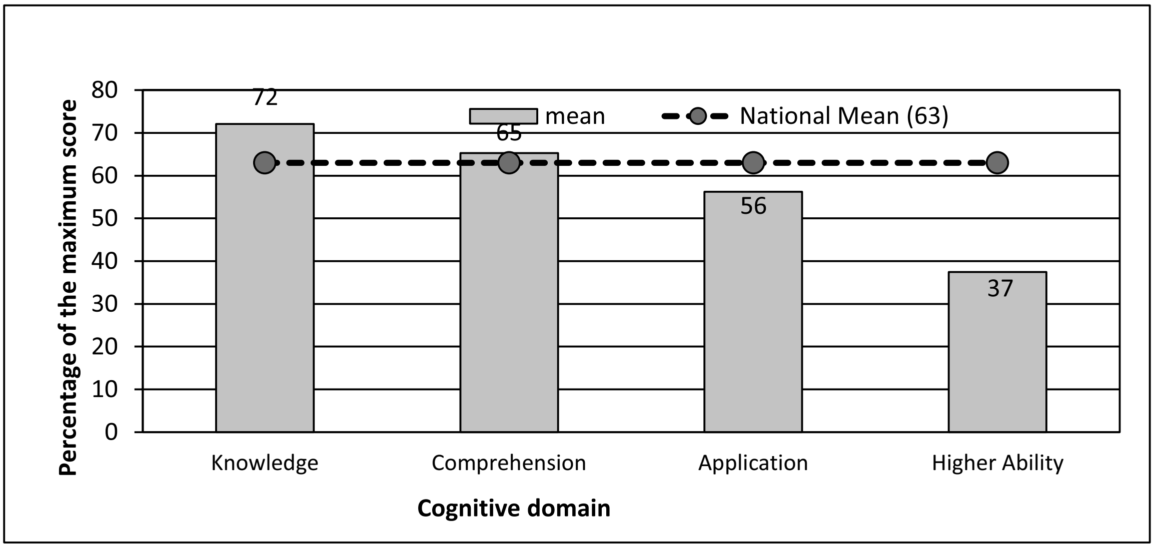

Figure 4.1.6 Achievement in various levels of cognitive domain

Remarkably a large number of students (12.5%) were able to solve only 15 percent or less of the practical problems, that is, the application type of items. Round about 13% of the students could solve just less than 10 percent of the tasks requiring higher cognitive abilities (analysis and evaluation) and 3% of the students did not solve any of these tasks.

The main trend is obvious that the students are much better in the recall type of questions than in the tasks requiring higher skills. Statistically, the tendency is seen in the effect size: though the differences are remarkable in any case, the difference is notably smaller in the area of Knowledge ($d = 0.62$) than in Application ($d = 1.15$) and Higher Ability ($d = 1.06$).

The dataset shows that the students' ability to solve complex problems is quite low; only 37% of the maximum scores on tasks requiring higher ability were reached. Students are much better in the recalling type of questions (72%). About 3% students were unable to solve any of the tasks requiring higher cognitive ability.

Types of items and achievement

There were basically two types of questions in the test: objective and subjective items. 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 was needed to get a correct answer. There were also 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 made in most of the hierarchical levels (knowledge, comprehension, application and higher ability) and a wide range of difficulty levels though the subjective items tended to be more demanding because of the higher demand of cognitive level. Tables 4.1.3 and 4.1.4 comprise the basic statistics of the achievement levels in various types of items.

Table 4.1.3 Score in different types of items

Type of items	Mean	SD	Minimum	Maximum
Objective	70.5	23.5	0	100
Subjective	51.0	26.5	0	100

Table 4.1.4 Item type and average achievement score by school type

Community schools (N = 14,712)				Institutional schools (N = 4,789)		
Type of items	**Mean**	**SD**	**CV**	**Mean**	**SD**	**CV**
Objective	65.7	24.0	36.5	85.2	13.8	16.1
Subjective	44.4	25.3	56.9	71.4	18.4	25.7
Total	**57.0**	**23.8**	**41.7**	**79.9**	**14.8**	**18.5**

It is obvious that the subjective test items – usually those with more demanding requirements for the correct answer – are solved much less (51%) than the objective items (70%). Most of the objective items were of knowledge, comprehension and application type, whereas subjective items were of application and higher ability type. Though the differences between the community (22%) and institutional schools (14%) are, in any case, wide, the effect size is notably higher when it comes to solve the subjective type of items (d = 1.14) compared to the objective type of items (d = 0.89).

The dataset suggests that the students are performing well in recognizing the correct answer and in recalling simple facts from the texts, fundamental thinking, the basic interpretation of paragraph, table and chart, and a few steps of logical thinking. They are much weaker in producing fluent texts or letters, or preparing synthesis and abstracts from a text. In many cases, the students tended to do the open ended task (like free writing, problem solving and analysis) but the skills were not high enough for achieving the highest marks.

Comparison of NASA 2012 with previous datasets

The National Assessments carried out in various years aim to assess the change in the achievement level and the progress over a period of the years. The datasets of previous grade 3 Nepali assessment are, however, somehow sparse (see, BPEP, 1995). There are two issues on the previous datasets regarding the comparison with the present dataset. First, the National Assessment of grade 3 students carried out by the Basic and Primary Education Project (BPEP, 1995, p. 14) shows that the national average of the students was 35 percent. The National Assessment carried out in 2012 by the ERO shows that the national average of Nepali in grade 3 is 63 percent. These figures are coming from Classical Test Theory, but they are not comparable with each other due to the lack of a proper linking procedure. The differences between the scores can easily be explained by difficulty levels of the test. Second, the previous datasets of grade 3 are not available; and, hence, any IRT modelling based procedures for comparison could not be made.

The comparison cannot be made in absolute sense as proportional comparisons can be made with caution on the basis of the previous results. The proportional differences are presented in table 4.1.5.

Table 4.1.5 Comparison of score distribution of 1995 and 2012 in Nepali subject

Level	1995 (BPEP, 1995)	NASA 2012
	% of cases	% of cases
ABOVE the mean	45.7	56.0
AT the mean	5.4	1.7
BELOW the mean	48.9	42.3

A rough comparison of score distributions of 1995 and 2012 datasets shows that in both years the distribution is not Normal. However, in 1995, there is longer tail in the higher part of the scale, whereas in 2012, there is a longer tail in the lower part of the scale. This means that in 1995 there were more cases at the lower part of the scale and in 2012 there were more cases in the higher part of the scale. Expressed in a schematic way, the difference between the years is shown in the following set of graphs:

Figure 4.1.7 Comparison of score distribution of 1995 and 2012

There is no need to make too strong interpretation of the figures; the shapes of the distribution may just tell that the test in 1995 was too difficult for the population and, in 2012, it was too easy. However, it informs that, within 17 years, a population of well performing pupils has grown.

Table 4.1.6 Comparison of scores of 1995 and 2012 in Nepali by gender

Indicators	1995 (BPEP, 1995)		NASA 2012	
	Boys	**Girls**	**Boys**	**Girls**
Mean	12.7[1]	12.5[1]	62.7[2]	64.2[2]
SD	6,24	6.45	24.0	23.5
CV	48.9	51.2	38.2	36.7
N	1,020	824	8,586	8,861
T	0.575		-4.168	
Sig	n.s		<0.001	
Cohen's *d*	0.00		0.06	

1) Raw scores, maximum 36 2)Percentages of marks

Compared to the 1995 dataset, the difference of achievement between the boys and girls has changed slightly; the non-existent difference between gender in1995 have turned to a slight female dominance in 2012; the difference between boys and girls is 1.5 percent. Though the differences are significant, both the effect size and the real difference is still small. The Coefficient of Variation (CV) shows that, over the 18 years, the distributions of boys and girls have been narrower (CV has lowered from round 50 to round 38). This is a positive sign from the national equality point of view.

Table 4.1.7 Comparison of scores of 1995 and 2012 in Nepali by Ecological zones

Indicators	1995 (BPEP, 1995)			NASA 2012		
	Mountain	**Hill**	**Tarai**	**Mountain**	**Hill[1]**	**Tarai**
Mean	14.5[2]	14.9[2]	10.7[2]	61.2[3]	59.9[3]	56.5[3]
SD	4.7[4]	7.1[4]	5.9[4]	22.0	24.1	23.8
CV	32.6	47.9	55,4	36.0	40.2	42.1
N	66	685	1,093	1,917	8,860	5,349

1) Students from Kathmandu Valley are excluded 2) Raw scores, maximum 36, weighted mean 3) Percentages of marks 4) Weighted mean

Compared to the 1995 dataset, the differences between the Ecological zones is found to have been widened slightly. The students in the Mountain zone have raised their achievement in comparison to the Hill zone (no difference in 1995 and somehow higher result in 2012). One may note, though, that there was only one district from Mountain zone in 1995. In 1995, the students in Tarai were about 10 percent behind the students from other zones. Tarai zone has become more homogenous during the years (CV has lowered from 55 to 42) in 2012. This is also a positive sign from the national equality point of view.

Table 4.1.8 Comparison of scores of 1995 and 2012 in Nepali by Development regions

Indicators	1995 (BPEP, 1995)					NASA 2012[1]				
	Eastern	Central	Western	Mid-Western	Far-Western	Eastern	Central	Western	Mid-Western	Far-Western
Mean	8.7[2]	13.7[2]	19.1[2]	12.7[2]	11,5[2]	53,7[3]	59,3[3]	67,0[3]	54,4[3]	59,2[3]
SD	5.6[4]	6.6[4]	8.2[4]	6.2[4]	5.9[4]	23.9	23.6	21.9	23.5	23.8
CV	64.8	48.5	43.2	49.1	51.7	44.6	39.8	32.7	43.2	40.2
N	561	295	279	364	345	3,043	4,568	3,023	2,328	3,164

1) Students from Kathmandu Valley are excluded 2) Raw scores, maximum 36, weighted mean 3) Percentages of the maximum marks, maximum 100 4) Weighted mean

When comparing the Development regions, two things should be kept in mind: First, the Central and Western developmental regions were represented by only one district from each region in 1995, and second, the Valley was not included in the 1995 sampling. However, to make the datasets comparable, the raw scores have been changed to the percentage of total score (24.2, 38.0, 53.03, 35.3, and 31.9), and the Western region (with the highest score) has been taken as a reference point for comparison. It is notable that, according to the 1995 dataset, the students from Eastern region achieved 29 percent lower than the students from the Western region; in 2012 the difference is only half of that, with 14 percent. The effect size has decreased from d = 1.46 to d = 0.52 showing that the difference has reduced remarkably. The same is found to have happened with Far-Western region in comparison with Western region: in 1995 the difference was 22 percent but in 2012 it is about one third of that; at 8 percent points. The effect size has decreased from d = 1.12 to d = 0.36 showing that the difference has reduced from high to moderate. These signs are positive from equality point of view.

Compared to the 1995 dataset by the ethnicity of the students, it is evident that the context has changed significantly during the 17 years. In the 1995 dataset, the Brahmans and Chhetris (implying mainly Nepali speaking) were at the higher level in achievement than the students from Urdu and Gurung/Magar/Tamang language groups. In 2012, the latter groups have reached much nearer to the Nepali speakers. Students from Tharu, Newari, and “Other” communities are found to have lowered down their position remarkably. Table 3.3.8 demonstrates the information.

Table 4.1.9 Situation in 2012 in comparison to the 1995 datasets

Selected background variables					
	Gender	**Ecological zone**	**Development region**	**School location**	**School type**
Main finding	Differences are at the same level; girls still over-perform boys slightly	Differences have increased moderately; students in the Mountain zone scored higher, while in Tarai scored lower.	Students in the Far-Western region scored higher while in the Eastern region scored lower.	Remarkable rise in performance within the Urban schools.	No remarkable change in difference between the community and institutional schools.

Compared to the 1995 results, the dataset in 2012 informs that there are more high performing students; girls have turned to perform better than boys, the students in the Mountain and Tarai zones scored higher and the difference between Hill and Tarai have reduced. In both years the students from the Eastern region performed the lowest; otherwise the difference between the regions would have been smaller. Nepali speaking groups have raised their position compared to the other language/ethnic groups; students from Tharu, Newari, and "Other" communities are found to have remarkably lowered down their position.

Comparison with international assessment results

The NASA 2012 was made comparable with the international PIRLS reading assessment. Four of the released PIRLS items were used as linking items to the international item bank. 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; when a student's ability level in NASA 2012 is zero, it corresponds to the average level of the international students of Grade 4 in their own native language. Because the text and the related items were targeted to the native speakers, the text and items were translated into Nepali.

Figure 4.1.8 shows the comparison of the students' achievement with the international standard. In the figure, the x-axis shows the content areas of Nepali and y-axis shows the ability shown by the students. The middle horizontal line of Theta θ = 0.00 indicates the international average. When the ability is below the average, the bars go down whereas when the ability was above the international average, the bars would have gone upwards.

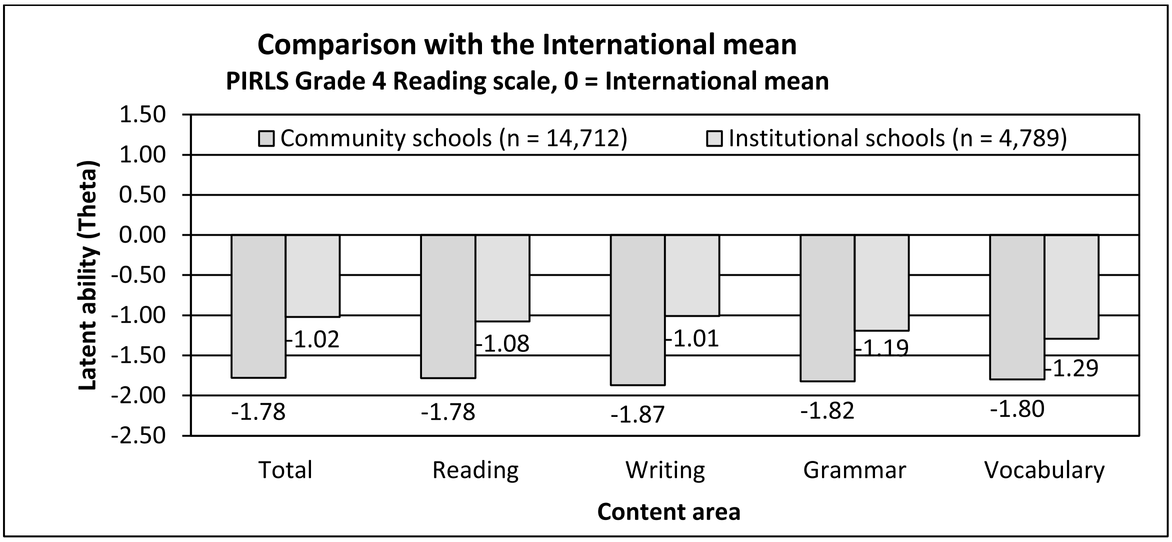

Figure 4.1.8 Student achievement in the international PIRLS reading scale

Figure 4.1.8 shows that the average ability shown by Nepali students in Reading is generally much below the international average. This indicates that the students in Nepal score remarkably lower than their international peers; the achievement level of an average grade 3 student in the community schools ($\theta = -1.78$) is very low compared to an average international student of grade 4. The achievement level of an average student in the private schools ($\theta = -1.02$) is higher than his/her peer in the community school but still far behind the international average. It is good to remember two things. First, *all the linking items came from the content area of Reading* and hence there actually is no real equating in the other areas. Especially incomparable are the grammar and vocabulary because in PIRLS these areas are not measured at all. However, they are modeled on the basis of proficiency in the reading test. Second, the difficulty level of the text (complex text suitable for 4th graders) was not the best suiting the 3rd graders. Hence, the lower proficiency is expected but the degree is remarkable.

The dataset suggests that, from the international comparison point of view, the average reading and writing proficiency in Nepal is much lower (-1.78 and -1.87) than the international average in PIRLS standard (0.00). The lower achievement level was expected because the linking items were suitable for more matured students of grade 4.

Comparison with the objective standards – CEFR levels

An international comparison was done on the basis of the standards of Common European Framework in Reference for Languages (CEFR). CEFR classification with the standard setting procedure (3TTW) (Metsämuuronen, 2013; see also ERO, 2013) was applied to assess the criterion-based proficiency in the Nepali language. The main results of Nepali Reading and Writing proficiency levels in Nepal are depicted in figures 4.1.9 and 4.1.10 and tables 4.1.10 and 4.1.11.

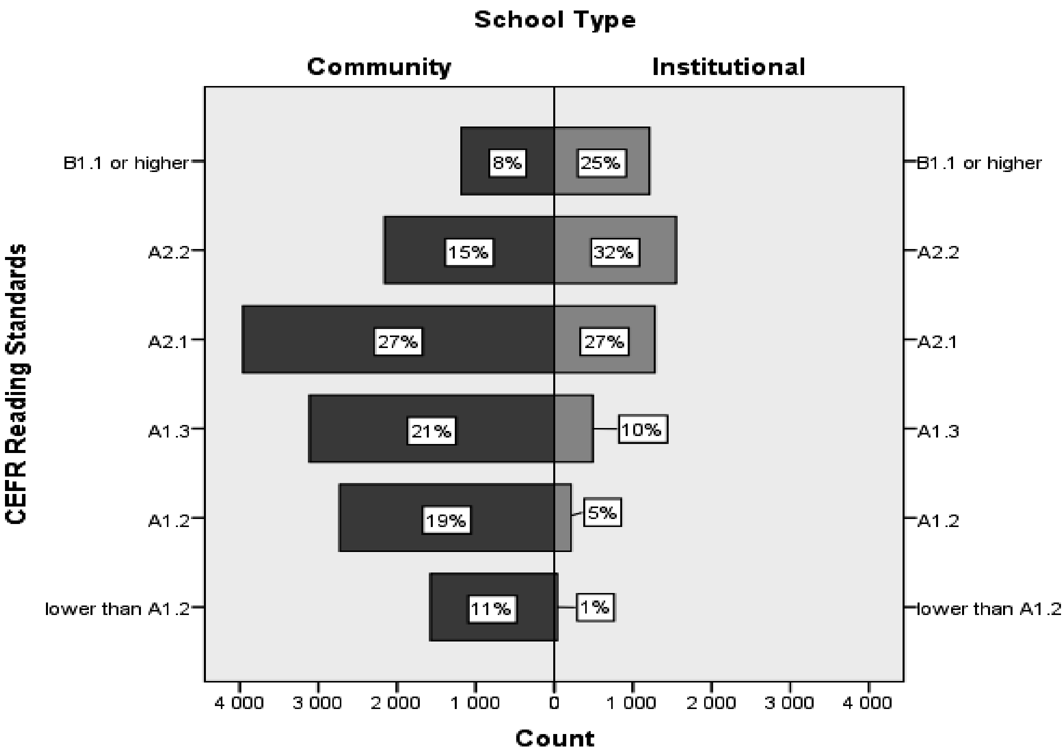

Figure 4.1.9 Student achievement in the international CEFR Reading standard

In the community schools, the most typical grade 3 reader of Nepali is at the level of A2.1. This means that the typical student "[can] *understand simple texts containing the most common vocabulary (personal letters, brief news items, everyday user instructions);can pick up the main points and some details of a few paragraphs of text; locate and compare specific information; and can draw very simple inferences based on context*. (FNBE, 2004: 278–295; see also ERO, 2013: table 4.6.4). The most typical grade 3 reader of Nepali is at the level of A2.2. This means that the typical student "[can] *understand the main points, some details of messages consisting of a few paragraphs in fairly demanding everyday contexts (advertisements, letters, menus, timetables) and factual texts (user instructions, brief news items); acquire easily predictable new information about familiar topics from a few paragraphs of clearly structured text* and *infer meanings of unfamiliar words based on their form and context*." (FNBE, 2004, 278–295; see also ERO, 2013, table 4.6.4). It is notable that the distribution is somehow the same as in grade 5 (see figure 4.2.7); in grade 3 dataset, though, the number of students is higher at the categories below A2.1.

In figure 4.1.9, it is seen that 55% of the 3rd graders in community schools are at level A1.3 or lower, which means that they can read only very short notices and postcard type of texts and only pick some facts out of the text. This kind of elementary reading skill was tested, for example, in the reading task called "Calendar". In this indicating task of level A1.3, a calendar of the month of Kartik 2068 BS was given. A true/false type of question states: "*The month starts on Tuesday*". The result showed that 8.9% of the students did not answer the question, and 19.1% students were not able to give the correct answer.

Based on the dataset, 82% of the whole student population has reached at least level A2.1 in Nepali Reading and hence they can understand (only) simple texts containing the most common vocabulary (Table 4.1.10).

Table 4.1.10 Percentages of 3rd graders reaching the specific CEFR levels in Reading

CEFR level	Brief description of ability	% reaching the level	% at each level
B1.1 or higher	Can read **a few pages of a wide variety of texts about *familiar* topics** (tables, calendars, course programmes, cookery books)	12.3	12.3
A2.2	Can understand the **main points and some details of** messages consisting of a few paragraphs in **fairly demanding everyday contexts** (advertisements, letters, menus, timetables) and **factual texts** (user instructions, brief news items).	31.3	19.0
A2.1	Can understand **simple texts containing the most common vocabulary**.	58.2	26.9
A1.3	Can understand **very short messages dealing with everyday life** and routine events or giving simple instructions.	76.6	18.5
A1.2	Can understand names, signs and other very short and **simple texts related to immediate needs;** **Can** identify specific **information in simple text provided one can read it as required.**	91.7	15.1
< A1.2			8.3

When it comes to Nepali Writing proficiency in the community schools, the typical grade 3 students of Nepali vary ranging from A1.2 (16% of the students) to A2.2 (16%) though the mode is at A2.1 (23%) (figure 4.1.10). The typical student at the CEFR level A2.1"[can] *manage in the most routine everyday situations in writing.* [(S) he can] *write brief, simple messages (personal letters, notes), which are related to everyday needs, and simple, enumerated descriptions of very familiar topics (real or imaginary people, events, personal or family plans).*" (FNBE, 2004, 278–295; see also ERO, 2013, table 4.6.4) In the institutional schools, the most typical 3rd grader writer of Nepali stands at the level higher than B1.1 (42% of students).[31] This means that the typical student can, at least, "*write an intelligible text about familiar, factual or imaginary topics of personal interest, also conveying some detailed everyday information; write a clearly formulated cohesive text by connecting isolated phrases to create longer sequences (letters, descriptions, stories, telephone messages); and can effectively communicate familiar information in the most common forms of written communication.*" (FNBE, 2004, 278–295; see also ERO, 2013, table 4.6.4) It is notable that the distribution is somehow the same as in grade 5 (see figure 4.2.8); in grade 3 dataset, though, the number of students is higher at the categories below A2.1. This means though that the difference between the grade 3 and 5 students is not big.

[31] 42% of the student at the level B1.1 or higher seems quite a high number for the grade 3. Compared to the grade 5 dataset (see section 3.4) and grade 8 (NASA 2011) the result seems credible though. In the institutional schools in grade 5, there are 60% students at the level B1.1 or higher. In NASA 2011, at grade 8, there were 74% students at the level B1.1 or higher in the institutional schools. The tendency (42 – 60 – 74) seems possible and credible.

The typical writer is somehow at the same level as in Reading, even though usually the receptive skills (as Reading is) are at the higher level as the productive skills (as Writing is). Also, the aimed level of the receptive skills may be set lower than for productive skills (see FNBE, 2004, 140, see also in ERO, 2013, table 4.6.3). In the institutional schools, it is seen that the typical writer in grade 3 is as high as B1.1 or higher. This means that, compared to the Reading proficiencies, there are quite many good 3rd grader writers in the institutional schools.

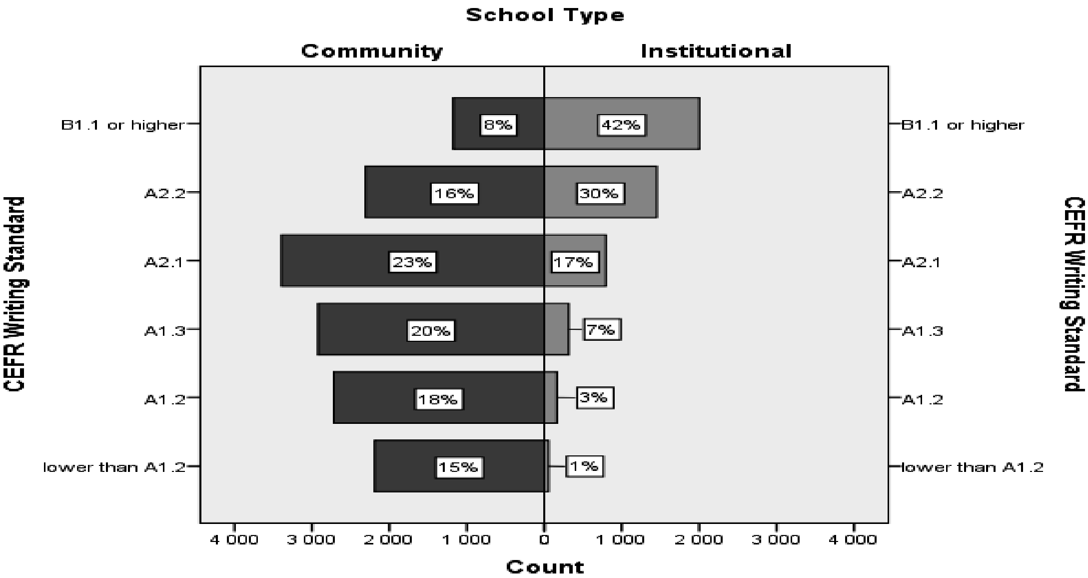

Figure 4.1.10 Student achievement in the international CEFR writing standard

On the basis of the grade 3 Nepali dataset, 57% of the students in the whole population have reached at least the level A2.1 in Nepali writing, and hence they can write **brief, simple messages** (personal letters, notes), which are **related to everyday needs**.

Table 4.1.11 Percentages of 3rd graders reaching the specific CEFR levels in Writing

CEFR	Brief description of ability	% reaching the level	% at each level
B1.1 or higher	Can write a **clearly formulated cohesive text** by connecting isolated phrases to create longer sequences (letters, descriptions, stories, telephone messages).	16.3	16.3
A2.2	Can write a **very short, simple description of events, past actions and personal experiences** or everyday things in his/her living environment (brief letters, notes, applications, telephone messages).	35.6	19.3
A2.1	Can write **brief, simple messages** (personal letters, notes), which are **related to everyday needs**.	57.1	21.5
A1.3	Can manage to write in the **most familiar, easily predictable situations** related to **everyday needs and experiences**. Can write **simple messages** (simple postcards, personal details, simple dictation).	73.7	16.6
A1.2	Can **communicate immediate needs inbrief sentences**. Can write a **few sentences** and phrases about him/herself and his/her immediate circle (such as answers to questions or notes).	88.5	14.8
< A1.2			11.5

Two points of the language proficiency in Nepali grade 3 are worth of highlighting. First, usually in language learning, the receptive skills (Reading and Listening) stay at the lower level than the productive skills (Writing and Speaking). Also, the aimed level of the receptive skills may be set lower than for productive skills (see table 3.3.9c; FNBE, 2004, 130, 135; see also ERO, 2013, table 4.6.3). In Nepali grade 3 dataset, however, Writing skills is seen at the *same* level as of Reading. One explanation for this is that, in general, the current curriculum prefers writing skill over the Reading skills. Second, there is no objective criterion regarded what should be the language proficiency level at the end of grade 3 or grade 5. One aid for evaluating the proficiency would be the one used in the Finnish core curriculum (FNEB, 2004) and the descriptions of good performance as the measurement stick. In the Finnish system, the criterion is given for the end of grade six and hence it is not fully comparable in Nepalese context and, especially, it is not the best in evaluating grade 3 proficiency. The criterion is given for the foreign languages instead of the mother language. Anyway, some clues of the required language proficiency levels can be obtained from table 4.1.12. The closest fit to the Nepali language comes the criterion for native-level Finnish or, in grade 3, may be from the criteria for Finnish as the first foreign language.

Table 4.1.12 Description of good performance in Languages at the end the sixth grade in the Finnish system

Language and level	Reading	Writing
Finnish as Native-level (bilingual) (FNBE 2004, 135)	B1.2 Fluent basic language proficiency	B1.1 Functional basic language proficiency
Finnish as the first foreignlanguage (FNBE 2004, 130)	A2.1 Initial stage of basic language proficiency	A1.3 Functional elementary language proficiency

Compared to table 4.1.13, the typical Nepali reader at grade 3 in the community schools is at the same level (A2.1) as the "good" reader in the Finnish system for the students studying Finnish as the first *foreign language* at grade 6. The typical Nepali writer (A2.1) is at the higher level than the "good" writer in the Finnish system for the students studying Finnish as the first foreign language at grade 6. In the institutional schools, the typical Nepali reader (A2.2) is at the higher level than the "good" reader in the Finnish system for the students studying Finnish as the first *foreign language* at grade 6. However, the typical Nepali writer at grade 3 (B1.1) looks to have reached almost the same level as required for the *bilingual native speaker at grade 6*. This is a very high level. More comparative studies and standard settings may be needed to confirm the results.

Dataset indicates that, in the community schools, the typical grade 3 reader of Nepali is at the CEFR level of A2.1. This means that the typical student can understand simple texts containing the most common vocabulary, the main points and some details of a few paragraphs of text. In the institutional schools, the most typical grade 3 reader is at the level B1.1, that is*, (s)he can understand the main points, some details of messages consisting of a few paragraphs in fairly demanding everyday contexts, factual texts and* *acquire easily predictable new information about familiar topics from a few paragraphs of clearly structured text*.

4.1.2 Results Based on Diversity Factors

Diversity is a relative and contextual term. In the context of Nepal, geographical/ecological, language, gender, religious, ethnic, cultural, disability and economic factors are considered as diversity (see ERO, 2013). Although the NASA 2012 background information questionnaire included six of the above diversities - three of these - the cultural, religious and disability backgrounds of the students, were not asked. Instead, three other diversities are handled in this section: district-wise, school type-wise (community/institutional), and school location-wise (rural/urban). These factors can be taken as equality factors as all children regardless of their sex, language, birth place, or family background should have equal opportunities to reach the same educational goals.

Districts and student achievement

Out of 75 districts, 25 were randomly selected to represent the Ecological zones and developmental regions and ultimately to represent the country as a whole. Additionally, 3 districts of the Kathmandu Valley (Kathmandu, Bakhtapur, and Lalitpur) were selected because they present a unique stratum in the country. It is good to keep in mind that there may be lower or better performing districts within those *not* selected in the sample. The district-wise differences are depicted in table 4.1.13. The table shows the district-wise achievement in ascending order according to the achievement. The mean represents the average achievement percentage of the particular district.

Of the randomly selected districts in the sample, the student performance was very low in Saptari (48%) and Khotang (51%) from the Eastern region; Mahottari (52%) from the Central region; and Bardiya (52%) and Salyan (52%) from the Mid-Western region. Except Kaski district (77%), the outperforming three districts come from the Central region, specifically from the Kathmandu Valley: Kathmandu (81%), Bhaktapur (80%), and Lalitpur (78%). Comparison may be unfair because 71% of the schools in the Kathmandu Valley are private ones, while in the other districts in the sample, on average, only 10% were private ones. Out of the eight lowest performing districts, two have no private school and four have only one for each. From this perspective, interesting districts are those where the number of institutional schools is low but the results are higher than the national average. Some examples of these districts are Solukhumbu (69%), Baitadi (68%), Manang (64%) and Parsa (63%).

Table 4.1.13 Average achievement score by sample districts

Districts	N	Mean	SD	CV	Districts	N	Mean	SD	CV
Kathmandu	2110	81.0	15.8	19.6	Kailali	1120	59.6	22.8	38.3
Bhaktapur	544	80.0	13.8	17.2	Darchula	535	58.7	23.4	39.8
Lalitpur	721	78.1	17.3	22.2	Makwanpur	979	57.1	24.6	43.1
Kaski	974	77.5	15.9	20.6	Rolpa	699	56.7	22.5	39.7
Solukhumbu	375	68.5	19.3	28.1	Kapilbastu	817	55.7	23.3	41.8
Baglung	790	68.3	19.5	28.5	Dhankuta	521	54.8	24.6	45.0
Baitadi	784	67.6	22.2	32.8	Jumla	190	54.0	18.4	34.0
Manang	16	63.7	25.0	39.3	Udayapur	768	52.8	23.2	43.9
Parsa	638	62.9	19.7	31.3	Salyan	677	52.4	23.9	45.6
Chitwan	842	62.7	22.4	35.6	Bardiya	548	51.6	25.7	49.8
Myagdi	426	62.1	23.3	37.5	Mahottari	664	51.5	25.2	48.9
Humla	214	61.2	21.7	35.5	Khotang	659	51.4	23.6	45.9
Dolakha	587	61.0	22.2	36.5	Achham	725	50.0	24.0	48.0
Sindhuli	858	60.5	24.2	40.1	Saptari	720	48.1	23.5	48.9
Total						**19,501**	**62.6**	**24.02**	**38.4**

The difference in achievement due to the district is statistically significant ($p < 0.001$). The variation explained in achievement due to the district is $\eta^2 = 0.189$, that is, the district explains 19% of the variation in the data which is a very high percentage. Effect size is $f = 0.48$ indicating that the difference between the lowest performing district (48%) and highest performing district (81%) is remarkably high.

The dataset suggests that there is a wide difference in achievement between the districts. The results are bound to 28 sample districts. The results are high in the districts where the proportion of institutional schools is high. However, there are some districts having achievement of community schools above the national average.

Ecological zone and student achievement

As the sample schools represent all Ecological zones of Nepal, achievement levels and their variations have been analysed taking them into account. The achievement level varies significantly according to the Ecological belts of schools. The variation of achievement in terms of the Ecological zones is presented in table 4.1.14 and figure 4.1.11.

Table 4.1.14 Achievement in Ecological zones

Ecological zone	Community schools				Institutional schools			
	N	Mean	SD	CV	N	Mean	SD	CV
Mountain	1770	60.7	22.3	36.8	147	67.6	16.7	24.7
Hill	7398	56.0	23.8	42.6	1462	79.8	12.7	15.9
Tarai	4491	53.4	23.4	43.9	858	72.7	18.2	25.0
Valley	1053	73.4	19.2	26.2	2322	83.3	13.0	15.7
Total	**14,712**	**57.0**	**23.8**	**41.7**	**4,789**	**80.0**	**14.8**	**18.5**

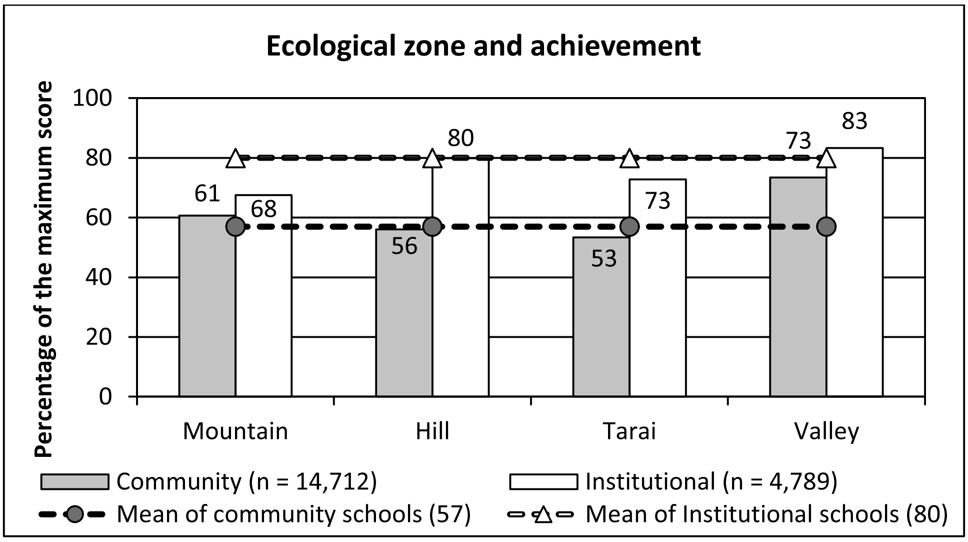

Figure 4.1.11 Differences in achievement across Ecological zones by school type

The data shows that, on average, the students from the Valley have outperformed those from the other Ecological zones. The difference is wider within the community schools (53 -73%) than within the institutional schools (68–83%). In the community schools, the students from Tarai zone performed the lowest (53%) though the difference is not notable compared with the Hill zone (56%). In the institutional schools, the students in Mountain zone performed the poorest (68%) though not much lower than Tarai (73%). One can also note the exceptionally low value for the Coefficient on Variation in the Valley, which, in the community schools, is almost half of that in the other zones. The obvious reason for this is the systematically high score in the Valley compared to the other areas.

While analyzing the achievement in terms of Tukey's *post hoc* test, the achievement in the zones differs significantly ($p < 0.001$) from each other within both the school types. Ecological zone explains 4% variance in the community schools ($\eta^2 = 0.045$) and 9% in the institutional schools ($\eta^2 = 0.089$). As a comparison, one can notice that the district explains more than 19% of the variation. The effect size is $f = 0.21$ within the community schools and $f = 0.31$ within the institutional schools showing moderate difference between the highest and lowest performing Ecological zones. The effect sizes are smaller if the Valley is taken out from the analysis ($f = 0.09$ and $f = 0.27$ respectively). This means that, in the community schools, the real differences are not remarkable between the Ecological zones but the Valley differs radically from the other areas. From equality point of view, this can be taken as a good sign.

The dataset informs that there is a moderate difference in the performance of students among the four Ecological zones applicable in both community and institutional schools. Students in the Kathmandu Valley outperform the other students. The achievement of community school is the lowest in Tarai zone.

Development region and student achievement

The student achievement varies according to the Development regions which are divided into Eastern, Central, Western, Mid-Western, and Far-Western. Additionally, the Kathmandu Valley is taken as the 6th Development region though administratively it falls under the Central Developmental region. The mean achievements in the Development regions are given in table 4.1.15 and illustrated further in figure 4.1.12.

Table 4.1.15 Achievement in the Development Regions

Development region	**Community schools**				**Institutional schools**			
	N	**Mean**	**SD**	**CV**	**N**	**Mean**	**SD**	**CV**
Eastern	2,847	52.5	23.8	45.4	196	71.1	17.6	24.8
Central	4,047	57.3	23.6	41.2	521	74.4	16.8	22.6
Western	1,934	59.5	22.6	37.9	1,089	80.2	12.5	15.5
Mid-Western	2,100	52.0	22.9	44.1	228	77.1	15.2	19.7
Far-Western	2,731	57.1	23.9	41.9	433	72.6	18.0	24.8
Valley	1,053	73.4	19.2	26.2	2,322	83.3	13.0	15.7
Total	**14,712**	**57.0**	**23.8**	**41.8**	**4,789**	**80.0**	**14.8**	**18.5**

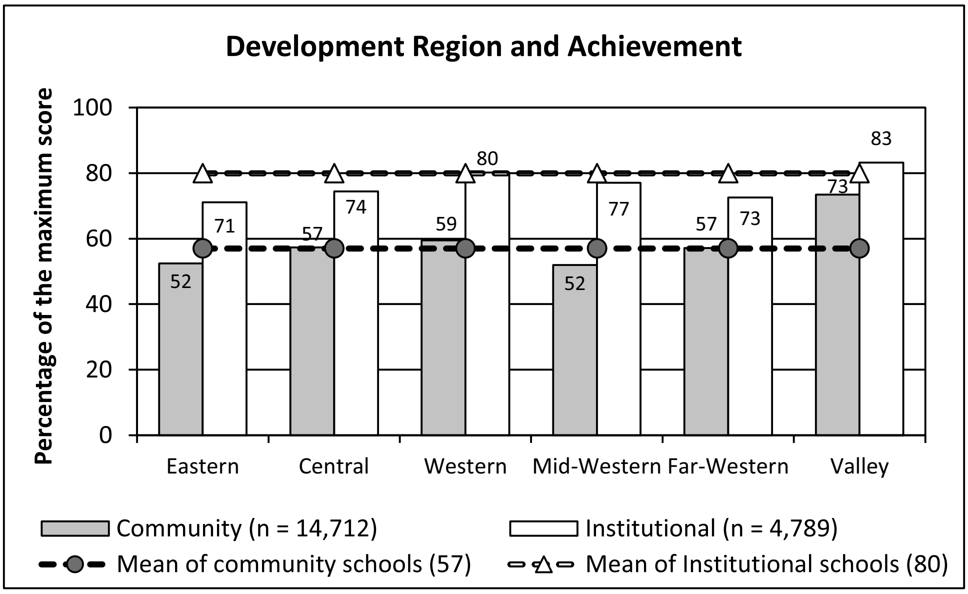

Figure 4.1.12 Comparison of student achievement in Development regions

The highest performance is found within the institutional schools in the Valley (83%) and in the Western Development region (80%). The performance is the lowest in the community schools in the Eastern (52%) and Mid-Western (52%) regions. The difference between the regions is statistically significant in both community and institutional schools ($p < 0.001$). Tukey's *post hoc* test shows that in the community schools, the average achievement level in the Western region is higher than the other regions except for the Valley. The achievement level in the Eastern region is significantly lower than in other regions ($p < 0.001$) except for Mid-western region. It is also lower in Mid-western region than in the other regions.

Development region explains 5% of the variance in the community schools ($\eta^2 = 0.049$) and 8% in the institutional schools ($\eta^2 = 0.079$).[32] This is the same proportion as found with the Ecological zone. It is noted that the district explains more than 19% of variation which means that in the Developmental regions there are lower and higher performing districts. The effect size is $f = 0.23$ in community schools and $f = 0.29$ in institutional schools showing moderate or wide difference between the highest and lowest performing regions. The effect sizes are moderate if the Valley is omitted from the analysis ($f = 0.12$ and $f = 0.23$ respectively).

The dataset indicates that there is inequality of children's opportunities among the Development regions to reach an adequate level in Nepali language. Especially the wide difference between community schools in the Valley and the rest of the country (21 percent difference between the lowest and highest) is a clear sign of inequality of opportunities in learning Nepali. There are also wide differences between the regions in the institutional schools; the difference in student performance in the institutional schools in the Valley and Eastern region is the highest (12 percent).

School type and student achievement

All the schools are categorized into community and institutional (or private). The differences in the Nepali achievement have been handled within the sections above. Here the main differences are condensed in table 4.1.16.

Table 4.1.16 Type of school and the average achievement

Community (N = 14,712)				**Institutional (N = 4,789)**				
Content area	**Mean**	**SD**	**CV**	**Mean**	**SD**	**CV**	**Mean difference**	**Cohen's *d***
Reading	59.3	25.2	42.4	79.6	16.8	21.1	20.3	1.04
Writing	47.3	25.8	54.6	75.0	18.0	24.1	27.7	0.87
Grammar	59.5	28.7	48.2	81.1	17.5	21.6	21.6	1.14
Vocabulary	55.1	34.5	62.7	75.8	25.8	34.1	20.7	0.82
Total	**57.0**	**23.8**	**41.7**	**79.9**	**14.8**	**18.5**	**22.8**	**0.63**

The achievement levels in the community and institutional schools differ from each other remarkably as presented above. The average performance in terms of the scores in the

[32] If the Valley is taken out of the analysis, the values for *Eta* squared would be 0.031 and 0.074 respectively, that is, only 3% and 7% explanation – one third and half of those with the Valley included in the analysis. The role of the Kathmandu Valley students in the whole national mean is remarkable.

private schools is 80% whereas in the community schools it is 57%. Here, the difference is 23 percent, which is remarkable. The difference is statistically significant (p < 0.001) and the effect size is medium (d = 0.63) – showing that the community schools are far below the institutional schools. Difference is the highest in the content area of Grammar (d = 1.14) and lowest in Vocabulary (d = 0.82). Division of the students into the community and institutional schools explains 12% of the student variation in Reading (η^2 = 0.122) and 19% in Writing (η^2 = 0.195). From the figure 4.1.3 it is known that the deviation within the community schools is remarkable ranging from almost 0% to 100%; contrarily, most private schools in the sample show high achievement. This may be explained partly by much higher socio-economic input into students' life in institutional schools and strict selection of the students.

The dataset suggests that, on average, the students in the institutional schools outperformed the students in the community schools. The difference is highest in Writing (28 percent). This variance can be explained partly by much higher socio-economic input into students' life and strict selection of the students in the institutional schools.

School location and student achievement

One of the strata of sampling in NASA 2012 was school location. The schools were divided into rural and urban based on the location. This information was given by the head teacher though some of the head teachers did not inform the school location. The achievement of the students in rural and urban schools is presented in table 4.1.17.

Table 4.1.17 Student achievement on the basis of the location of schools

School Location	Community				Institutional			
	N	Mean	SD	CV	N	Mean	SD	CV
Rural	12,376	56.5	23.9	42.3	1,736	78.5	16	19.9
Urban	1304	63.7	21.1	33.1	2,686	80.4	14	17.8
Mean difference		7.2				1.9		
Cohen's d		0.30				0.13		
Total	**14,712**	**57.0**	**23.8**	**41.7**	**4,789**	**79.9**	**14.8**	**18.5**

In urban community schools, the achievement level of the students (64%) is 7 percent higher than in the rural community schools (56%). The difference is statistically significant (p < 0.001) and the effect size is moderate (d = 0.30). If the community schools from the Valley are omitted, the score of the urban community schools lowers to 58 and the difference (2 percent points) is still statistically significant (p = 0.010), but the effect size is low (d = 0.09). The main difference in the community schools is, hence, caused by the high achievement level of the students in the Valley schools. The division into rural and urban schools explains 1.2% of the student variation within the community schools (η^2 = 0.012), and without the Valley schools it explains only 0.2% variation (η^2 = 0.002). The latter is a good sign from equality point of view. But, if the Valley is excluded there is no difference between rural and urban community schools.

In urban institutional schools, the achievement level of students (80%) is 2 percent higher than that of rural institutional schools (78%). Though the difference is statistically significant ($p < 0.001$), the effect size is small ($d = 0.13$). If institutional schools from the Valley are omitted, the difference remains the same (2 percent) which still is statistically significant ($p = 0.001$) and effect size is also the same ($d = 0.13$). In the institutional schools, the effect of the Kathmandu Valley is not remarkable. The division into rural and urban schools explains only 0.4% of the student variation in the institutional schools ($\eta^2 = 0.004$).

The dataset clearly shows that the students in urban community schools have achieved 7 percent more than the students in the rural areas. Excluding the Valley schools, the difference is only 2 percent. However, from the educational inequality point of view, the difference between them is not a good sign though the real difference is not wide in the community schools. But, in institutional schools, the difference between the rural and urban areas is not so wide.

Language at home and student achievement

In the context of Nepal, student achievement is found to be depending on the language spoken in their homes i.e., the mother tongue of the students. The mother tongue also reflects, in many cases, the ethnical background and hence any ethnic difference can be considered as a possible source for inequality in society.

Based on the entire data, 40.5% of the 3rd graders speak a language other than Nepali as their first language. These “other” languages are quite fragmented; the largest groups in the dataset are Tharu (4.2%), Urdu (4.0%), and Newar (3.2%) speaking groups. After dividing the languages into ten groups excluding Nepali, there were still 15.4% of the students classified into the group “other”. Because the languages are very fragmented and the majority of the students are Nepali speakers, for the purpose of the statistical analysis, all the other languages are first grouped into “Non-Nepali”. The results are presented in tables 4.1.18 and 4.1.19 and further illustrated in figure 4.1.13.

Table 4.1.18 Student achievement on the basis of home language

Language group	Community				Institutional			
	N	Mean	SD	CV	N	Mean	SD	CV
Nepali	7,807	59.1	22.7	38.4	3,803	80.6	14.1	17.5
Non-Nepali	6,905	54.7	24.7	45.2	986	76.9	14.1	18.4
Mean difference		4.4				3.7		
Cohen's *d*		0.19				0.25		

When combining all the minor language groups as “Non-Nepali”, there is a slight difference between the language groups in the community schools (4 percent favoring the Nepali speakers). Though the difference is statistically significant ($p < 0.001$), the effect size is low ($d = 0.18$). The same difference of 4 percent in the institutional schools is moderately high ($p < 0.001$, $d = 0.25$).

On the basis of the original categorization of the minor languages, the issue looks quite much interesting. It is evident that the Tamang and Magar students are at quite much higher level in Nepali than the Nepali speaking students (68% and 66% compared to 59% in the community schools). On the other hand, the students from Gurung speaking have performed much lower than the average. Rai speaking students outperformed the other language groups in institutional schools (average of 88%). Of the larger language groups, the Magar students performed higher than the others (84%). It is somehow interesting that the Newari speaking students perform relatively low in community schools (54%) as well as institutional schools (68%); in 1995 their achievement level was almost the same as that of the Nepali speaking students.

Table 4.1.19 Achievement in the different language groups

Language [1]	**Community**				**Institutional**			
	N	**Mean**	**SD**	**CV**	**N**	**Mean**	**SD**	**CV**
Tamang	365	67.7	17.3	25.6	10	82.2	15.2	18.5
Magar	20	66.1	16.8	25.4	102	84.1	9.3	11.1
Urdu	726	63.6	21.9	34.4	56	81.9	12.5	15.2
Sherpa	10	62.7	14.7	23.4				
Rai	140	61.7	20.4	33.0	3	88.7	6.1	6.9
Nepali	7,807	59.1	22.7	38.4	3,803	80.6	14.1	17.5
Tharu	795	55.3	22.5	40.8	22	82.2	15.2	18.5
Mathili	20	54.9	15.8	28.8	15	84.8	4.1	4.8
Newari	559	53.6	22.0	41.1	70	67.8	18.7	27.6
Gurung	70	42.4	27.1	63.9	13	72.1	19.9	27.6
Other	4,197	51.9	25.8	49.7	693	76.1	17.3	22.7

1) Those language groups in which number of the students was less than 10 are omitted.

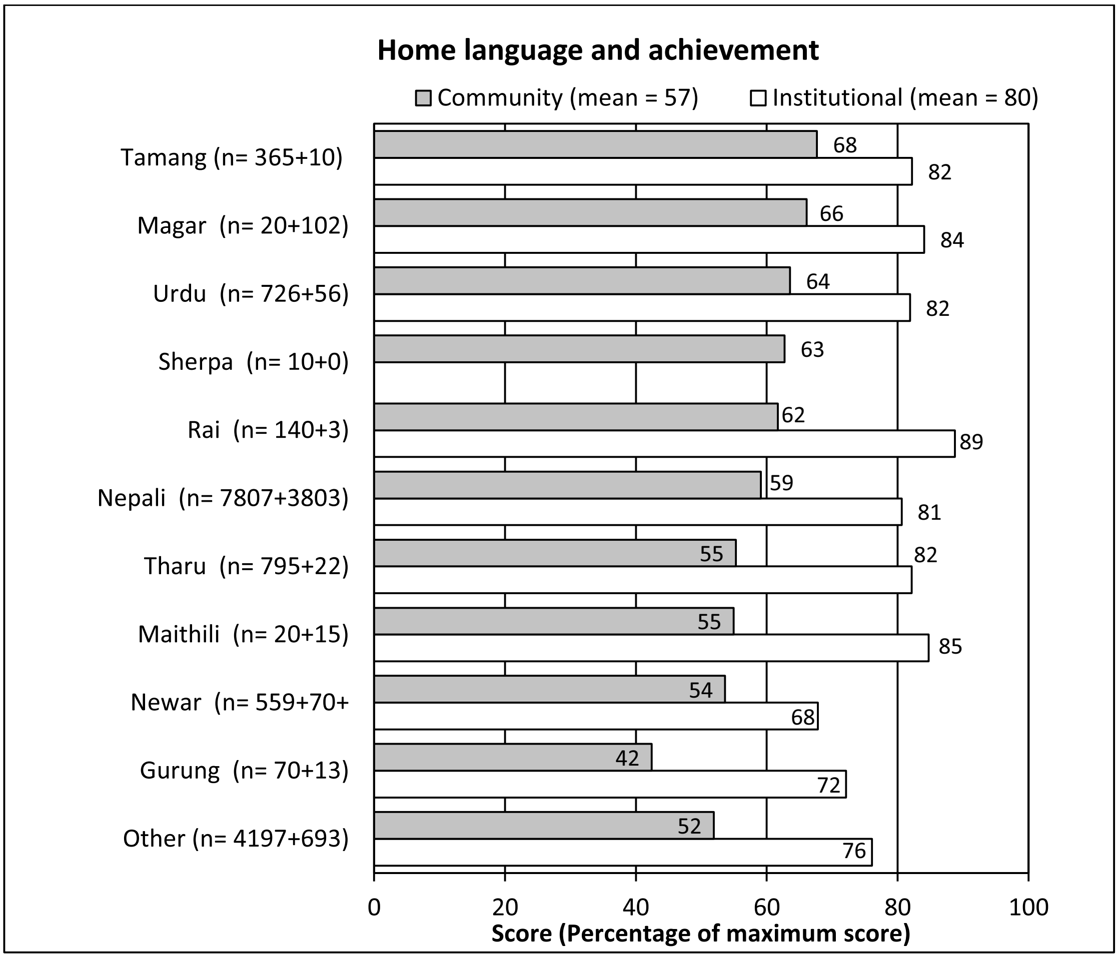

Figure 4.1.13 Relation between language at home and achievement

The differences between the students in the highest and lowest performing language groups are statistically significant ($p < 0.001$) and notable; the effect sizes are moderately high ($f = 0.17$ in the community schools and $f = 0.17$ in the institutional schools). The division into smaller language groups explains about 2% and 3% of the variation in the data ($\eta^2 = 0.029$ in the institutional schools and $\eta^2 = 0.031$ in the community schools). Though the differences are wide between the extreme groups, it would be worth to consider that the number of students is quite small in some of the other language groups, hence, the effect size is moderate. When analyzing only the minority languages, and hence excluding the Nepali speakers and the group "Other", the effect size is high ($f = 0.27$) in the community schools – indicating a remarkable difference between the highest performing minority group (Tamang, 67%) and the lowest performing group (Gurung, 42%).

The dataset shows that there is an educational inequality among the students of various language groups in Nepali subject. In community schools, the students from Magar (81%) background have performed very high in Nepali, while the students from Newari (55%), Tharu (56%), and Gurung (47%) speaking students performed lower than the average. The differences between the lowest and highest performing language groups are remarkable.

The dataset shows that there is an educational inequality in the language groups in Nepali subjet. In the community schools, the students from Magar (81%) speaking performed high in Nepali while the students from Newari (55%), Tharu (56%) and Gurung (47%) speaking performed lower than the average. The differences between the lowest and highest performing language groups are remarkable.

Ethnicity and student achievement

The latest household survey (CBS, 2012) shows that the participation of Hill Dalits has increased remarkably in primary level education but their participation in the secondary and higher education is still very low. The results concerning the castes and achievement are presented in table 4.1.20 and illustrated in figure 4.1.14.

Table 4.1.20 Achievement of the students of various ethnic/castes groups

Caste/ethnicity	Community				Institutional			
	N	Mean	SD	CV	N	Mean	SD	CV
Brahman	932	67.4	20.8	30.8	1,094	82.0	13.3	16.2
Chhetri	2,504	61.5	21.8	35.5	1,033	80.5	14.0	17.4
Janjati	3,997	58.3	23.5	40.2	1,348	80.8	14.1	17.5
Madhesi	960	57.7	22.6	39.1	178	78.1	18.1	23.2
Dalit	1,966	57.5	22.8	39.6	186	75.7	17.7	23.4
Others	1,854	54.8	23.4	42.7	548	78.3	15.3	19.5
Total	**12,213**	**59.0**	**23.0**	**39.0**	**4,387**	**80.4**	**14.5**	**18.0**

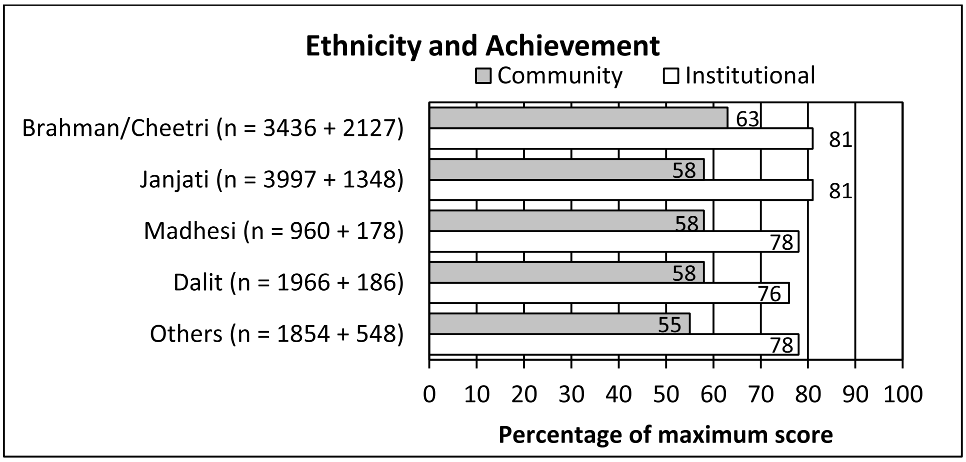

Figure 4.1.14 Relation between ethnicity and achievement

In the community schools, the students from "Other" ethnic/caste group performed the lowest (55%) in Nepali followed by Dalit, Madhesi, and Janjati students (58% each). Dalit students performed below the average in community as well as institutional schools. The overall difference between the groups is statistically significant ($p < 0.001$) though the effect size is small ($f = 0.13$) in the community schools.The division of students according to their ethnic/caste background explains just 1.9% of the student variation ($\eta^2 = 0.019$). In the institutional schools, the effect size is also small ($f = 0.10$); and explains just 1.1% of the student variation ($\eta^2 = 0.011$). From the equality point of view, this is a good sign though there is still a lot to do to reduce the achievement gap between the ethnic groups.

Dalit students have been followed up because they were historically deprieved of education. A positive sign from equality point of view is that Dalit students performed better than the national mean (59%)[33] in the Eastern (64%), Central mountain (61%) and Western Hill (60%), as well as in the entire Central region (> 60%) (table 4.1.22). However, the results are found much lower than the average in the Mid-Western (< 54%) and Far-Western regions (< 57%), Western Tarai (53%), Eastern Hill (54%) and Tarai (56%).

A few Dalit students in the institutional schools (n = 186) have performed always lower than the average (80%). Especially the low performance was found in Eastern Hill (50%), Central Mountain (62%) and Tarai (66%) as well as Far-Western Tarai (71%).

Table 4.1.21 Dalit students' achievement in different Ecological zones and Development regions

School type	Ecological zone	Developmental region					
		Eastern	Central	Western	Mid-Western	Far-Western	Total
Community Schools	Mountain	64.3	61.6	22.1[1]	54.1	48.5	57.8
	Hill	54.1	62.6	60.0	51.9	57.5	57.2
	Tarai	55.6	60.1	53.1	51.0	53.5	56.5
	Total	**56.4**	**61.0**	**58.5**	**52.1**	**55.9**	**57.5**
Institutional Schools	Mountain		61.9				61.9
	Hill	50.4	83.9	79.9	75.5	72.1	77.5
	Tarai	19.5[1]	65.6	81.8[1]	89.1	70.3	69.3
	Total	**45.2**	**68.4**	**80.0**	**80.3**	**70.8**	**75.7**

1) *Groups with a small number of students*

The dataset informs that there are statistically significant differences between the ethnicities in their performance in Nepali grade 3. Students form the "Other" ethnic group are performing the lowest (55%) followed by Dalit, Madhesi, and Janjati students (all 58%); the highest results are in the Brahmin (67%) and Chhetri (61%) castes. In the community schools, Dalit students perform somehow lower than average in the Mid-Western (< 54) and Far-Western regions (< 57), Western Tarai (53%), and Eastern Hill (54%) and Tarai (56%). They perform better than the national mean in the Eastern (64%) and Western Hill (60%), as well as in the entire Central region (> 60%).

Gender and student achievement

Effort have been put globally in reducing the differences between boys' and girls' in school level students' achievement. Because the gender equality is considered to be important in the modern discourse, the matter is handled here extensively than in the previous sections of equality. Basic results are presented in table 4.1.22 and figure 4.1.15.

---

[33] A number of students did not tell their caste/ethnicity. Hence, there is the situation of missing values. Because of the missing values, the total score (58.95 > 59) is somehow higher than that of the total student sample size (57.01 > 57).

Table 4.1.22 Student achievement of boys and girls by school type

Gender	Community				Institutional			
	N	Mean	SD	CV	N	Mean	SD	CV
Boys	6,340	56.9	23.7	41.7	2,246	79.1	15.4	19.5
Girls	6,651	58.4	23.3	39.9	2,210	81.5	13.6	16.7
Total	**14,712**	**57.0**	**23.8**	**41.7**	**4,789**	**79.9**	**14.8**	**18.5**

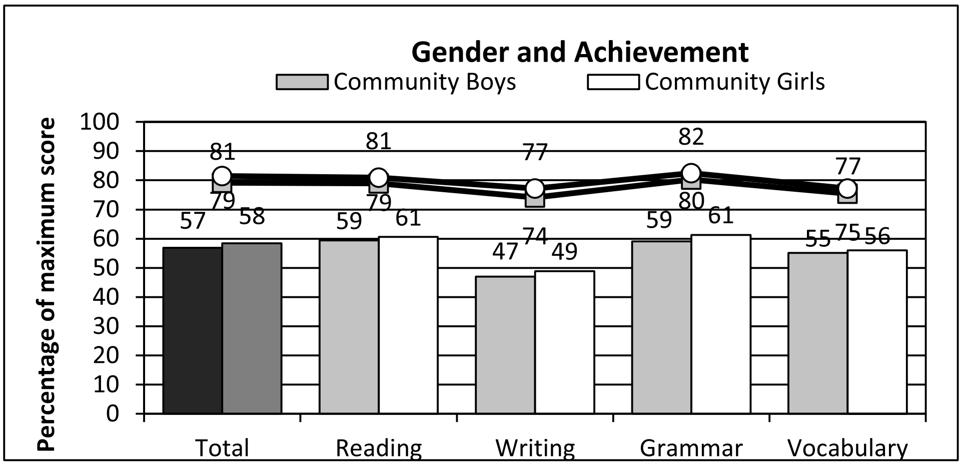

Figure 4.1.15 Comparison of the achievements of boys and girls in various content areas

There are no statistically significant differences between the achievement of boys and girls in the community schools in any of the content areas of Nepali. The differences within the institutional schools are also very small though the girls are performing slightly better than boys in all content areas. The differences are statistically significant in Total score, Reading, Writing, and Grammar ($p < 0.001$). The effect sizes are, however, small (d = 0.15, 0.17, 0.11 and 0.12 respectively), which is a positive sign from the equality point of view.

In the community schools, the achievement difference between girls and boys is moderate – less than 3 percent (figure 4.1.16). Within the institutional schools, girls are performing better. When it comes to the Ecological zones, the differences between boys and girls are very small.

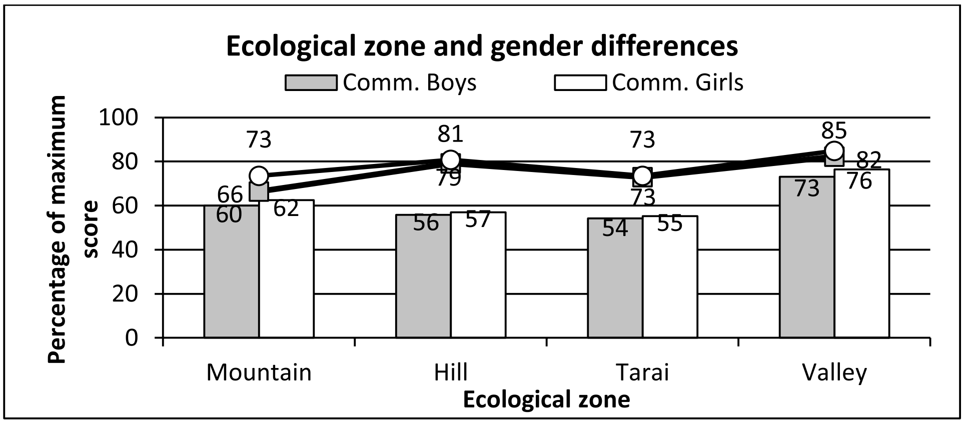

Figure 4.1.16 Ecological zone and gender-wise differences in achievement by school type

There are no notable differences between the Development regions when it comes to boys' and girls' equal opportunities to achieve the same educational goals (figure 4.1.17). The difference between boys and girls is found somehow wider (3 percent) in both the community and institutional schools in the Kathmandu Valley.

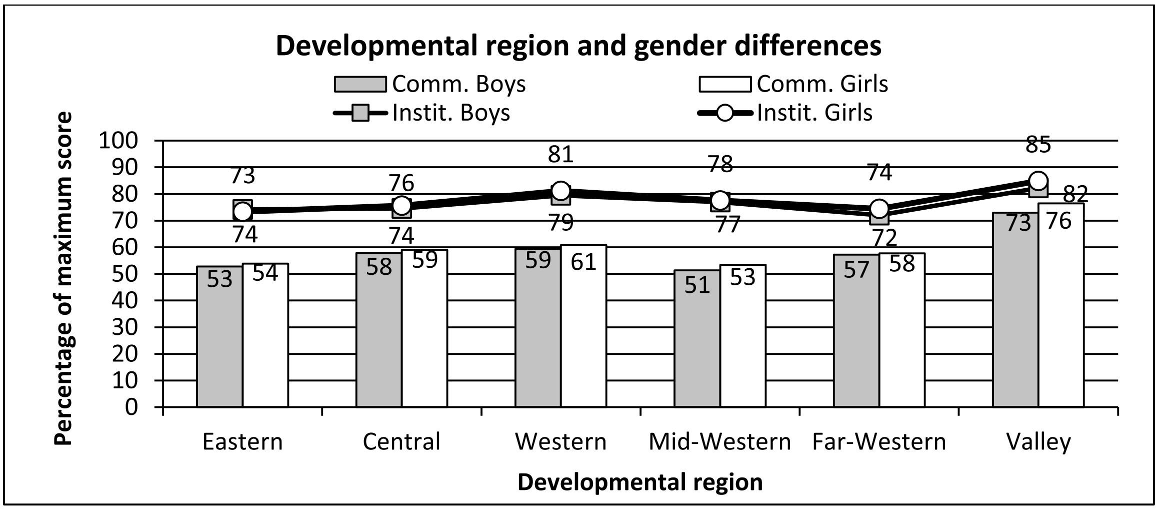

Figure 4.1.17 Development region and gender differences in achievement

The dataset shows that the differences between boys and girls in Nepali proficiency are moderate. From equality point of view, this is a positive sign. A tendency is observed that the girls are slightly out-performing the boys in Nepali at grade 3.

Gender and caste/ethnicity

Within the community schools, the difference between boys and girls is the highest for the Others caste (difference is 4 percent points) and Janjati caste (3 percent points) where the girls outperform the boys. Otherwise, there is no difference between the castes. In the institutional schools, the differences between the sexes are highest within the others students; girls outperform boys by 3 percent points. It is worth noting that within the institutional and community schools, the girls outperform boys in Nepali from all the castes except the Madhesi and Dalit communities. Generally, girls perform better in languages.

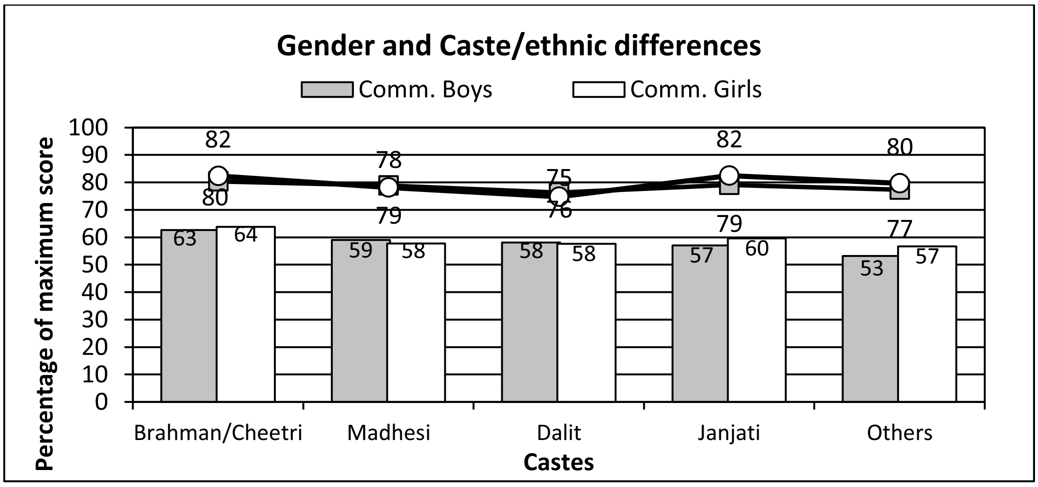

Figure 4.1.18 Caste/ethnicity-wise comparison of achievement of boys and girls

4.1.3 Selected Explanatory Factors and Achievement

The simplistic model in section 2.4 (figure 2.2) represents several possible factors which explain the differences in student achievement. The relation between the student achievement and geographical factors such as Districts, Ecological zone, Development region; school related factors such as school type, school location; and some individual related factors such as home language, ethnicity/caste, gender have already been discussed in previous sections. In this section, some other family and individual related factors such as the socio-economic status (SES) of the students' families, involvement of students in paid and unpaid work beyond the school hours, students' attitude towards Nepali as a school subject, age of the student, and the support provided for studies are taken into consideration. As a sample of deepening school and teacher related factors, the availability of textbooks, homework assigned and checked by the teacher, and other selected activities in the school are also included in this section.

While discussing these factors two things are worth mentioning, particularly for grade 3. First, students had difficulty to response meaningfully to some of the background questionnaire and therefore teachers were asked to help the students to fill the questionnaire. Second, there are missing values in the background questionnaires. For example, in the question of mother's education, 3265 students (16.7%) did not answer the question. This evidently has an effect in the analysis, and therefore the readers need to be critically aware of the situation when it comes to the grade 3 students. In most cases, however, the results are considered to be credible and comparable with grade 5 datasets.

Parents' education, occupation, home possessions and student achievement

There are several variables indicating the socio-economic status. In NASA 2012, these 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 (how many mobile phones, televisions and computers there are 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 economic, educational, and occupational background of the family (see section 2.5). In this section, parents' education is further elaborated, so that the literacy level of the parents is analyzed in relation to the Nepali language achievement.

Several SES-related variables were analyzed by using a data mining tool of SPSS and DTA. The method is 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 NASA 2012 background questionnaire, the parents' education is divided into eight categories: 1) illiterate, 2) literate, 3) grade 10 pass, 4) SLC pass, 5) IA pass, 6) BA pass, 7) MA pass, and 8) Above MA pass. As the information was obtained from the students, some deviant responses were also observed, though, with the huge dataset the results are found to be be credible.

DTA classifies mother's education into four groups with statistically significant differences in students' achievement levels: illiterate (students' average is 62%), just literate (64%), grade 10 passed (72%), and SLC passed or higher (76%). The difference between each group is statistically significant ($p < 0.001$). It means that the mother who is at least grade 10 passed, she can give, on average, +10 percent advance for her child in the national test compared to illiterate mother. If she is SLC passed, she can give + 14 percent advance for her child. These figures are much higher in Nepali than in Mathematics.

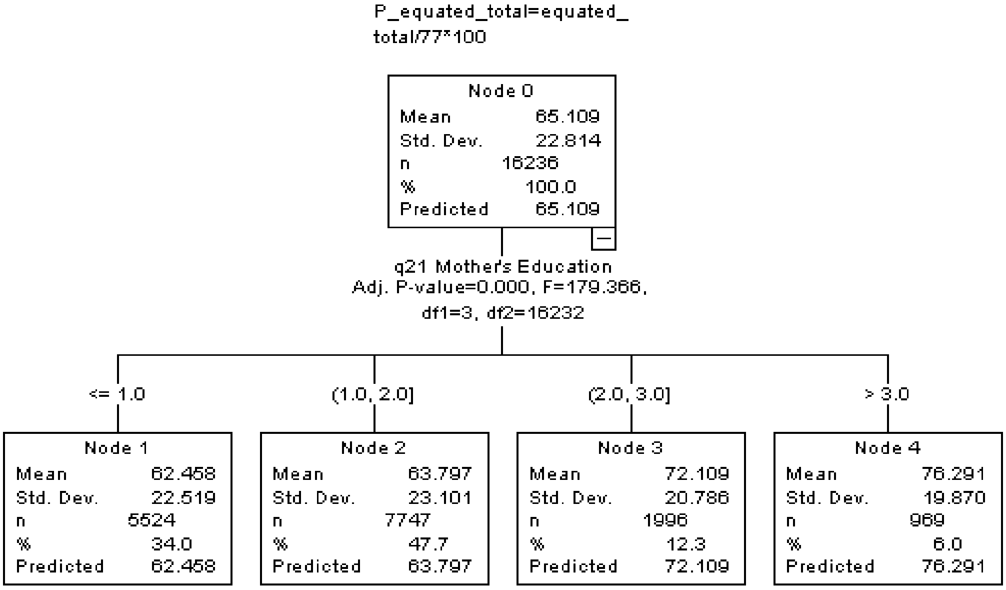

Figure 4.1.19 DTA of mothers' education and students' achievement

Figure 4.1.19 shows that if the mother was MA passed or higher, the advance was + 23 percent over the illiterate mother. Mother's education explains 3% of the student variation ($\eta^2 = 0.034$) which indicates a medium effect size ($f = 0.18$). Obviously, the result shows that the children of the highly educated mothers are mainly found in the private schools.

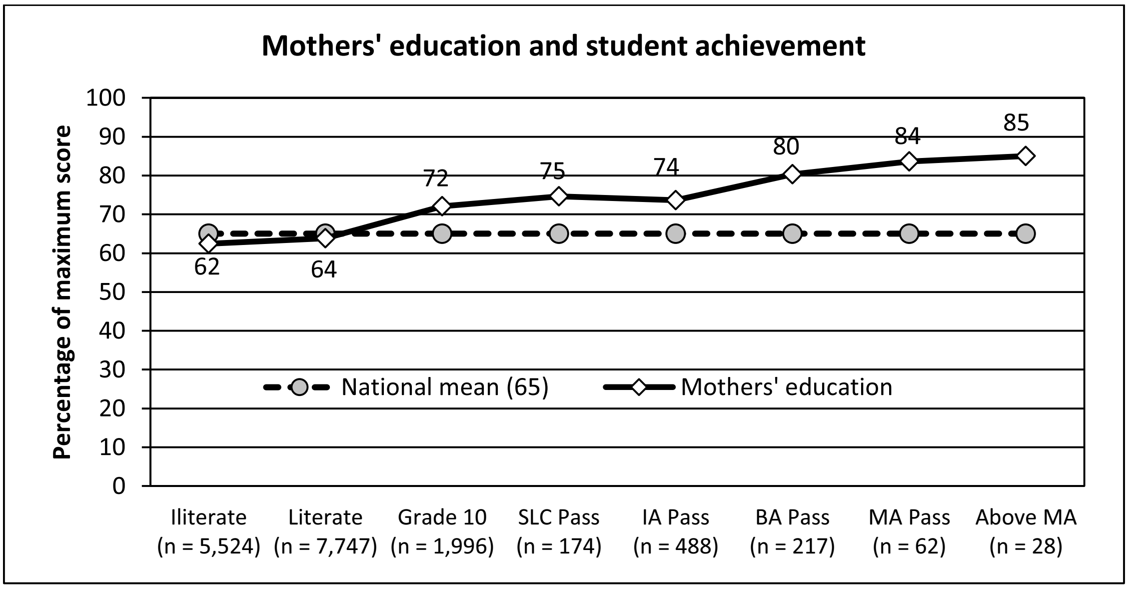

Figure 4.1.20 Mothers' education and students' achievement

In parallel, DTA divides father's education into four categories: illiterate (62%), (just) literate (63%), grade 10 passed (71%), and SLC passed or higher (74%) (figure 4.1.21). The difference between each group is statistically significant (p < 0.001). In practical words, the father who has passed grade 10, can give, on average, + 9 percent point advance for his child in the national test compared with illiterate father.

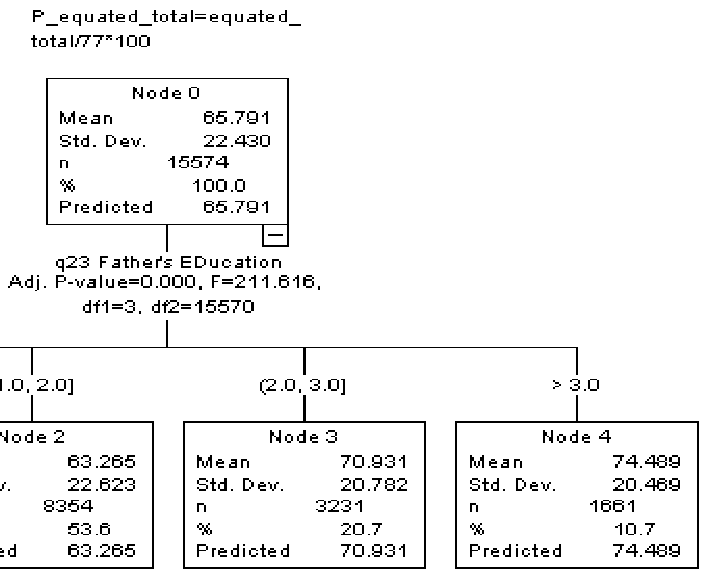

Figure 4.1.21 DTA of fathers' education and students' achievement in Nepali.

Figure 4.1.21 shows that if the father was MA passed or higher, the advance was + 22 percent points over the illiterate father. Obviously, the high average means that the children of the highly educated fathers (as well as of mothers) are mainly found in the private schools. Fathers' education explains 4% of the student variation (η^2 = 0.043) which indicates a moderate or high effect size (f = 0.21).

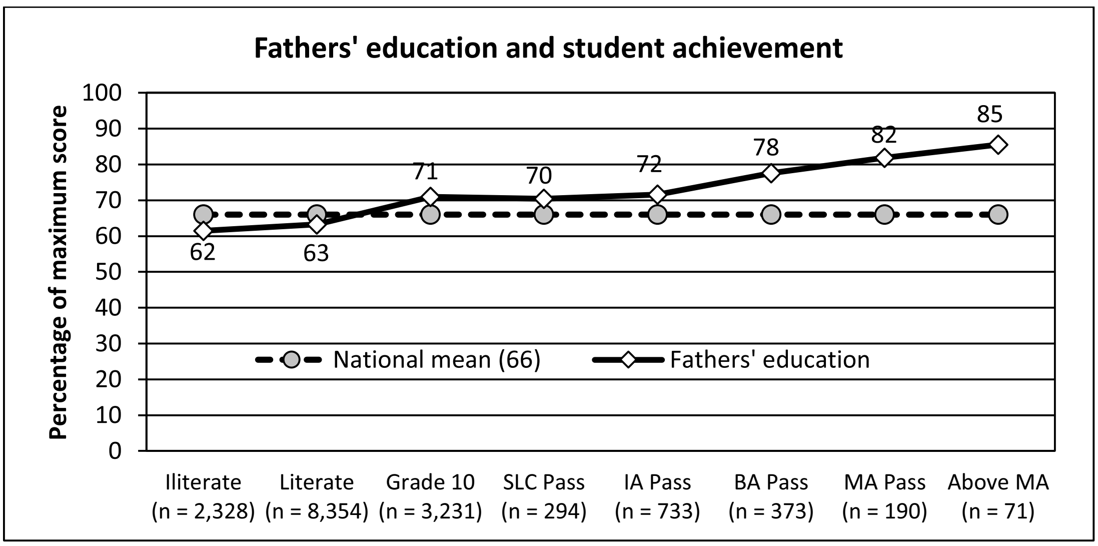

Figure 4.1.22 Fathers' education and students' achievement

It is noteworthy that mother's education has given the same leverage to children's achievement in grade 3 Nepali as that of father's (33 percent at the highest). In both cases, the effect size is moderate or high ($f = 0.18$ to 0.21) showing that the difference between the highest and lowest group is remarkable.

When combining the mother's and father's education, the highest results are in the group where both the father and mother have passed (at least) the grade 10 (74%) or when the father has passed SLC and mother has passed (at least) the grade 10 (65%). This indicates a general trends that the higher the parents' education the children achieve better results.

In what follows with the final SES variable, the cut-off for parental education was set to "grade 10", that is, when being passed the grade 10 (or higher), the indicator for mothers' (and fathers') education for SES was set to 1, and the education lower than SLC passed gave the value 0.

Parent's occupation

The occupation of parents was categorized into eight groups: 1) working abroad, 2) farming and working at home, 3) only working at home, 4) teaching, 5) services, 6) business, 7) daily wages, and 8) working at other's home. The result related to mother's occupation is presented in figures 4.1.21 and 4.1.22 and to father's occupation in figures 4.1.23 and 4.1.24.

While comparing the students' responses by DTA, the achievement is the lowest when mothers' occupational background comes from working abroad (60%). Statistically speaking, it is significantly lower than when the mother works at others homes (60%). The achievement of the students high when mother works as a teacher (81%) or in services (80%). Mother's occupation explains 3.7 percent of the student variation ($\eta^2 = 0.037$) which indicates a moderate effect size ($f = 0.19$).

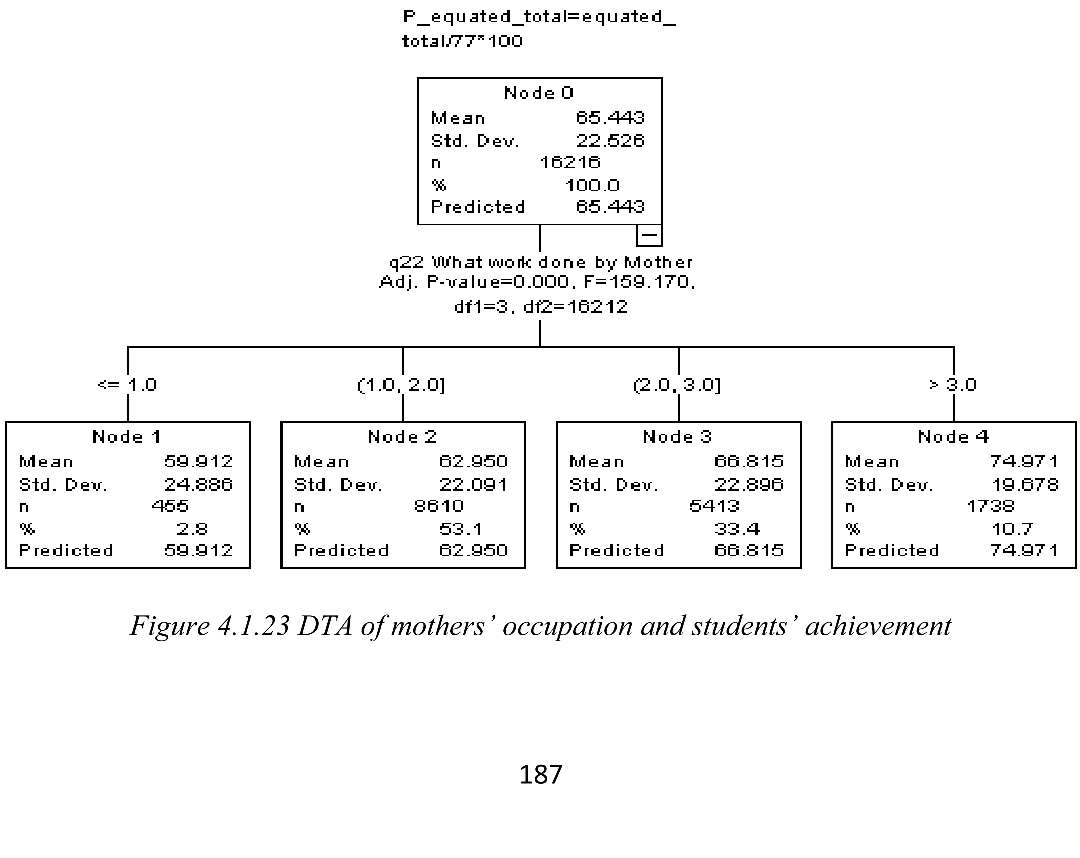

Figure 4.1.23 DTA of mothers' occupation and students' achievement

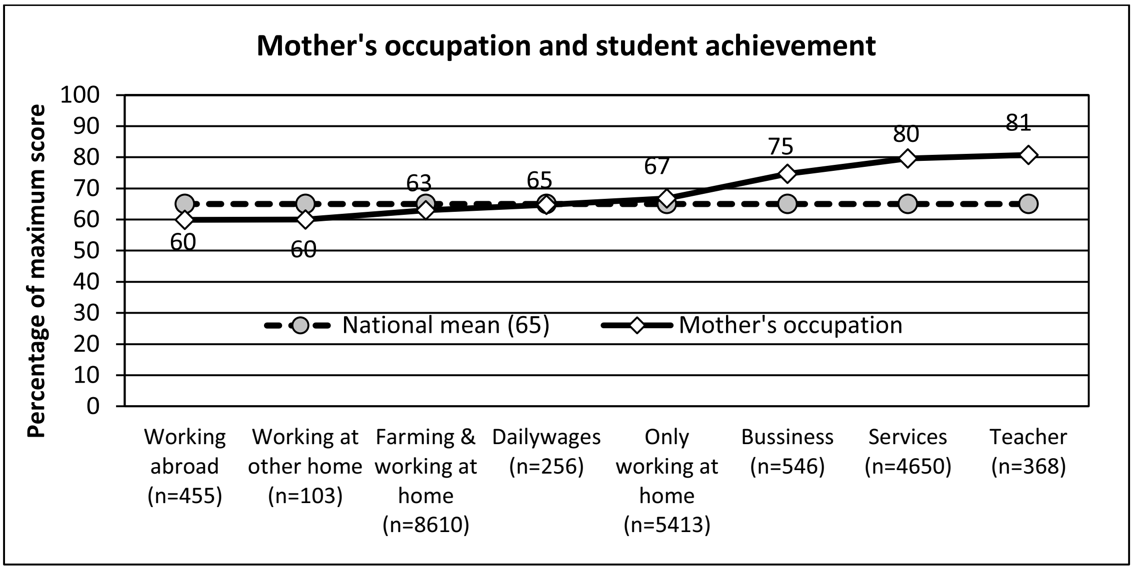

Figure 4.1.24 Mothers' occupation and students' achievement

When it comes to fathers' occupation, on the basis of DTA and ANOVA (figure 4.1.24 and 4.1.25), the main division is whether the father works in agriculture (61%) or not (> 64). More precisely, if the father is involved either in agriculture related works or in household chores (that is, is probably unemployed), the Nepali skills are remarkably lower (59%) compared with the father from business (75%), teaching (78%) or service profession (77%). The difference between the lowest and highest group is 19 percent which is a remarkable difference. Fathers' occupation explains 8.5 percent of the student variation ($\eta^2 = 0.085$) which indicates a high effect size ($f = 0.30$).

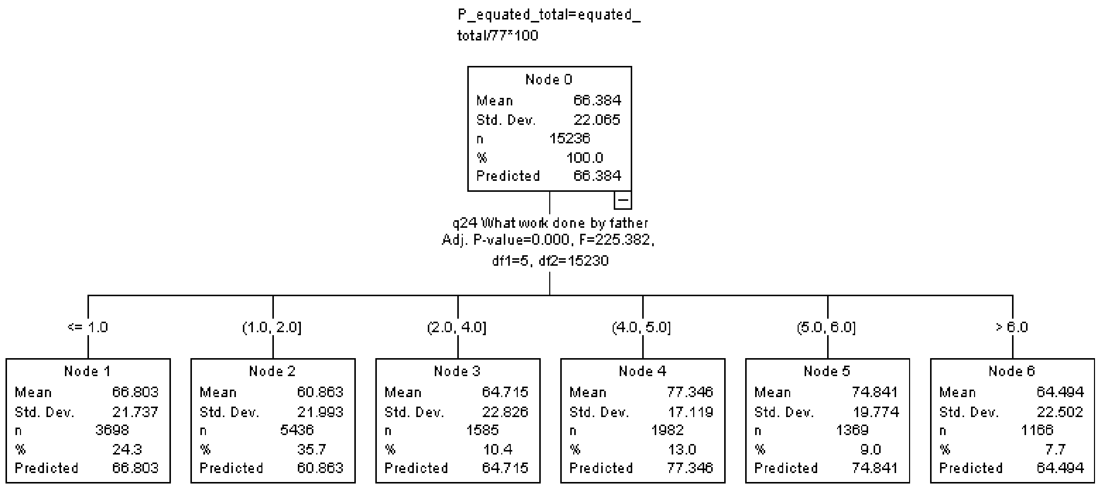

Figure 4.1.25 DTA of fathers' occupation and students' achievement

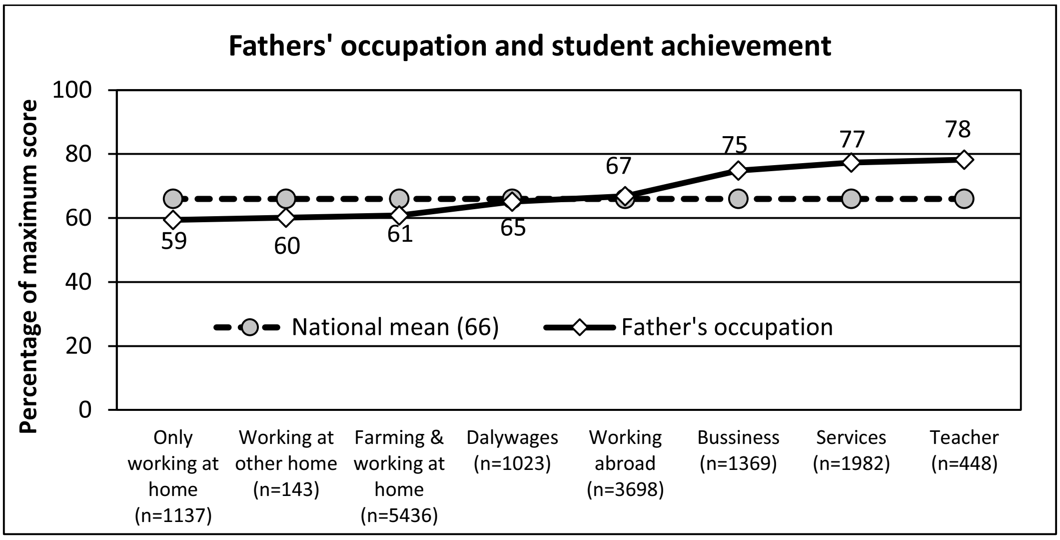

Figure 4.1.26 Mothers' occupation and students' achievement

When combining the occupation of fathers' and mothers', DTA shows that the lowest student achievement is found in the families where the fathers' occupation is unknown but the mother works abroad (40, n = 65), or in agriculture (51, n = 1,218), or where the father comes from an agricultural background and the mothers' occupation is unknown (46, n = 150), or she works in home (59, n = 1,109). The highest achieving students come from the families where the father comes from the services (77, n = 1,982) or business (75, n = 1,369) regardless of mothers' occupation, or if the father is a teacher and the mother is either teacher, service holder, or business woman (76, n = 203). It is worth noting that service and business occupations are more probably urban than rural occupations.

For the later use as a SES indicator, the cut-off for the parents' occupation was set 0 for agriculture and 1 for all other options.

Home possessions and accessories

Facilities and resources available in home influences 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 help in studying at home: whether they have a table for study, a separate room or peaceful place for study, a computer for school work, software for the computer assisted learning, internet facilities, their own calculator, access to classical literature and poetry books, dictionary, or artistic things like pictures and books that help them for study such as dictionary. Another type of home possessions includes different types of normal home accessories (and hence, in what follows these are called home *accessories* to differentiate them from home *possessions*) such as the number of mobile phones, televisions and computers.

There are 11 questions in the student background questionnaire related to home possessions. Each was scored 1 if the student had the access to this possession (e.g. having a separate room or a table for study). Adding these items up, the maximum score was 11,indicating that the student reported to have access to all of the possessions, and the lower the score the fewer possessions they have at home. Figure 4.1.27 shows the connection of home possessions and achievement level. Except for the highest category, the achievement level of the students raises logically if there is access to more home possessions.[34] Pearson correlation between the achievement level and the factor ($r = 0.16$) is statistically significant ($p < 0.001$) and indicates medium effect size ($d = 0.34$).

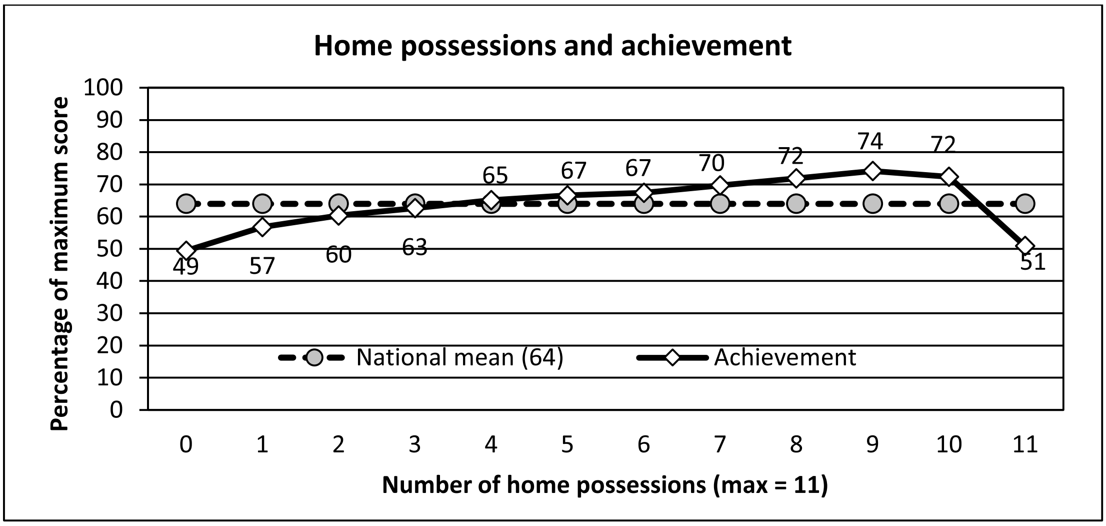

Figure 4.1.27 Relation between home possessions and achievement

For the later use in SES, the cut-off for the factors was set for 6 possessions[35]: if 6–10 items were met in the background questionnaire, the student was given 1, otherwise 0.

The same pattern, i.e., the more possession, the better results, can also be seen with home accessories, as seen in figure 4.1.27. The question in the background questionnaire was set differently compared with home possessions. Regarding the accessories it was asked "how *many* of the following accessories do you have in your family?" with the options 0 – 3 (or more). For the indictor, the availability of the home accessories is dichotomized in the same way as the home possessions. After dichotomizing the items individually by using meaningful cut-offs found with ANOVA and DTA (and maximizing the differences in achievement level, see table 4.1.24), all three indicators were summed up.[36] The maximum score was 3, which indicates that the student's home possessed a set number of *all* of the accessories.

[34] The same phenomenon, though not as radical as here, was seen also in 2011 datasets (see ERO, 2013, figure 3.1.24 and 3.2.22): the students who selected all the possibilities may not have understood the question in the same way as the other students. Most probably, in any case, they actually did not have all the possessions though they claim that.

[35] The cut-off was selected to be 6 because the willingness to keep the boundaries comparable over the subjects. In grade 3 dataset, the cut-off could have been three possessions.

[36] There was also fourth item in the questionnaire – the number of radios in home. However, this item behaved pathologically in the analysis: the more there were radios in home the less achievement. Hence, it was not taken as an indicator for SES.

Table 4.1.23 Dichotomizing the indicators for home accessories

Accessory	Cut-off for 1	Cut-off for 0
Mobile Phone	2,3	0,1,missing
Television	1–3	0, missing
Computer	1–3	0, missing

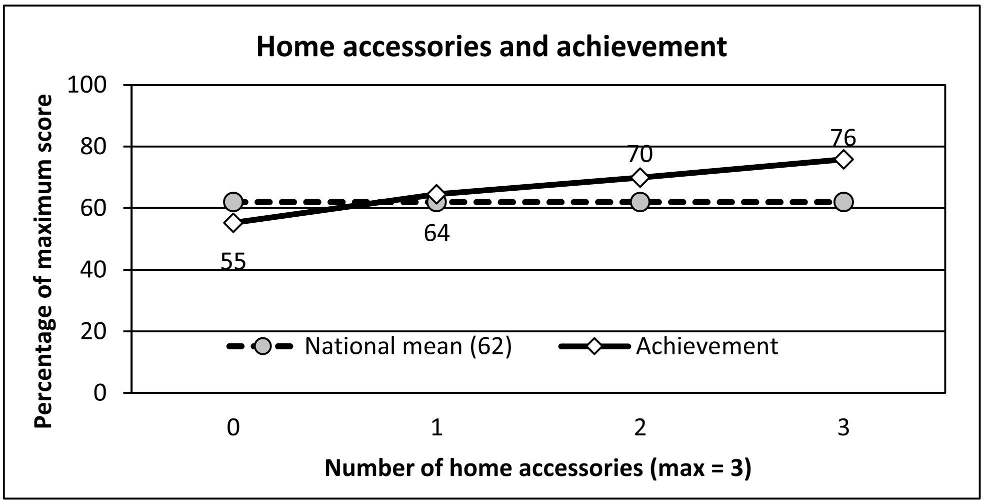

Figure 4.1.28 Relation between number of home accessories and achievement

Figure 4.1.28 clarifies that as the number of home possession or accessories increases, students' achievement raises from 55% (if none of them are available) to 76% (if all three of them are available). Availability of all the stated facilities indicates the higher SES of the family. Correlation between the number of home accessories and achievement is r = 0.30 ($p < 0.001$) which is certainly positive and indicates a moderate or high effect size (d = 0.72).

The dataset suggests that parents' educational level predicts the children's future achievement level in Nepali. Especially harmful for the achievement level is the situation where the father or mother or both are illiterate. The data shows that 34.0% of the students have an illiterate mother and 14.9% an illiterate father.

The dataset is also evident that either economic or intellectual capacity or both at home helps children to increase their Nepali proficiency. If the father or mother or both are coming from an agricultural or related occupation, the students' achievement in Nepali is significantly lower than with the children coming from other occupational groups. Of the total sample size, 53.1% of the mothers and 35.7% of the fathers worked in agriculture or only at home.

The dataset shows that when children have very few home possessions – zero to three out of the 11 – the achievement level is remarkably lower than the national average (< 63%). With nine to ten possessions, the average score is very high (> 72%) compared to the national average. The same is true of home accessories: When none or only one accessory indicator out of three is met, the results are lower than average (55–64%) and when two or more are met, the results are remarkably higher (70–76%). It is found that 7.1% the students did not have any of the home possessions and 44.7% had no accessory.

SES and Achievement

The socio-economic status of the family was formed on the basis of seven indicators which were all first dichotomized. The variables (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 up as SES and changed into the percentage of the maximum score (P−SES). Deeper description of the transformations is seen in section 2.5. The P−SES represents the percentage of SES of the student's family; 100 means that the student has the highest possible SES measured with these variables and with these transformations (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 P−SES by using Univariate GLM (that is, the Regression modeling) shows the strong relation between SES and achievement. Figure 4.1.29 presents the relationship between SES of the students and their achievement.

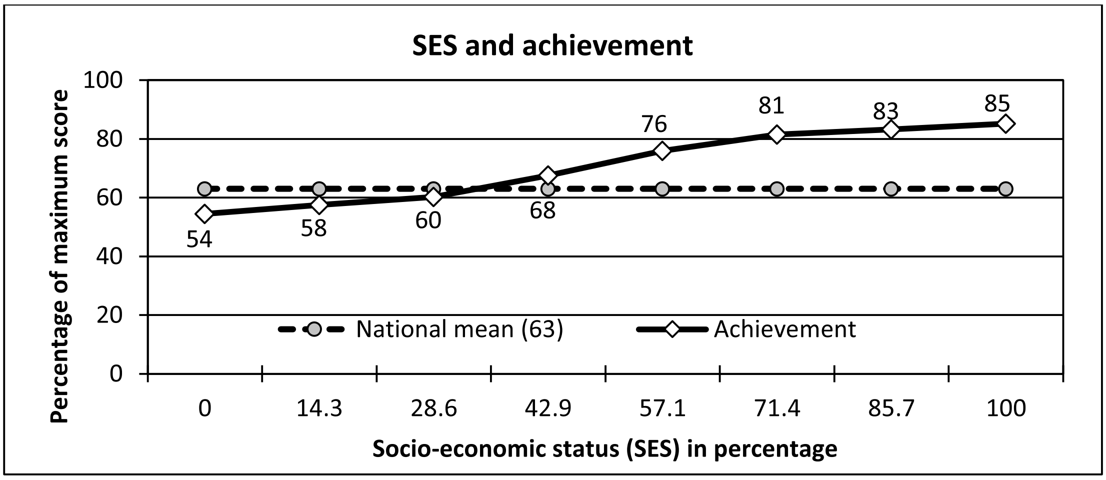

Figure 4.1.29 Relation between SES and achievement

Figure 4.1.29 shows a positive relationship between SES and the Nepali achievement – the higher the SES the higher the achievement. Pearson correlation between the variables is $r = 0.37$ which is a high value ($p < 0.001$) and indicates very high effect size ($d = 0.93$). The difference in achievement between the lowest SES group (54%) and the highest one (85%) is remarkable. SES explains somehow 14% of the student variation ($\eta^2 = 0.144$) which is not very high a percentage compared with, for example, the English dataset ($\eta^2 = 0.311$), but it has remained at the same level as was found in Mathematics in the grade 3 dataset (0.129, see section 3.1).

It is worth noting that SES as a variable is more school-related than being the student-related factor. The correlation of SES and Achievement is $r = 0.37$ in the student dataset but $r = 0.48$ in the school-wise dataset. It is also worth noting that even though the SES is controlled in the student-wise dataset[37], there is still a statistically significant difference

[37] Because the attending of students to the private school is imbedded in the SES, the school type does not explain the achievement in ANCOVA when controlling the SES. For the ANCOVA, another SES – without the school type – was created.

between the community and institutional schools (p = 0.001). However, the effect size is reduced from f = 0.44 to f =30, that is, from high to moderate.

From sociological point of view, it is interesting to know which of the individual indicators of SES are not met in those families where the children perform the lowest. Figure 4.1.30 illustrates the fact that in the families meeting less than four SES indicators, challenge lies mainly among the three factors marked with the dark circle, triangle, and square: both mother's and father's education is low and the child does not attend the private school.

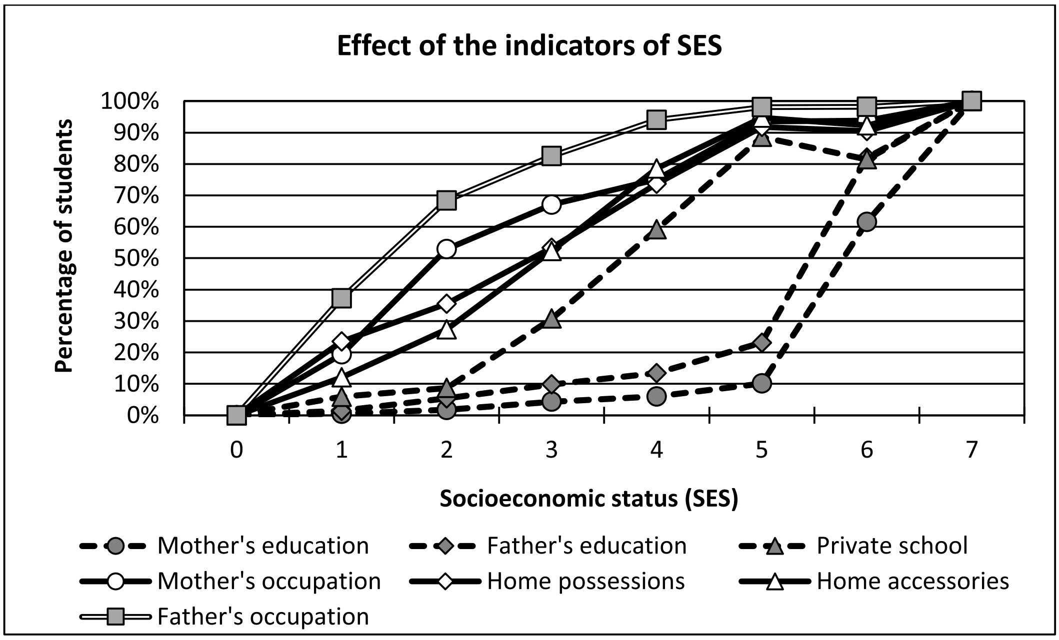

Figure 4.1.30 Effect of individual SES indicators in achievement

The dataset suggests that the socio-economic status plays a vital role in the students' achievement in Nepali. The difference between the lowest and highest SES groups is remarkable (31 percent). This means that if the SES of the lowest performing students rises into a decent level, the results in these groups will also improve remarkably. Especially challenging is the situation in the families where the father or both parents are illiterate or just literate. There is no huge difference between illiterate and just literate families, but the gap is huge from illiterate/literate to grade 10 or above. As many as 25.6% of the students are at the lowest level of SES.

Working beyond the school hour and achievement

Several questions were set in the student background questionnaire regarding the students' activities outside the school. Two of them are briefly handled here: (1) *Working after the school for a paid job* and (2)*Participating in household work/chores*. The values of the variables are divided into five categories: 0 (no time at all), 1 (less than 1 hour per day), 2 (1–2 hours per day), 3 (2–4 hours per day), and 4 (more than 4 hours per day).

The DTA indicates that, when it comes to working after school, the cut-off is on whether the students work for a paid job or not. The DTA shows that when the children have no paid work at all, the results are above the national average in both community and institutional schools. If the students are working for the paid job – even if it is less than one hour – the results are statistically significantly lower than the average. The ANOVA shows that the relationship is firm (p = 0.008 in the community schools and p = 0.033 in the institutional schools)[38] though mildly negative (f = 0.08 in the community schools and f = 0.18 in the institutional schools) when students need to be engaged in paid work before and after school. It is notable, though, that most of the grade 3 children do not need to be engaged in paid work. Working after school implies that the family is poor and the extra income is needed. It is obvious that when the student needs to work for more than 4 hours per day, there is no time or energy to concentrate in study and to handle school homework. Within institutional schools, there is notable difference in achievement between the children working over 4 hours per day (58, n = 511) and that of those with no need to work at all for paid job (82, n = 2,913) (see figure 4.1.31 institutional schools). The achievement of working children is notably lower than those who do not need to work.

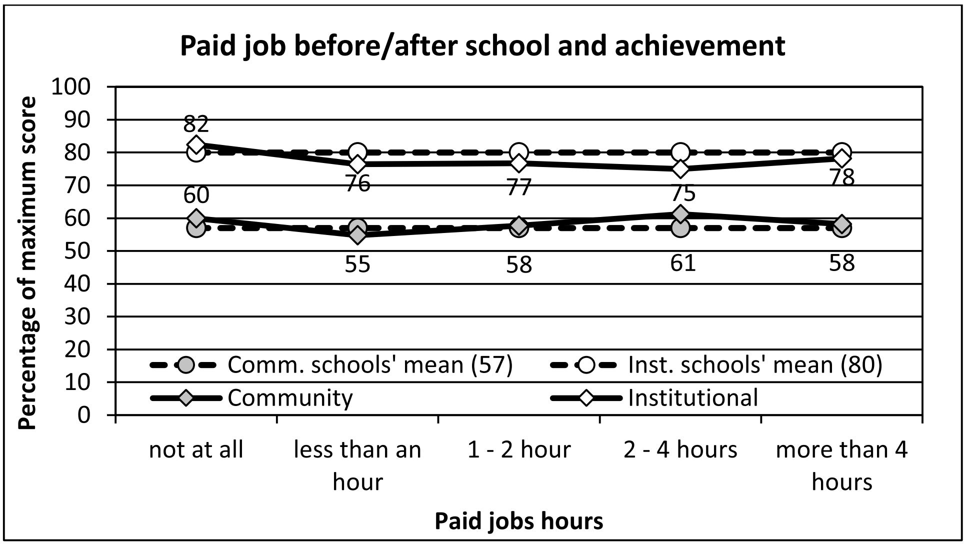

Figure 4.1.31 Relationship between achievement and paid job beyond school time

In relation to the involvement in unpaid household work, it is usual – and a supportive practice in families – that the children take part in household chores at home, which is also part of the socializing process of the children. The DTA shows that when the child spends some time (less than two hour) for the household chores, the results are statistically higher (56–70%) than those who spend no time for house work at all (51%). The effect of not participating in the household chores is larger in the community schools than in the institutional schools. However, in the institutional schools, it does not make any difference whether the children work for two hours or less or not at all; the effect is seen when spending four or more hours in household chores. Within the community schools, the results are significantly lower if school children are not participating in the chores. Differences are significant (p = 0.033 in community schools and p = 0.008 in institutional schools)[39] though the effect size is small or moderate (f = 0.18 in community schools and f = 0.08 in institutional schools). It is somehow interesting that more than 3.5% of the

[38]Multilevel modeling with factorial design shows significance of $p < 0.001$.
[39]Multilevel modeling with factorial design shows significance of $p < 0.001$.

students (n = 684) reported that they spend more than 4 hours per day doing household work but they achieve high score. In the rural area, it is obligatory for children to be involved in cattle raising when the cattle is far from home. It is understandable that, in these cases, there is not much energy to concentrate on their study. On the other hand, it is also possible that those who do not work for household chores might involve in paid job. This explains partly why, in community schools, the students achieve lower score than in the institutional schools.

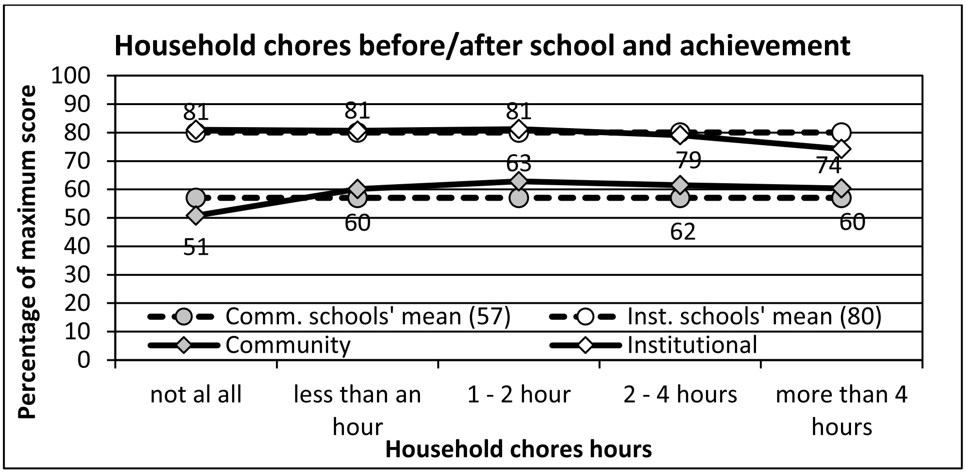

Figure 4.1.32 Household work and achievement

The dataset reveals that either working for a paid job or for four hours per day in an unpaid or household work before or after school significantly reduces the school achievement of the student. However, some amount of household work up to two hours per day is not found to have disturbed the learning of students in Nepali subject. It is found that 35.4% of the students worked for the paid job and 14.8% spent more than 2 hours in household chores.

Attitude towards the subject and achievement

In the context of the assessment of Nepali language achievement, attitude explains what the students think about Nepali, its usefulness in their daily life and future. There is a more or less firm relationship between the attitude of the students and achievement. Though the connection is not always clear, the correlation between achievement and attitude towards the subject as well as self-efficacy in the subject is widely studied (see in Mathematics, for example, House & Telese, 2008; Shen & Tam, 2008; Kadijevich, 2006; 2008). Some researchers have noticed remarkable differences in correlation between countries (e.g., House & Telese, 2008; Kadijevich, 2006; 2008; Wilkins, 2004; Shen, 2002; Papanastasious, 2000; 2002; Stevenson, 1998). In some countries, the correlation between attitude and achievement is found near zero, like in Macedonia (Kadijevich, 2008, in the Philippines (Wilkins, 2004), in Indonesia (Shen, 2002) or in Moldova (Shen, 2002) whereas in some other countries, the correlation is found as high as 0.60 (e.g., in Korea, Shen, 2002). In NASA 2011, it was noticed that the grade 8 students were not consistent in the attitude test and the reliability of the international test remained low (see ERO, 2013, table 2.11).

In NASA 2012, technically speaking, the same shortened version of Fennema–Sherman Attitude Scales (FSAS, Fennema & Sherman, 1976) as is used in several international comparisons like TIMSS and PISA studies was used. The original scales included nine dimensions but in these international comparisons only three are used with four items on each dimension and two negative items on each of the first two dimensions (see the detail in section 2.3 and 2.5). The names of the original factors were "*Liking Nepali*", "*Self-Efficacy in Nepali*", and "Experiencing utility in Nepali" (compare naming in, e.g., Kadijevich, 2006; 2008). Because of students' inconsistent manner in answering the attitude scale in NASA 2011, only the dimension of "Experiencing utility in Nepali" was taken into the measurement instrument of grade 5 students. Reliability of the score of five items is sufficient ($\alpha = 0.89$). The relation between the attitudes (divided into seven groups with somehow an equal number of the students, that is, septiles[40]) and achievement score is shown in figure 4.1.33.

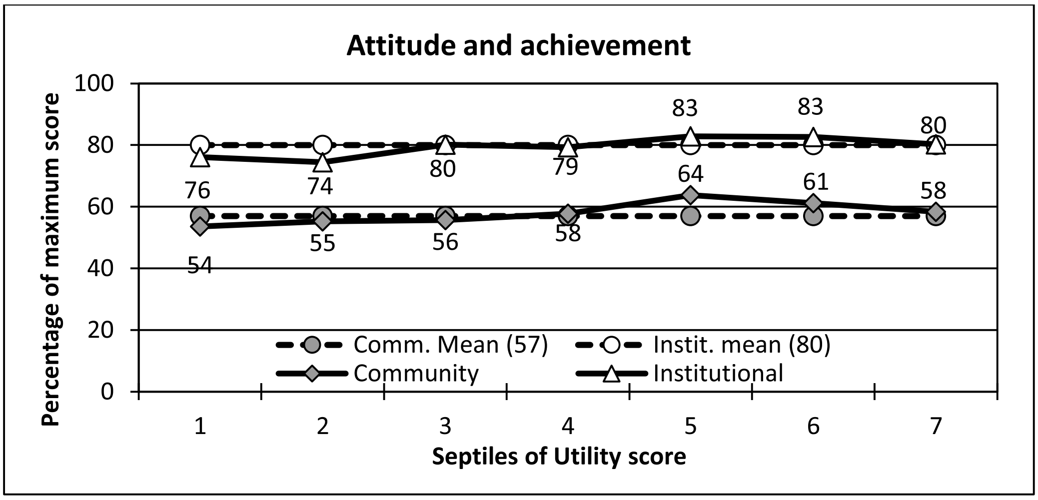

Figure 4.1.33 Relation between attitude and achievement by school type

There is a positive correlation between the students' attitude and achievement in Nepali in the whole dataset ($r = 0.14$). The influence is small ($d = 0.24$) though, which indicates that difference between the means of the lowest attitude group (57%) and highest one (70%) is remarkable. In the whole dataset, the categorisation of attitude to seven groups explains the achievement level as being somehow 3% ($\eta^2 = 0.027$). Influence is higher in the institutional than community schools. The possible reason is that, in the community schools, the highest attitude group does not correlate logically; it is also possible that, within the highest attitude group, there are many students who have either fooled in the test or did not understand the questions (see the same kind on phenomenon in the SES analysis above). The difference between the lowest and highest attitude group is 10 percent in community schools ($f = 0.12$) and 7 percent in institutional schools ($f = 0.16$).

[40] The original score is short (maximum was 15 points) and quite many students (36%) gave the maximum score. Hence it was not possible to form more precise classification such as deciles. Seven classes (septiles) was the most precise alternative with the given dataset.

The relation of the sense of utility in Nepali and achievement is clear though it is not known whether the positive attitude is a consequence of high achievement or the other way round. From statistical point of view, on the basis of simple ANOVA GLM procedure, attitude explains 2.7% achievement. Hence, it is likely that, in grade 3 Nepali subject, the better achievement is a consequence of more positive attitude than other way round.

The dataset indicates that positive attitude towards the subject correlates with higher achievement in Nepali. The better achievement is more probable consequence of more positive attitude rather than other way round.

Age of students and achievement

The age of the students attending grade 3 varies widely. Some students have mentioned their age even below seven years and some above 13. All the students below 7 were encoded as 'up to 7 years', and all students above 13 were encoded as '13 years or above". The descriptive statistics of the mean in each year are given in tables 4.1.24, and 4.1.25 and depicted in figure 4.1.34.

Table 4.1.24 Descriptive statistics of the students' achievement in different age groups

Age	N	Mean	SD	CV
Up to 7 years	436	55.9	24.7	44.2
8 years	2,252	61.4	23.8	38.8
9 years	5,341	64.4	24.2	37.6
10 years	5,714	64.2	23.8	37.1
11 years	2,428	62.2	23.7	38.2
12 years	1,294	60.1	23.4	38.8
13 years or higher	599	57.8	24.6	42.6
Total	**18,064**	**62.6**	**24.0**	**38.4**

Table 4.1.25 Student achievement in different age groups by the type of school

Age	Community				Institutional			
	N	Mean	SD	CV	N	Mean	SD	CV
Up to 7	397	53.8	24.2	44.9	39	77.6	19.1	24.6
8 years	1,933	58.4	23.7	40.6	319	79.6	14.6	18.3
9 years	3,633	56.7	24.1	42.5	1,708	80.8	14.1	17.4
10 years	3,990	57.2	23.6	41.3	1,724	80.4	14.6	18.1
11 years	1,901	57.8	23.6	40.9	527	78.3	16.1	20.5
12 years	1,107	57.1	23.1	40.4	187	78.1	15.8	20.2
13 or above	556	56.2	24.5	43.6	43	77.7	14.7	18.9
Total	**14,712**	**57**	**23.8**	**41.8**	**4,789**	**80**	**14.8**	**18.5**

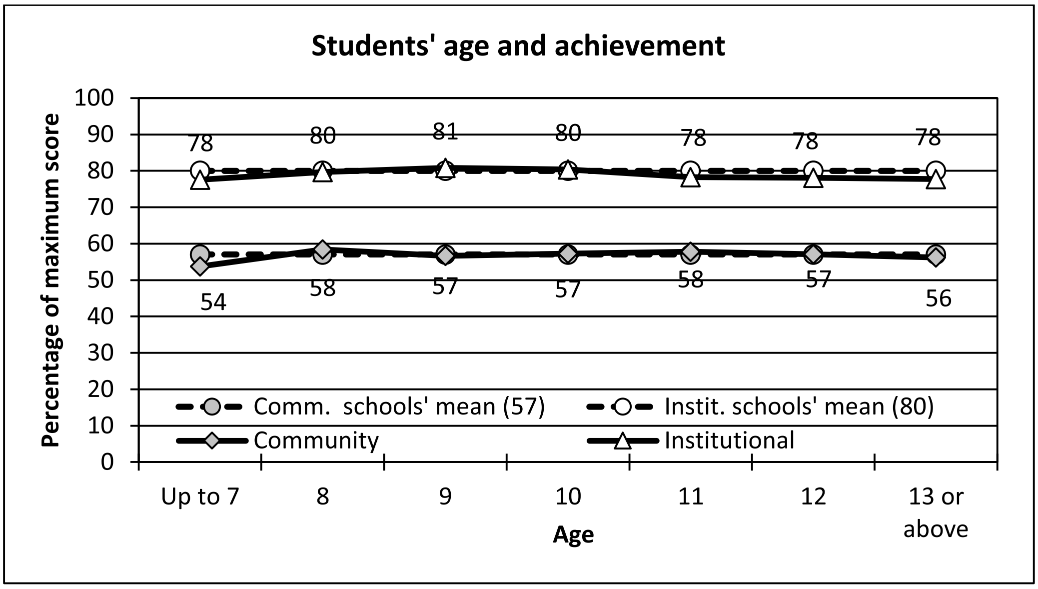

Figure 4.1.34 Relation between age and students' achievement

The data shows that the best achievers are those students who are at their proper age for grade 3 studies (8 to 10 years old). The higher age means that the students have either started much later than they should have, or they have repeated the same grade or early grades. Results are weaker in this case. The phenomenon is not as clear in Nepali language in grade 3 as it is in Mathematics grade 3 and in English grade 5. The achievement level is somehow lower than the average when the students are at the age of 13 or higher (78 in institutional schools and 56 in community schools). Correlation between the age and achievement is $r = 0.00$ ($p = n.s.$) in community schools and is $r = 0.04$ ($p = 0.008$) in institutional schools indicating low effect size ($f = 0.03$). It is good to consider that these “over-aged” students are at school to learn; though they should have been provided at a much earlier age with extra tuition or additional support.

The dataset suggests that the highest performance is found with the students studying at their normal age group, that is, at the age of 8 to 10 years. Otherwise, the achievement decreases as the age increases.

Support to study and student achievement

The relation between the support received for studies and achievement was analyzed based on information provided to the question “*Who helps you when you do not understand what you have read?”* Only one option was selected as the response to this question. In many cases, there might be several support providers, which is not possible to detect here in. However, the descriptive statistics of major supports are given in tables 4.1.26 and 4.1.27.

Table 4.1.26 Support for the study to the student and achievement

Support received	N	Mean	SD	CV
Tuition	1,272	67.4	23.1	34.3
Teacher	1,903	66.7	21.4	32.1
Mother	1,909	65.7	24.4	37.2
Brother/Sister	6,773	65.4	22.5	34.3
No one	334	60.9	25.4	41.7
Father	4,905	59.9	24.4	40.8
Total	**17,096**	**64.1**	**23.4**	**36.6**

Table 4.1.27 Support to the student and achievement by type of school

Support received	Community				Institutional			
	N	Mean	SD	CV	N	Mean	SD	CV
Teacher	1,306	60.8	21.8	35.8	597	79.5	13.5	17.1
Brother/Sister	4,988	60.1	22.5	37.5	1,785	80.4	14.0	17.4
Tuition	784	59.2	23.6	39.9	488	80.5	14.9	18.5
Mother	1,204	57.2	24.7	43.2	705	80.3	15.4	19.2
Father	4,066	55.7	23.9	42.8	839	80.3	15.1	18.8
No one	261	54.6	24.4	44.8	73	83.8	12.4	14.8
Total	**12,609**	**58.3**	**23.3**	**40.0**	**4,487**	**80.3**	**14.5**	**18.0**

An external support is, in many cases, necessary for the students to gain better than average marks on the test. In the whole dataset, there is about 6 percent difference between those who do not get any kind of support (61%) and those who receive the tuition support (67%). It is more likely that the children receiving the private tuition in the institutional school and the community school spend more time on the homework which explains the high score. Those who received support from their father or teacher gained notably lower than the average – even lower than those with no tuition at all.

The support provided to the students of community schools by their teachers, brothers or sisters, or tuition is seen the most effective which has helped raise achievement from 59% – 61%. In institutional schools, on contrary, the highest results come when the students have studied just by themselves (84%). In institutional schools, it is seen that when the teacher is the main to support, the result is less than the average. The effect of the support is, in any case, very low: effect size is $f = 0.09$ in the community schools and $f = 0.03$ in the institutional schools, indicating that the difference in mean is not notable.

The dataset shows that the support provided by the mother, brother and sister raises the achievement level more than the support provided by the father. In the whole sample, the highest achieving group is the one who receive private tuition and support from the teacher. However, the difference between the highest and lowest performing groups is not notable. It is possible that the group receiving private tuition also spends more time on their homework, explaining the higher score.

Availability of textbook and student achievement

The data shows that there were still some students who did not have the proper textbook up to the end of the academic session. Table 4.1.28 shows the descriptive statistics regarding the availability of the Nepali textbook and the achievement.

Table 4.1.28 Availability of textbook of Nepali and the achievement

Availability of Nepali textbook	N	Mean	SD	CV
Yes	16,114	64.8	23.1	35.7
No	741	54.3	25.0	46.1
Total	**16,855**	**64.3**	**23.3**	**36.2**

Out of 16,855 students who responded to the question, 4.4% (4.6% in the community schools and 3.7% in the institutional schools) did not have a textbook available with them. The relation between the textbook and achievement is significant ($p < 0.001$) though the effect size in the whole dataset is small ($f = 0.09$). The difference in achievement is 9.7 percent in community schools and 9.0% in the institutional schools.

According to the dataset, 4.4% of the students lack textbook in Nepali. The achievement level of the students without textbook is significantly lower than those who have access to the textbook.

Homework given/checked and achievement

Based on the students' responses to the questionnaire, statistics related to the assignment of homework and its checking is presented in tables 4.1.29 and 4.1.30 and further illustrated in figure 4.1.35.

Table 4.1.29 Assigning and checking the homework and the achievement

Status of homework	N	Mean	SD	CV
Given Someday-Checked Everyday	3,079	67.1	23.4	34.8
Given Everyday-Checked Someday	1,247	65.0	23.9	36.8
Given Everyday-Checked Everyday	12,083	63.7	23.2	36.4
Given Everyday-Not checked	200	56.4	23.8	42.2
Given Someday-Not checked	116	56.0	23.8	42.4
Not given	101	51.8	26.6	51.3
Given Someday-Checked Someday	83	50.9	24.6	48.4
Total	**16,909**	**64.2**	**23.4**	**36.5**

Table 4.1.30 Homework given and checked by the type of school

Status of homework	Community				Institutional			
	N	Mean	SD	CV	N	Mean	SD	CV
Given Everyday-Checked Everyday	9,454	59.2	23.1	39.0	2,629	80.1	14.7	18.4
Given Someday-Checked Everyday	1,855	58.0	23.8	41.0	1,224	80.9	14.1	17.4
Given Everyday-Checked Someday	748	53.9	23.1	42.9	499	81.6	12.8	15.7
Given Everyday-Not checked	151	50.5	23.0	45.6	49	74.4	15.7	21.1
Given Someday-Not checked	72	49.3	24.0	48.8	44	67.1	18.9	28.2
Not given	84	48.0	25.6	53.4	17	70.8	23.4	33.0
Given Someday-Checked Someday	62	43.4	22.3	51.4	21	72.9	17.0	23.3
Total	**12,426**	**58.3**	**23.3**	**39.9**	**4,483**	**80.2**	**14.6**	**18.1**

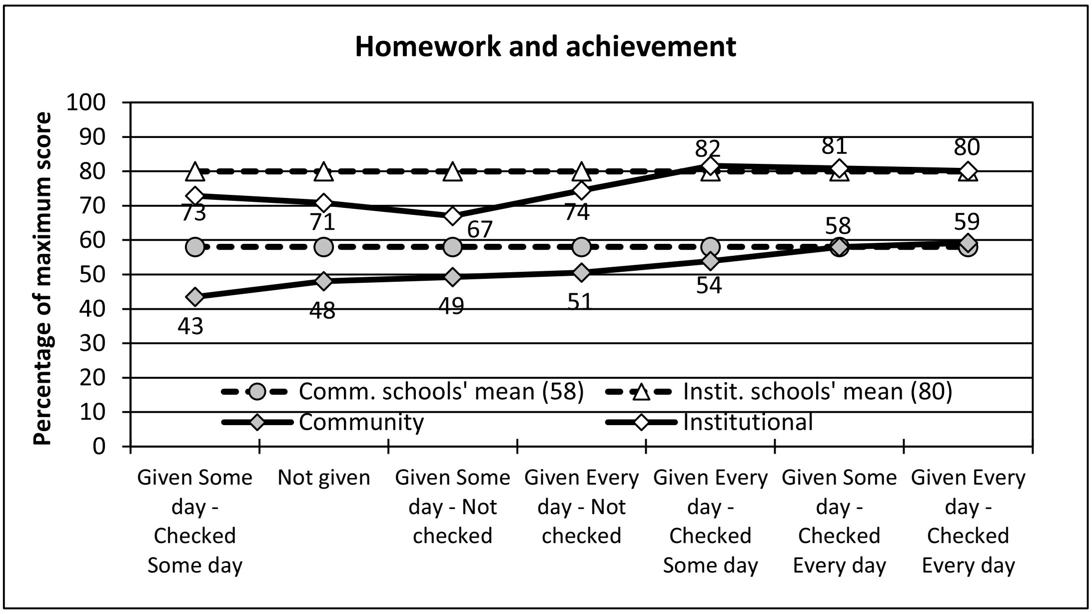

Figure 4.1.35 Relation between the homework and achievement subject

It is evident that if the students claim that the teachers do not assign them homework or the assigning and checking of homework was non-systematic, the students' achievement is notably lower compared to the students receiving and checking homework regularly. The differences are statistically significant ($p < 0.001$). However, those groups without having and getting it checked are very small; hence, the effect size is small (f = 0.10 for community schools and f = 0.12 for institutional schools); grouping explains only 1% of the variation in data (η^2 = 0.009 for community schools and η^2 = 0.014 for institutional schools).

The dataset is evident that if the teacher assigns and checks the homework regularly, the achievement is higher than without checking or assigning homework. By assigning homework daily along with its checking, even if not done every day, it contribute students raise the scores up to 16 percent. However, data shows that 2.5% of the students neither got homework nor got it checked.

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 activities of the students in the school that affects learning. In the student background information questionnaire, several student and school related activities were asked – some of which are positive and some are negative. Here, bullying is handled as one of the negative indicators and students' positive impressions of schools and teachers activities are taken as examples of positive indicators.

Negative activities - Bullying

In NASA 2012 student questionnaire, five questions indicate the varieties of bullying that are likely to happen in the school. All the questions were stemmed by the phrase "*Which of the following activities happened in your school in the last month?*" The students' responses are presented in tables 4.1.31 and 4.1.32 and depicted in figure 4.1.36. '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 reported the particular type of bullying happened within a month. Around 29% of the student mention that, in a month, something of their own was stolen, which is an alarming sign for the system.

Table 4.1.31 Frequencies of encountered bullying

Types of Bullying	No (%)	Yes (%)
I was made fun of or called names	72.5	27.5
Something of mine was stolen	71.4	28.6
I was hit or hurt by other student(s)	75.1	24.9
Fellow students kept me outside without involving in activities	74.2	25.8
I was made to do things I didn't want to do by other students	80.7	19.3

Table 4.1.32 Bullying and the achievement by the type of school

Intensity of bullying	Community				Institutional			
	N	Mean	SD	CV	N	Mean	SD	CV
No bullying	4,989	63.2	21.9	34.7	2,291	81.9	13.0	15.9
20% bullying	1,961	62.2	22.0	35.4	1,024	81.3	14.2	17.4
40% bullying	1,506	60.8	21.7	35.7	594	79.2	14.6	18.5
60% bullying	1,073	57.1	22.7	39.7	351	74.4	17.8	23.9
80% bullying	394	54.6	22.4	41.1	110	74.1	17.3	23.4
100% bullying	812	45.2	23.2	51.2	63	68.5	20.2	29.5
Total	**10,735**	**60.4**	**22.6**	**37.5**	**4,433**	**80.4**	**14.4**	**17.9**

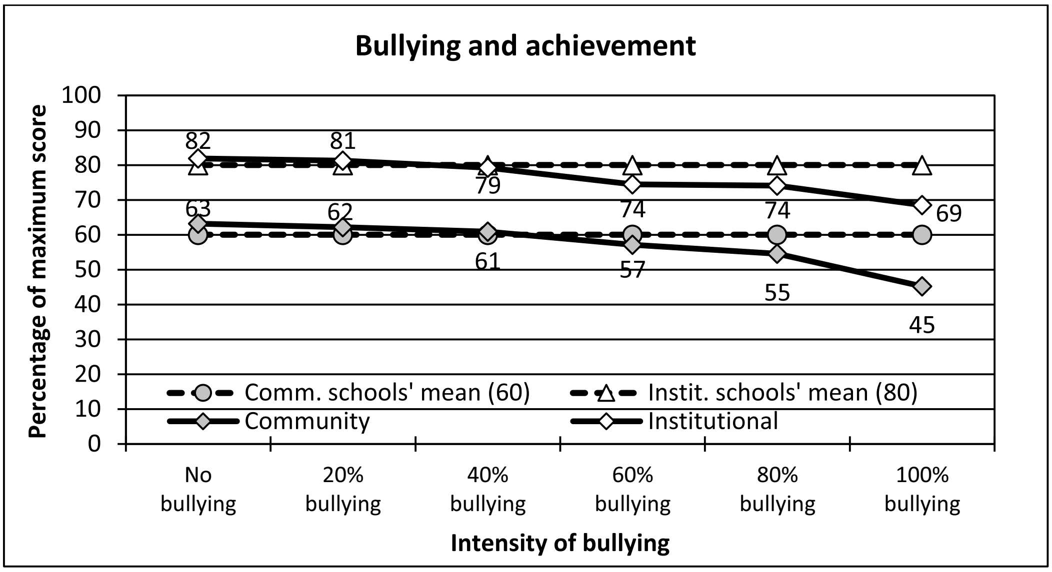

Figure 4.1.36 Relation between bullying and achievement

All five items are summed up as an indicator of bullying. Figure 4.1.36 shows the achievement of the students in each category of bullying. If only one activity of bullying is reported, it is categorized as 20% bullying, and if all five activities are reported it is categorized as 100% bullying. When knowing that 45% of the students did *not* encounter any bullying in a month, one can infer that the remaining 55% did encounter at least one type of bullying, which is a remarkable number of students. As many as 8.8% students – 10.6% in the community schools and 3.9% in the institutional schools – are experiencing at least four bullying out of five. This means, in practice, that more than 75,000 grade 3 students[41] in Nepal have been encountering physical, psychological, and social bullying every month. Learning outcomes are notably lower for the students who have encountered more than two types of bullying. Students who do not experience bullying and those who encountered extreme bullying of four or five kinds have at most 21 percent achievement gap. The difference is statistically significant (p = 0.001) though the effect size is medium (f = 0.22) in community schools and small (f = 0.19) in institutional schools. Though extreme cases of severe bullying are rare, bullying is found quite common in schools. This negative phenomenon causes needless harm to young children and has to be rooted out from the schools.

Positive activities in school

The activities that can boost the learning achievement of the students are categorized as positive activities. Such positive activities about the school were obtained from the students from two sets of questions presented in table 4.1.33. The table shows the responses of the students in all four categories; the responses are in the 4 point rating scale anchored to fully disagree (0) and fully agree (3). Generally speaking, the 3rd grade students express content with the school and student related activities in school. However, remarkably high

[41] According to the “Primary level total enrollment in all types of schools by district, Flash I_2012–2013. p.20”, there were 859,593 grade 3 students. 8.8% of these is 75,644 students.

number of students (8.9%) expressed that they feel that the teacher is not treating them fairly. The same phenomenon was also seen in 2011 datasets with grade 8 students: 11% students in Mathematics, 12% in Nepali and 13% in Social Studies (see ERO, 2013) felt – outstanding from the other questions – unfair behavior of teacher.

Table 4.1.33 Students' response towards teacher and school-related activities

Teacher and Students activities[1]	Respondents in % (valid percentage)			
	Fully agree	**Partially agree**	**Partially disagree**	**Fully disagree**
q28a I like to come and stay in school	91.4	5.6	1.3	1.7
q27a Students get along well with most teachers	87.8	9.0	1.7	1.5
q28c Teachers in the school care about the students	87.6	7.3	2.8	2.4
q27b Most teachers are interested in student's well-being	86.9	8.7	2.2	2.1
q27d If I need extra help, I will receive it from my teacher	86.2	8.7	2.5	2.6
q27c Most of the teachers really listen to what I have to say	84.5	9.4	3.5	2.6
q28b Students in my school like me	80.6	14.8	2.6	2.0
q27e Most of my teachers treat me fairly	79.5	11.6	3.5	5.4
Average	**72.5**	**8.0**	**2.1**	**2.1**

1) *The activities are ordered on the basis of percentage in "Fully agree."*

Further analysis is carried out by merging the variables into two categories (2–3 = 1, that is "agree", and 0–1 = 0, that is "disagree"). Furthermore, the sum of eight indicators is converted into the percentage of maximum score to analyze the level of positive activities and its relation to achievement.

DTA finds four attitude groups in the indicator. These boundaries and descriptive statistics are seen in tables 4.1.34 and 4.1.35 and illustrated in figure 4.1.37. The overall result is that the feeling of the positive actions in the school relates positively with student achievement. The correlation between the sum of positive activities and achievement is (r = 0.18), which is statistically significant ($p < 0.001$) and moderately high (d = 0.41).

Table 4.1.34 Teacher and school related activities and the achievement

Percentage of positive actions	N	Mean	SD	CV
62.5% or lower	1,772	51.9	24.4	47.1
75.0%	826	59.4	22.6	38.1
87.5%	1,877	63.0	23.7	37.6
100%	12,625	66.3	22.6	34.1
Total[1]	**17,100**	**64.1**	**23.4**	**36.5**

1) *Total includes also the cases without giving their opinion (missing n = 2,401)*

Table 4.1.35 Teacher and school related activities and the achievement by school type

Percentage of positive actions	Community				Institutional			
	N	Mean	SD	CV	N	Mean	SD	CV
62.5% or lower	1,490	48.3	23.9	49.4	282	71.3	17.4	24.5
75.0%	630	54.1	22.2	41.1	196	76.4	14.0	18.3
87.5%	1,298	56.0	23.7	42.3	579	78.8	14.5	18.4
100%	9,177	60.6	22.6	37.3	3,448	81.4	14.1	17.3
Total[1]	**12,595**	**58.4**	**23.2**	**39.8**	**4,505**	**80.2**	**14.6**	**18.2**

1) *Total includes also the cases without giving their opinion (missing n = 2,401)*

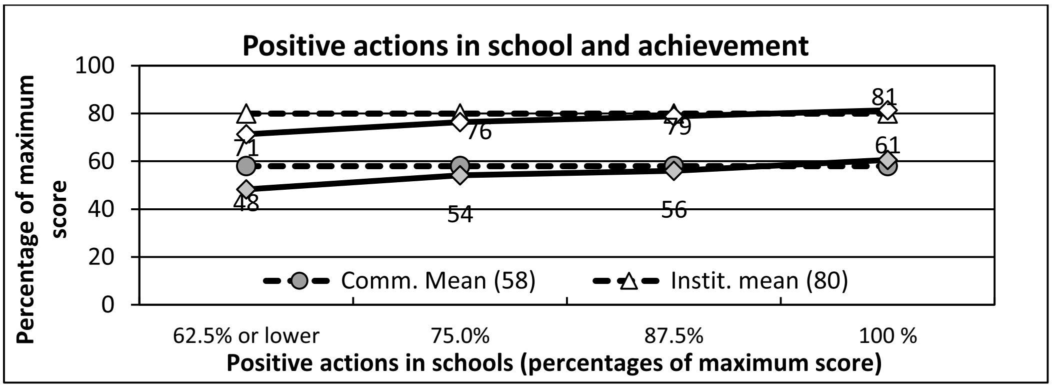

Figure 4.1.37 Relation between positive actions in school and achievement

The data shows that there is a positive relation between the students' positive feeling towards the teacher and school related activities and the achievement. The increase in achievement is directly proportional to the increase in intensity of such activities. After dividing the indicator into four groups on the basis of DTA, the differences between the groups are statistically significant ($p < 0.001$), however, the effect size is moderate (f = 0.18 in both the community and institutional schools). The difference between the most positive group and the most negative group is notable (10 –13 percent). Only when the students are extremely positive towards school and teachers' behavior, the learning achievement is higher than the average. Students with a negative feeling against five or more of the eight indicators (62.5%) are in great danger of achieving much lower than the average in Nepali.

The dataset shows that a high number of the students (55%) have encountered bullying in school within a month and 8.8% of students are experiencing a severe kind of bullying. This means that more than 75,000 grade 3 students in Nepal have been encountering physical, psychological, and social bullying every month. The phenomenon has also been affecting the learning outcomes in almost all the groups of the students who felt bullying; so all possible efforts have to be made to root the phenomenon out from schools.

The dataset also indicates that when the students feel that the actions of the teachers and the schools are ultimately good, the results are better than average in Nepali (61 in community schools and 81 in institutional schools). At the same time when students feel extremely negative, the results are far below the average. Of the total samples, 8.9% students feel that their teachers do not treat them fairly.

4.1.4 Synthesis of the Analysis

Several individual student related and geographically related factors have been analysed, which individually explains the difference in achievement between the students. These factors are collected in table 4.1.36. It is notable that, except for gender, all the factors showed statistically significant difference between the groups when analysed individually.

Table 4.1.36 Individual variables handled within the text and their effect in one-way ANOVA

Variable and values[1]	**Leverage**[2]	**Eta squared**[3]	**Effect size**[4]
Ecological zone (1 = Mountain, 2 = Hill, 3 = Tarai, 4 = Valley)	+23.69	0.116	0.36
Development Region (1= Eastern…5 = Far-Western, 6 = Valley)	+26.54	0.141	0.41
School location (0 = Rural, 1= Urban)	+15.71	0.074	0.28
School type (0 = Community,1= Institutional)	+22.85	0.168	0.45
Gender (0 = girls, 1= boys)	+1.50	0.001	0.03
Ethinicity/caste (1 = Janjati, 2 = Dalit, 3 = Madhesi, 4 = Brahman, 5 = Chhetri)	+16.18	0.043	0.21
Language at home (1 = Nepali, …, 12 = Other)	+36.98	0.049	0.23
Mother's Education (1= Illiterate, …, 8 = Above MA)	+22.61	0.034	0.19
Father’s Education (1= Illiterate,…, 8 = Above MA)	+23.98	0.043	0.21
Mother's Occupation(1= working abroad.., 8 = working at other home)	+20.82	0.037	0.20
Father’s Occupation(1= working abroad., 8 = working at other home)	+17.95	0.085	0.30
Home possessions (sum; max 11)	+24.80	0.079	0.29
Home accessories (sum; max 3)	+20.56	0.092	0.32
SES (sum max 7)	+30.68	0.144	0.41
I do jobs at home (1 = not at all, …, 4 = more than 4 hours)	+7.87	0.015	0.12
I work at a paid job (1 = not at all, …, 4 = more than 4 hours)	+8.25	0.021	0.15
Attitude Utility in Nepali (sum max 15)	+19.20	0.029	0.17
Age	+8.48	0.007	0.08
Who helped you …? (1 = Father, …, 6 = Teacher	+7.46	0.014	0.12
Do you have textbook of Math subject (0 = no, 1 = Yes)	+10.53	0.009	0.10
Homework (0 = not given.., 6 = Given everyday, checked everyday)	+15.23	0.009	0.10
Bullying (sum; max 5)	+21.28	0.053	0.24
Positive Activities in school (sum; max 8)	+17.64	0.038	0.20

1) *The order of the variables is the same as handled in the Sections above*
2) *Difference between the lowest and highest group-mean*
3) *On the basis of one-way ANOVA 4)Cohen’s f*

On the basis of univariate ANOVA, school type, closely followed by the Development region and socio-economic status, are found to be the most effective single factors in affecting the achievement level of the student as the effect sizes are $f = 0.45$, $f = 0.41$, and $f = 0.41$ respectively. Some of the variables in table 4.1.37 may be strongly related to each other and hence they may not add value in explaining why some students are performing much better than others. In what follows, the synthesis of the analysis is done in two ways and presented below: All the variables are presented as a result of Multilevel Modelling; and statistically best factors are collected by using the Regression Modelling. For the analysis, grouping factors are changed to be so called Dummy variables when needed; for example, Ecological zone is transformed into three variables: variables indicating for Mountain, for Hill, and for Tarai.

Modelling the overall achievement by Multilevel Modelling

The datasets collected from schools are always clustered, that is, the students within the school are more alike with each other in comparison to the case that the same amount of students would have been sampled totally from the population. Multilevel modeling is used to acquire the correct test values while taking into account the clustering effect of the school. Table 4.1.37 shows the corrected estimates for the variables while modelling the phenomenon in a multivariate manner; by using the multivariate ANOVA, the hidden commonalities of the factors are revealed.

When taking into account the clustered structure in the dataset and the conjoint effect of the factors, quite many of the factors do not show main effect. Such variables are living in a Hill zone, and Central, Mid-Western or Far-Western region, school location, home language, and none of the options for given help for studies. These factors could be omitted from the model explaining the differences in achievement in Nepali among the grade 3 students.

Table 4.1.37 Individual variables and their effect in Multi-level analysis

Source[1]	df_1	df_2	F	Sig.
Intercept	1	1884.7	421.3	<0.001
Eco zone Mountain Dummy (Mountain = 1, other = 0)	1	759.9	7.74	0.006
Ecol zone Hill Dummy (Hill = 1, other = 0)	1	705.0	0.21	0.650
Dev region Central Dummy (Central = 1, other = 0)	1	700.0	2.53	0.112
Dev region Western Dummy (Western = 1, other = 0)	1	680.6	8.75	0.003
Dev region Mid-Western Dummy (Mid-Western = 1, other = 0)	1	700.4	2.48	0.116
Dev region Far-Western Dummy (Far-Western = 1,other = 0)	1	734.0	0.51	0.474
Dev region Valley Dummy (Valley = 1, other = 0)	1	679.5	30.65	<0.001
School location (0 = Rural, 1 = Urban)	1	627.8	0.1	0.702
School type (0 = Community, 1 = Institutional)	1	648.0	58.31	<0.001
Gender (0 = girls, 1 = boys)	1	8337.9	14.40	<0.001
Caste Brahman & Cheetri Dummy (Brahman & Chhetri , other = 0)	1	8578.8	43.99	<0.001
Caste Janjati Dummy (Janjati = 1, other = 0)	1	8631.9	6.28	0.012
Caste Madhesi Dummy (Madhesi = 1, other = 0)	1	8740.7	4.44	0.035
Caste Dalit Dummy (Dalit = 1, other = 0)	1	8582.7	2.29	0.130
Language Dummy (Nepali = 1, other = 0)	1	8792.4	0.00	0.955
Homework Dummy 1or 2h (1 – 2 hours = 1, other = 0)	1	8545.0	46.47	<0.001
Paid work Dummy (0 hours = 1, other = 0)	1	8716.9	20.46	<0.001
Attitude "Utility in Nepali" (Sum, max 15)	15	8551.9	1.74	0.038
Age Dummy 11 to 12y (11 – 12 years = 1, other = 0)	1	8407.3	7.37	0.007
Help by Father Dummy (Father = 1, other = 0)	1	8353.6	0.01	0.939
Help by Mother Dummy (Mother = 1, other = 0)	1	8323.9	0.15	0.703
Help by Brother & Sister Dummy (Brother/Sister = 1, other = 0)	1	8326.8	0.08	0.777
Help by Tuition Dummy (Tuition = 1, other = 0)	1	8361.4	0.14	0.705
Help by Teacher Dummy (teacher = 1, other = 0)	1	8411.4	1.69	0.193
Do you have a textbook in Nepali (Yes = 1, No = 0)	1	8444.3	20.23	<0.001
Homeworks Not Given Dummy (Not given = 1, other = 0)	1	8491.4	21.94	<0.001
Bullying (Sum, max 5)	5	8512.5	46.08	<0.001
Positive Activities in school (Sum, max 8)	8	8404.9	6.15	<0.001
SES[2] (Sum, max 6)	6	8389.5	8.33	<0.001

1) In Ecological zone, developmental region, ethnicity/caste, and help given for studies, one of the classes needs to be omitted in the analysis because of singularity reasons. Tarai zone, Eastern region, "other" ethnic/caste, and "No one helps" are omitted; there showed no statistical significance in regression analysis. 2)Shortened SES; school type is taken away; this enables estimating the parameters for school type.

Statistically the best factors by using Regression Modelling

Traditional linear regression analysis with stepwise regression is used to explain the total score by the same variables, which are described above. Table 4.1.38 shows the results.

Table 4.1.38 Statistically the best model of linear regression analysis explaining student achievement (Method: stepwise)

Model	Coefficients				
	Unstandardized Coefficients		Standardized Coefficients		
	B	Std.Error	Beta	T	Sig.
(Constant)	32.32	1.999		16.17	<0.001
School type (0=Community,1=Institutional)	10.73	0.529	0.25	20.27	<0.001
Bullying (Sum, max 5)	-2.56	0.139	-0.17	-18.38	<0.001
Dev region Valley Dummy (Valley = 1, other=0)	11.58	0.600	0.23	19.30	<0.001
Positive Activities in school (Sum, max 8)	0.10	0.013	0.07	7.72	<0.001
Homework Dummy 1or 2h (1 –2 hours =1, other = 0)	3.72	0.441	0.08	8.44	<0.001
Caste Brahman & Chhetri Dummy	3.63	0.408	0.08	8.90	<0.001
Dev region Western Dummy(Western= 1, other=0)	5.22	0.555	0.10	9.41	<0.001
Dev region Central Dummy(Central=1,other=0)	3.88	0.536	0.07	7.24	<0.001
Homeworks Given and Checked everyday Dummy	7.36	1.197	0.06	6.15	<0.001
SES2 (Sum, max 6)	0.94	0.158	0.07	5.95	<0.001
Do you have a textbook in Math(Yes =1, No=0)	5.66	1.070	0.05	5.29	<0.001
Paid work Dummy (0 hours=1, other=0)	1.63	0.420	0.04	3.87	<0.001
Attitude ”Utility in Nepali” (Sum, max 15)	0.18	0.061	0.03	2.89	0.004
Gender(0=girls,1=boys)	1.11	0.378	0.03	2.93	0.003
Ecol zone Mountain Dummy(Mountain = 1, other=0)	1.98	0.767	0.02	2.58	0.01
Help by Tuition Dummy (Tuition = 1, other = 0)	1.85	0.743	0.02	2.49	0.013
Help by Teacher Dummy (teacher = 1, other = 0)	1.20	0.589	0.02	2.04	0.041

The model in table 4.1.38 can be interpreted as follows: The average mean of the students is 32.3% of the maximum score which implies that the student was in the lowest group in all the factors. If the school was an institutional one (value = 1), the student's score was, on average, + 10.7 percent higher (note the sign of the coefficient). Additionally, if the student came from the Valley, the additional score was +11.6 percent. Those who were assigned homework and got it checked regularly gained +7.4 percent more. Similarly, in the case of those having no homework or only two or less checked per week, the score was 5.7 percent higher. On the other hand, the student who faced bullying of five types, the achievement level dropped by 2.6 percent; the difference between the lowest and highest group is 5 × 2.56 = 12.8 percent.

4.2 Assessment Results in Nepali for Grade 5

This section analyses the assessment results of grade 5 Nepali. It starts with analyzing basic results including the overall distribution of scores, results in the different content areas of Nepali in general and goes to the analysis of the effects of different diversity factors from equality point of view. It then analyses the influences of factors explaining the differences in the achievement in Nepali.

4.2.1 Basic Results of Assessment in Nepali for Grade 5

As the basic results of assessment in Nepali for grade five, this sub-section analyses the overall distribution of scores, result in the various content areas and various levels of cognitive domains in Nepali subject, results variations in item types, and comparison of results with previous assessments and international assessment results.

Distribution of overall scores

The grade 5 Nepali sample was big enough to form a normal distribution (13,971 students). However, the distribution of this population is not normal (see fig. 4.2.1), which means that there are several populations embedded in the sample.

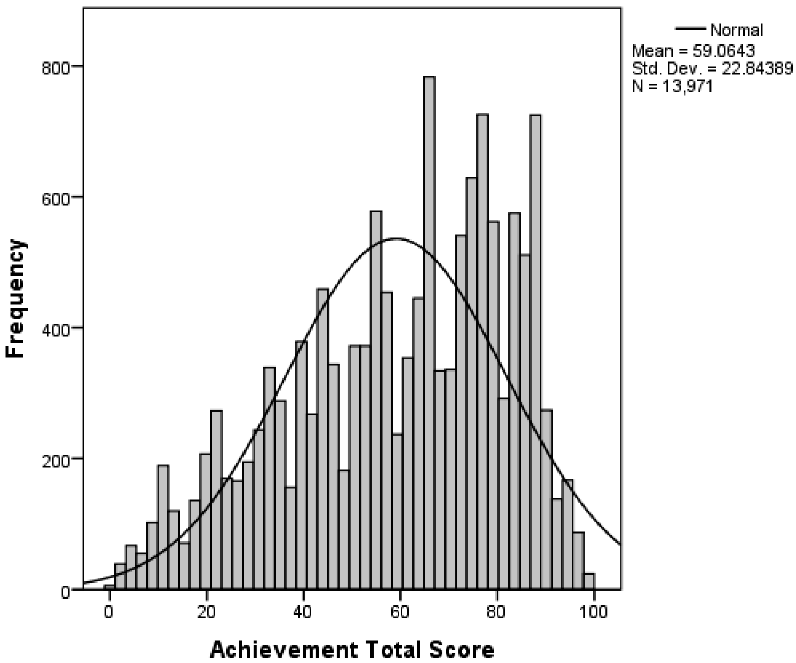

Figure 4.2.1 Non-Normally distributed population in Nepali

In the total score of Nepali, the majority of students' population is found lying slightly in the high-performing part of the distribution. Compared to grade 3 Mathematics and Nepali and grade 5 English, the distribution looks very similar. A closer look to the distributions shows that there are two normal populations with long tail in the dataset: students from the community schools and students from the institutional schools (figure 4.2.2).

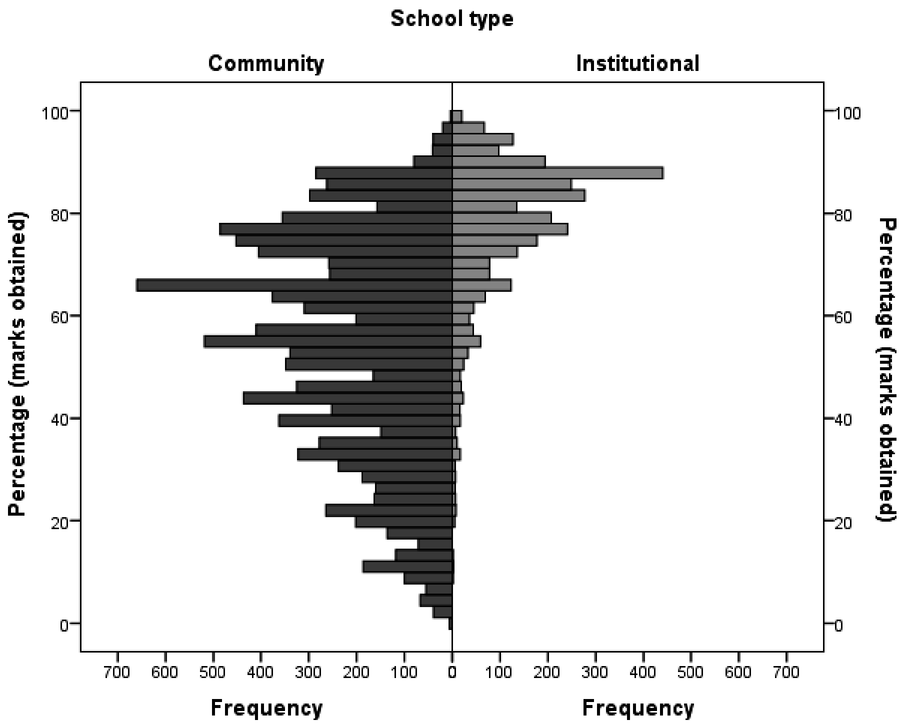

Figure 4.2.2 Distribution of students' means in community and institutional schools

In figure 4.2.2, the left hand side distribution shows that of community school students and the right hand side that of institutional school students. The whole system in the right hand is shifted to the better performing level in grade 5 Nepali. There are students in community schools getting equally high marks as in institutional schools. Figure 4.2.2 clearly shows that students in community schools are varying from the low performers to the highest performers whereas most of the students in institutional schools are the higher performers, though there is also a long tail of the low-performing students in institutional schools. Compared to the English dataset of grade 5, in grade 5 Nepali there are high number of low-performing students in institutional schools. The distributions in both community and institutional schools can be taken normal enough for the parametric statistical analyses.[42]

Another related fact is that the schools are also clearly divided into two "populations": the high-performing and the low-performing – both populations are normally distributed. On the basis of the school mean of the student performance, there are two categories of schools in which the difference between the populations is remarkable.

By analyzing using scatter plot and combining the socio-economic status (SES) with the average achievement in the school, figure 4.1.3 shows that two types of schools (community school in circle and institutional schools in triangle) fall into two groups: most of the institutional schools are performing very well but the community schools vary from very high performing to very low-performing ones.

[42] Most of the classical methods for statistical analysis assume the normal distribution of the population and in the sample.

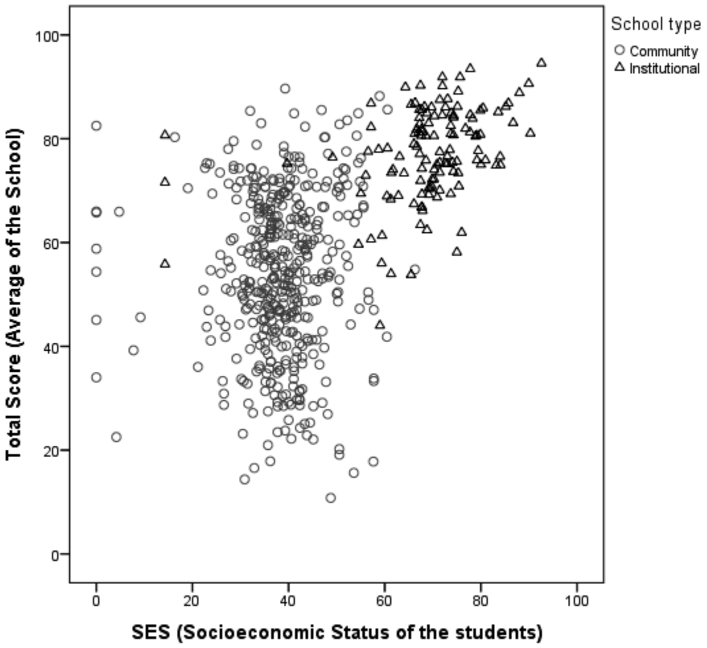

Figure 4.2.3 Achievement and Social economic status with the type of schools

The dataset is evident that the grade 5 Nepali population is not normally distributed. There are two groups of student populations: low and high performing students from the community schools and (mainly) higher performing students from institutional schools with a long tail of low performing students. The variance between the community schools is remarkable.

Different content areas and Achievement

The whole Nepali test was a combination of four content areas: 1) Reading, 2) Writing, 3) Grammar, and 4) Vocabulary. The maximum marks of Reading and Writing were more or less proportionally equal to the weightage given by the curriculum. Grammar is mentioned as the "functional grammar" in which Vocabulary is also integrated, so it gets very low weightage. In this test too low weightage is given to Grammar and Vocabulary; however, their weightage are more than that given in the curriculum. In order to compare the achievement in all the contents, the sub-scores are converted into the percentage of the maximum score of the content area. Figure 4.2.4 shows the students' achievement in Nepali as a whole and the achievement level in four content areas.

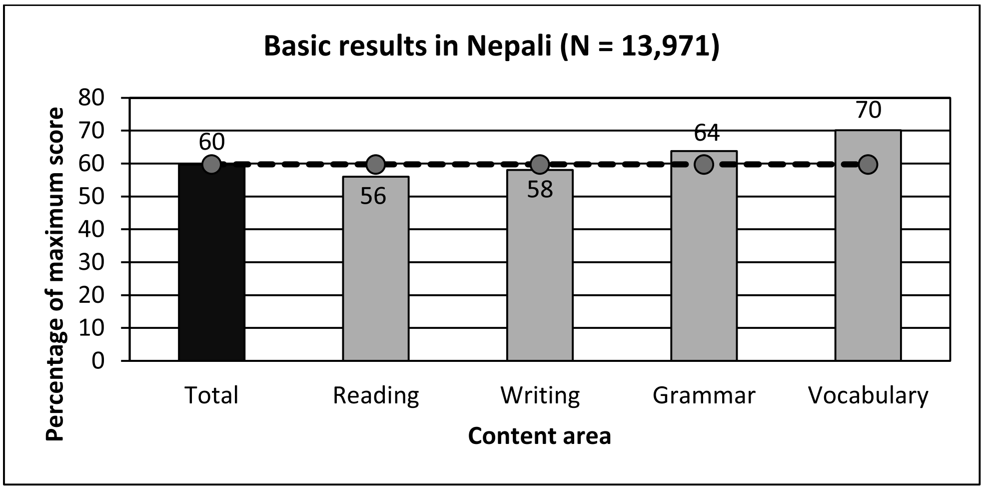

Figure 4.2.4 Achievement in various content areas

The percentage of achievement score shows that the national average of Nepali is 60. Of the different content areas, students are the weakest in Reading (56). They perform better than average in Vocabulary (70), Grammar (64), and nearer to average in Writing (58).

As there are differences between the community and institutional schools in average scores, it is noteworthy to compare whether there are proportional differences in the content areas between the students. Table 4.2.1 shows the comparisons.

Table 4.2.1 Achievement in various content areas in grade 5 Nepali by types of schools

Community schools (N = 10,842)				**Institutional schools (N = 3,129)**		
Content area	Mean	SD	CV	Mean	SD	CV
Reading	50.4	24.3	48.1	75.5	17.5	23.1
Writing	52.8	23.8	45.0	76.3	16.0	21.0
Grammar	59.0	25.4	43.0	80.3	16.5	20.6
Vocabulary	65.6	26.0	39.6	85.4	15.9	18.6
Total[1]	**54.4**	**22.3**	**41.0**	**78.2**	**14.7**	**18.8**

1) *Note that the total score is not the mean of the content areas because it has been equated independently from the content areas.*

Although the achievement in the community schools is lower than in the institutional schools in all the content areas, it is evident that the differences between the highest and the lowest score by content areas are wider in the community schools in comparison to the institutional schools. While the difference between the lowest scored content area is 16 percent in the community schools, it is 10 percent in the institutional schools. In all content areas, the gaps are wider in community schools than institutional schools. Partly this can be explained by ceiling effect in the institutional schools as the test was too easy for the students in the institutional schools. Hence, the best students were not able to show how far they could have been able to increase score.

The dataset clearly shows that the learning outcomes are the weakest in the content areas of Reading (56%) and the highest in Vocabulary (70%). The differences between the content areas are wider in the community schools than in the institutional schools. This can be caused by the ceiling effect since the test was too easy for the students in the institutional schools.

Achievement in various cognitive domains

The entire Nepali test was constructed based on Bloom's taxonomy of hierarchical level of cognitive domains (Bloom *et al.,* 1956; Metfesser, Michael & Kirsner, 1969) that is, *knowledge*, *comprehension*, *application*, and *higher ability* (reasoning/problem solving). The achievement of the students by the hierarchical levels is shown in figure 4.2.5.

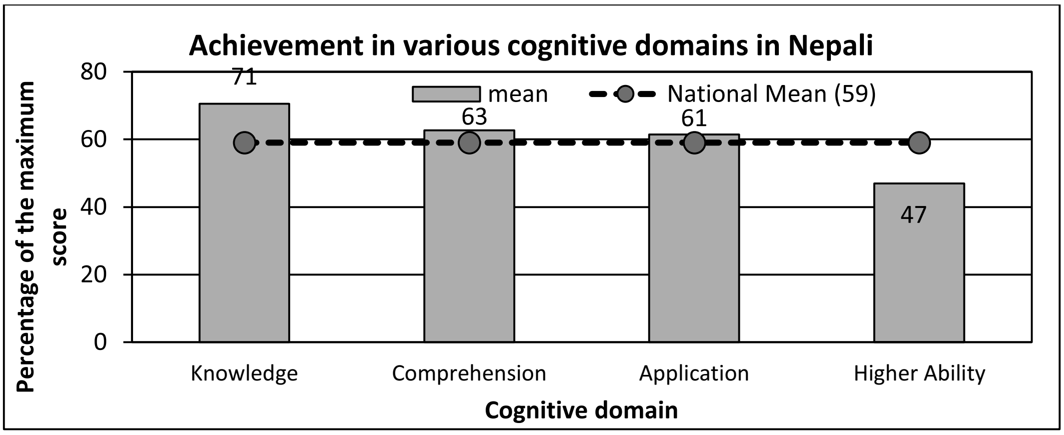

Figure 4.2.5 Achievement in various levels of cognitive domain

It is notable that the achievement levels are above the average in all the hierarchical levels except in the higher ability. Remarkably high number of students were able to solve less than 15 percent or less of maximum score in the practical problems, that is, the application type of items (7.4% of the students). Around 50% of the students have gained just one or two points out of 11 requiring the higher cognitive abilities and 18% of them did not solve any of the higher level cognitive tasks.

Because of the difference in the average level between the community and institutional schools, it is worth to know whether there is proportional difference in the hierarchical level between the students. Table 4.2.2 illustrates the differences.

Table 4.2.2 Achievement in various content areas by school type

Hierarchical level	Community schools (N = 10,842)			Institutional schools (N = 3,129)		
	Mean	SD	CV	Mean	SD	CV
Knowledge	66.4	27.1	40.8	84.8	18.6	21.9
Comprehension	57.8	22.5	39.0	79.5	13.8	17.3
Application	56.0	26.0	46.3	80.1	18.1	22.6
Higher Ability	41.7	23.9	57.3	65.6	19.9	30.3
Total	**54.4**	**22.3**	**41.0**	**78.2**	**14.7**	**18.8**

The main trend is obvious in both the community and institutional schools: in both cases the students are much better in recall type of questions than in the tasks requiring higher skills. However, there is another tendency worth mentioning. In the institutional schools, the difference between scores of knowledge type and higher ability type of questions is much smaller (19 percent) than in the community schools (25 percent). Moreover, the difference between the school types increases systematically as the more complex problems the students need to solve (18 – 22 – 24 – 24).[43] Nonetheless, the differences are not as high as found in English subject (31 – 33 – 34 – 45). Statistically, the tendency is seen in the effect size though the differences are remarkable in any case, the difference is notably smaller in the area of knowledge ($d = 0.72$) than in comprehension ($d = 1.04$) and higher ability ($d = 1.04$). This means that, for one reason or another, the students in institutional schools are more able to solve complex problems in relation to the simple tasks than their peers in community schools.

One more point is notable regarding table 4.2.2 that the exceptionally high value of Coefficient of Variation within the community schools is in higher ability (57.3). The moderate standard deviation (23.9) with very low mean (41.7) indicates that, within the community schools, there are also reasonably highly performing students.[44]

The dataset shows that the students' ability to solve complex problems is quite low as only 47% of the maximum scores of tasks requiring higher ability were reached. Students are much better in the recalling type of questions (71%). Remarkable number of students (18%) was not able to solve any of the tasks requiring higher ability. The students in the institutional schools are more able to solve complex problems than their peers in the community schools.

Type of item and achievement

Nepali test comprises both objective and subjective types of items. Objective items covered a wide range of content areas and were very specific to judge because there was only one correct answer for a question or one explicit piece of information was required to get a correct answer. There were also subjective items on each test version requiring a longer procedure to get the full marks. Both the objective and subjective types of items were made in all the hierarchical levels (Knowledge, Comprehension, Application, and Higher Ability) and all the difficulty levels, though the subjectively scored items tend to be more demanding because of the higher demand of cognitive ability. Tables 4.2.3 and 4.2.4 comprises the basic statistics of the item type-wise achievement levels.

---

43 Similar type of tendency is there when analyzing only the students from the Kathmandu Valley: the differences are 6 – 10 – 10 – 11 indicating, first, the differences are notably milder between the community and institutional schools; second, the difference is notably higher in the area of higher skills (11), and third, the students in the community schools in the Valley seems to be better than average in the comprehension and application type of tasks (with difference only by 10 percent in each).

44 In the community schools of the Kathmandu Valley, the mean of higher ability items is 71.9 (CV = 25.0). Hence, the CV is not exceptionally high compared to the CVs of the other cognitive levels in the total sample.

Table 4.2.3 Item type-wise mean scores

Type of items	N	Mean	SD	Minimum	Maximum
Objective	13,971	65.5	22.1	0	100
Subjective	13,971	55.1	25.1	0	100

Table 4.2.4 Item type-wise mean scores by school type

Type of items	Community schools (N=10,842)			Institutional schools (N=3,129)		
	Mean	SD[1]	CV[2]	Mean	SD	CV
Objective	61.1	21.9	35.9	80.9	14.1	17.4
Subjective	49.3	24.3	49.3	75.0	16.0	21.3
Total	**54.4**	**22.3**	**41.0**	**78.2**	**14.7**	**18.8**

It is obvious that the subjectively scored tasks – usually those with more demanding requirements for the correct answer – are solved much lower (55%) than the objective items (65%). Most of the objective items were knowledge, comprehension and application type whereas subjective items were application and higher ability type. Though the differences between the community and institutional schools are wide in any case, the effect size is somehow higher when it comes to solve the subjective type of items (d = 1.13) compared to the objective type of items (d = 0.97).

The dataset suggests that the students are performing well in recognizing the correct answer and in recalling simple facts from the texts, fundamental thinking, the basic interpretation of paragraph, table and chart, and a few steps of logical thinking. They are much weaker in producing fluent texts or letters, or preparing synthesis and abstracts from a text. In many cases, the students resorted to doing the open ended tasks (like free writing, problem solving and analysis) but the skills were not high enough for achieving the highest marks.

Comparison of achievement in NASA to previous datasets

The National Assessments carried out in various years aim to assess the achievement and the progress over a period of the years. The datasets of previous Nepali assessment are, however, somehow fragmented and achieved by using various strategies for sampling which makes the comparison difficult.[45] The previous datasets also carry two other challenges hindering the comparison with the present dataset. First, three studies conducted earlier by different institutions report different results of average scores. The study organized by Basic and Primary Education Project (BPEP, 1998) shows the average score as round 52 and the assessment of achievement conducted by Primary Education Development Project (PEDP) reports the average score as round 46. Similary, the assessment conducted by Research Center for Education Innovation and Development

[45] For example, BPEP and PEDP (1998) were sampled in the project areas representing the whole country. CERID (1998) was sampled exclusively within the six districts covering all five Development regions. EDSC (1999) was sampled covering five Development regions, three Ecological zones and valley as the separate cluster and FBC (2008) was sampled in 18 districts covering all development regions and Ecological zones. This deviance in samplings makes the comparison somehow difficult.

(CERID) for the grade 5 students (CERID, 1998, p 18) in Nepali shows that the national average of the students was round 46. Further, the National Assessment of grade 5 students carried out by the Educational Development Service Centre (EDSC, 1999) shows that the national average of the students was round 52%. The CERSOD (2001) study has also reported round 52% average score. Later, in 2008, National Assessment of grade 5 students conducted by Full Bright Consultancy and CHIRAG showed average score (45%) lower than previous studies. These figures are coming from Classical Test Theory (CTT) and, unfortunately, they are not comparable with each other because of the lack of a proper linking procedure. The differences between the scores can easily be explained by the different difficulty levels of the tests. Second, the previous datasets of grade 5 are not 185available; and, hence, any IRT modeling based procedures for comparison could not be made.

Though the comparison cannot be made in the absolute sense, proportional comparisons can be made on the basis of the previous results. The proportional differences are presented in tables 4.2.5 to 4.2.8.

Table 4.2.5 Comparison of achievements of 2008 and 2012 in Nepali by gender

Indicators	2008 (FBC, 2008)		NASA 2012	
	Boys	**Girls**	**Boys**	**Girls**
Mean	44.3	45.8	59.0	60.7
SD	20.3	20.8	23.0	23.1
CV	45.8	45.4	39.0	38.0
N	1,503	1,695	6,449	6,977

Comparing the data of 2008 with that of 2012, differences between the boys and girls have not changed significantly within their groups; girls still out-perform boys slightly (less than 2 percent). The lowered Coefficient of Variation (from 45–46% to 38–39%) shows that the distributions of boys and girls have changed moderately over the five years; achievement level has raised more than the standard deviation. This is a good sign from the equality point of view. The difference between boys and girls is not remarkable in both years ($d < 0.07$).

Table 4.2.6 Comparison of achievements of 2008 and 2012 by Ecological zones

Indicators	2008 (FBC, 2008)				NASA 2012			
	Mountain	**Hill**	**Tarai**	**Valley**	**Mountain**	**Hill**	**Tarai**	**Valley**
Mean	44.7	43.8	40.5	62.5	61.1	56.6	52.8	77.5
SD	18.2	21.5	17.3	21.1	21.4	21.9	23.4	16.2
CV	40.7	49.1	42.7	33.8	35.0	39.3	44.2	20.8
N	456	995	1319	429	1407	6209	3857	2497

Compared with the 2008 dataset, the differences between the Ecological zones is found to be widened moderately. The difference between the students' achievement in the Mountain and Hill zones has also increased from 0.9 to 4.4 percent points and between Mountain and Tarai from 4.2 to 8.2 percent. The Coefficient of Variation (CV) shows that the distribution of students from the Mountain area has changed less (from 41 to 35). In Hill and Valley areas CV has lowered remarkably (from 49 to 39 in Hill area and from 34 to 21 in Valley) while it has increased slightly in Tarai (from 43 to 44). In both years, the differences between the zones are not remarkable (the pair-wise values for *d* < 0.25).

Table 4.2.7 Comparison of achievement of 2008 and 2012 by Development region

Indicators	2008 (FBC, 2008)					NASA 2012[1]				
	Eastern	Central	Western	Mid-Western	Far-Western	Eastern	Central	Western	Mid-Western	Far-Western
Mean	42.5	49.8	44.1	47.7	37.7	51.7	64.7	62.7	52.3	57.9
SD	21.7	21.7	18.7	18.9	17.4	24.4	22.6	21.2	21.8	21.3
CV	51.1	43.6	42.4	39.6	46.2	47.1	35.3	33.8	41.8	36.8
N	580	1052	537	467	563	2207	5603	2185	1702	2274

1) Students from the Kathmandu Valley are included in the Central Developmental region

The Kathmandu Valley was included in sampling in both the year 2008 and 2012 having considered it as the separate strata. For the comparison, the Valley has been merged with the Central developmental region in order to make the comparison possible. Otherwise, the samples are not necessarily comparable because of the differences in sample size and coverage. The student performance in the Eastern region was remarkably low compared to the Central region in both the years 2008 and 2012. The effect size shows a negligible difference (*d* = 0.15) in 2008 but in 2012 it shows a moderate difference (*d* = 0.42) between the regions. The students from the Mid-Western region have stayed in the low position, in the Far-Western region they have raised their performance closer to the average and in the Eastern region they have got further lower down from the mean. From equality point of view, the Far-Western region is a positive one: the students in this region are seen to have gained the same results as in the Central and Western regions. However, the decline in scores within the students from the Eastern region is not a good sign.

Table 4.2.8 Comparison of achievements of 2008 and 2012 by school type

Indicator	2008 (FBC, 2008)		NASA 2012	
	Community	Institutional	Community	Institutional
Mean	43.6	63.8	54.4	78.2
SD	19.7	22	22.3	14.7
CV	45.2	34.5	41.0	18.8
N	2,961	238	10,842	3,129

It is also seen that the difference between the community and institutional schools has not reduced from 2008 to 2012. The difference was 20 percent in 2008 whereas it is 24 percent in 2012. The difference also has not changed much in the effect size from *d* = 1.02 to *d* = 1.14. Though the variance has reduced in community schools (CV has reduced from 45 to 41), it has not reduced as remarkably as in institutional schools (CV has reduced from 35 to 19). It indicates that the inequalities in community schools are higher than the institutional schools. Table 4.2.9 summarise the information provided in above tables.

Table 4.2.9 Situation in 2012 in comparison to the previous datasets

Selected background variables					
	Gender	**Ecological zone**	**Development region**	**School location**	**School type**
Main finding	No change in difference; girls still over-perform boys slightly.	Differences have increased moderately. Students in the Mountain zone score higher while in Tarai they score lower.	Students in the Far-Western region score higher while in the Eastern region they score lower.	Remarkable rise is noticed in performance within the Urban schools	No remarkable change is seen in the difference between the community and institutional schools.

The dataset indicates that, compared to the 2008 results, there is no change in difference between boys and girls, the students in the Mountain zone and Western region score higher, and the students in urban schools score remarkably higher in 2012. The Far-Western region is found to have progressed remarkably, whereas the Eastern region and Tarai zone still perform lower. The gap between rural and urban schools has widened remarkably within five years; the change in the phenomenon is remarkable. Dataset also indicates that the performance in the community schools has risen remarkably but still the gap between community and institutional schools is high.

Comparison with international achievement results

The NASA 2012 was made comparable with the international PIRLS reading assessment. Six of the released PIRLS items 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. It means that, when a student's ability level in NASA 2012 is zero, it corresponds to the average level of the international students of grade 4. Though the international test is in the English language, the reading passage and items were translated into Nepali. Here, it is to note that the items were selected from the item bank of grade 4 students, not grade 5, so the grade 5 in Nepali has also been compared with grade 4 standards of PIRLS.

Figure 4.2.6 shows the comparison of the students' achievement with the international standard. In the figure, x-axis shows the content areas of Nepali 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 go down whereas when the ability is above the international average, the bars would have gone up.

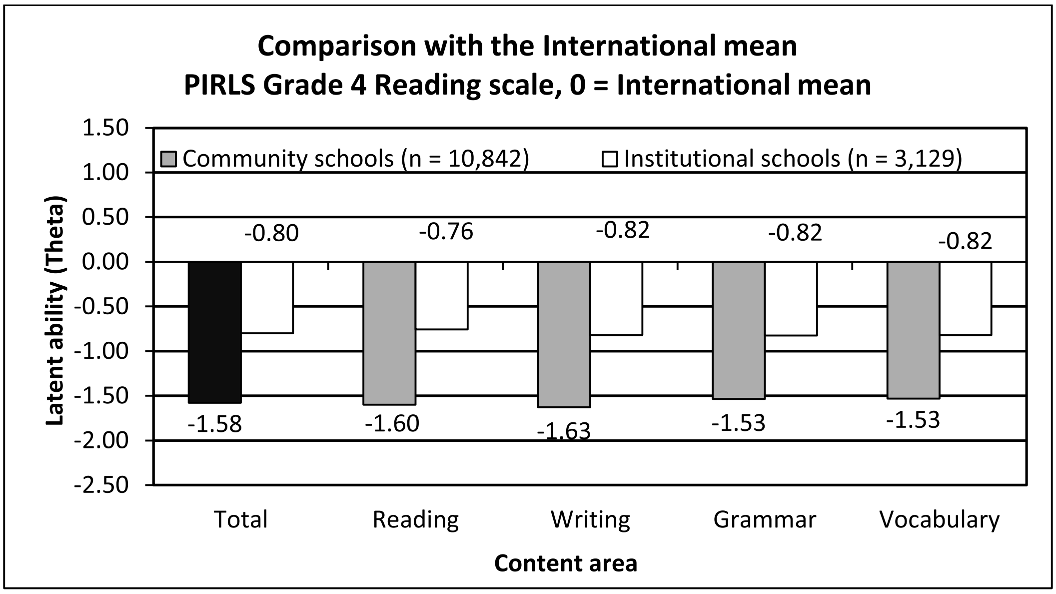

Figure 4.2.6 Student achievement in the international PIRLS reading scale

Figure 4.2.6 shows that the average reading proficiency of grade 5 students in Nepal is remarkably lower than that of the international level. The students in Nepal are remarkably lower in all the content areas of language compared to the international average. The average achievement level of grade 5 students in the community schools ($\theta = -1.58$) is very low compared to the average level of international students. The achievement level of an average students in the institutional schools ($\theta = -0.80$) is also remarkably lower than the international mean.

It is good to remember that all the linking items came from the content area of Reading and hence there actually is no real equating in the other areas. Especially, incomparable are Grammar and Vocabulary, because these areas are not measured at all in PIRLS. However, they are modeled on the basis of proficiency in the reading test.

Compared to the data with the international standard, the average reading proficiency of grade 5 students in Nepal is found to be much lower than the international average in PIRLS standards.

Comparison results with the objective standard – CEFR levels

Another type of international comparison was done on the basis of Common European Framework of Reference for Llanguage(CEFR) testing. CEFR classification with the standard setting procedure (3TTW) (Metsämuuronen, 2013; see also ERO, 2013) was applied to assess the criterion-based proficiency in Nepali language. The main results of Nepali Reading and Writing proficiency levels are presented in figures 4.2.7 and 4.2.8 and tables 4.2.10 and 4.2.11 respectively.

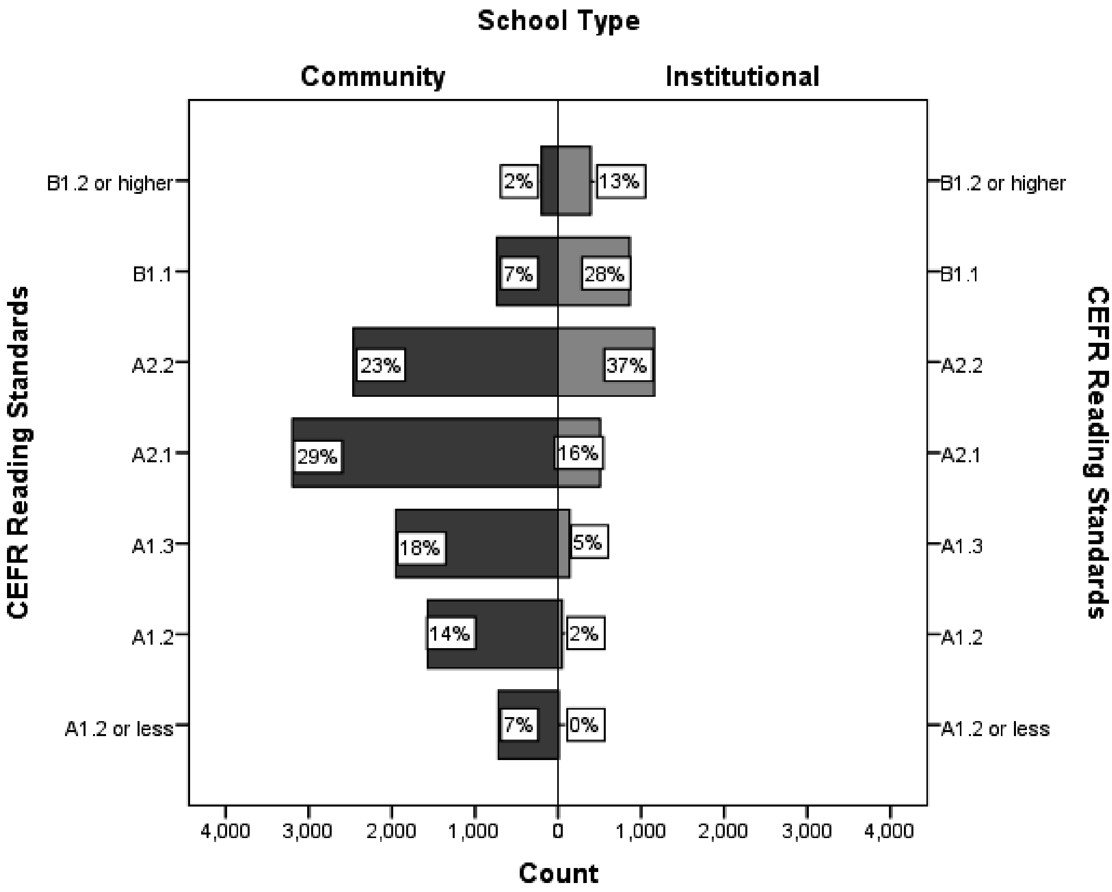

Figure 4.2.7 Student achievement in the international CEFR Reading standard

In the community schools, the most typical 5th grader (29%) of Nepali is at the level of A2.1. This means that the typical student "*can understand simple texts containing the most common vocabulary (personal letters, brief news items, everyday user instructions).* (S)he can *understand the main points and some details of a few paragraphs of text, can locate and compare specific information and can draw very simple inferences based on context.* Her/his *reading and understanding of even brief passages of text is slow.*" (FNBE, 2004, 278–295; see also ERO, 2013, table 4.6.4). In the institutional schools, the most typical 5th grader student of Nepali is at the level of A2.2 (37%). This means that the typical student "*can read a few pages of a wide variety of texts about familiar topics (tables, calendars, course programmes, cookery books) following the main points, key words and important details even without preparation.* S/he can *follow the main points, key words and important details of a few pages of text dealing with a familiar topic.*" (FNBE, 2004, 278–295; see also ERO, 2013, table 4.6.4).

In figure 4.2.7, it is seen that 69% of the 5th graders in community schools are at level A2.1 or lower – which implies that they can read and understand simple everyday texts and factual information and interpret them at slow pace. This kind of elementary reading skill was tested, mainly from the contents areas like child literature, simple description of information, letter, application, news and free writing. For example, in the reading task of A2.1 level (simple description), a simple notice of school "School request: No entry, without permission during the school hours" was given. The question is the multiple choice type: "Why does a permission need during the school time?" The options are: *(a) Students may bunk off the school, (b) To follow the school time, (c) To avoid disturbance in learning* and *(d) Teacher may bunk off the school*. Round 20% students did not respond to the question (missing value), 24% had selected option *(c),* and 12% had selected *(a).* Students have to contextualize the message to answer this question as answer is not apparently given in the text.

On the basis of the dataset, 4% of the whole student population has reached at least at the level B1.2 or higher in Nepali Reading, and hence they can read a few long paragraphs independently, interpret texts and relevancy of information but they face difficulties with specific vocabularies in longer texts (table 4.2.10).

Table 4.2.10 Percentages of 5th graders with the specific CEFR levels in Reading

CEFR level	Brief description of ability	% reaching the level	% at each level
B1.2 or higher	Can read **a few paragraphs of text about many *different* topics** (newspaper articles, brochures, user instructions, simple literature)	4.2	4.2
B1.1	Can read **a few pages of a wide variety of texts about *familiar* topics** (tables, calendars, course programmes, cookery books)	15.7	11.5
A2.2	Can understand the **main points and some details of** messages consisting of a few paragraphs in **fairly demanding every day contexts** (advertisements, letters, menus, timetables) and **factual texts** (user instructions, brief news items).	41.6	25.9
A2.1	Can understand **simple texts containing the most common vocabulary**.	68.1	26.5
A1.3	Can understand **very short messages dealing with everyday life** and routine events or giving simple instructions.	83.1	15
A1.2	Can understand names, signs and other very short and **simple texts related to immediate needs,**identify specific information in **simple text**, provided s/he reread it as required	94.7	11.6
< A1.2			5.2

When it comes to Nepali Writing proficiency in community schools, the typical 5th graders of Nepali vary ranging from A1.3 to A2.2 or higher though the mode is at A2.1 (figure 3.4.8b). This means that, compared to the Reading proficiencies, there are poorer 5th grader writers in the community schools. In Writing, community school students' performance is slightly lower than their Reading skills in Nepali. The typical community school student is at the level of A2.1 (26%). At this level, s/he can "*manage in the most routine everyday situations in writing.* S/he can *write brief, simple messages (personal letters, notes), which are related to everyday needs, and simple, enumerated descriptions of very familiar topics (real or imaginary people, events, personal or family plans).* S/he can *use concrete vocabulary related to basic needs, basic tenses and co-ordinate sentences joined by simple connectors (and, but).* S/he can *write the most simple words and structures with reasonable accuracy, but makes frequent basic errors (tenses, inflection) and uses many awkward expressions in free writing.*"(FNBE, 2004, 278–295; see also ERO, 2013, table 4.6.4.)

In institutional schools, the most typical 5th grader writer of Nepali is at the level of B1.1 (43%). There are also some students at a level of B1.2 (9%). This means that the typical student can, at least "*write an intelligible text about familiar, factual or imaginary topics of personal interest, also convey some detailed everyday information.* (S)he can *write a clearly formulated cohesive text by connecting isolated phrases to create longer sequences (letters, descriptions, stories, telephone messages).*(S)he can *effectively communicate familiar information in the most common forms of written communication, has sufficient command of vocabulary and structures to formulate most texts used in familiar situations, even if interference and evident circumlocutions occur. Routine language material and basic structures are by now relatively accurate, but some more demanding structures and phrases still cause problems."*(FNBE, 2004, 278–295; see also ERO, 2013, table 4.6.4.)

The typical writer is somehow at the same level (A2.1) as in Reading even though usually the receptive skills such as Reading is at the higher level as the productive skills (as Writing is). Also, the aimed level of the receptive skills may be set upper than for productive skills (see FNBE, 2004, 140, see also in ERO, 2013, table 4.6.3). In institutional schools, it is seen that the typical writer in grade 5 is as high as B1.1 or higher. This means that, compared to the Reading proficiency, there are many good writers among the 5th graders in institutional schools. Compared with the grade 3 (see section 4.1 and figure 4.1.10), there are also many good writers in the Grade 3, however, the percentage of students at the level B1.1 or higher is 51% at grade 5 and 42% at grade 3.[46]

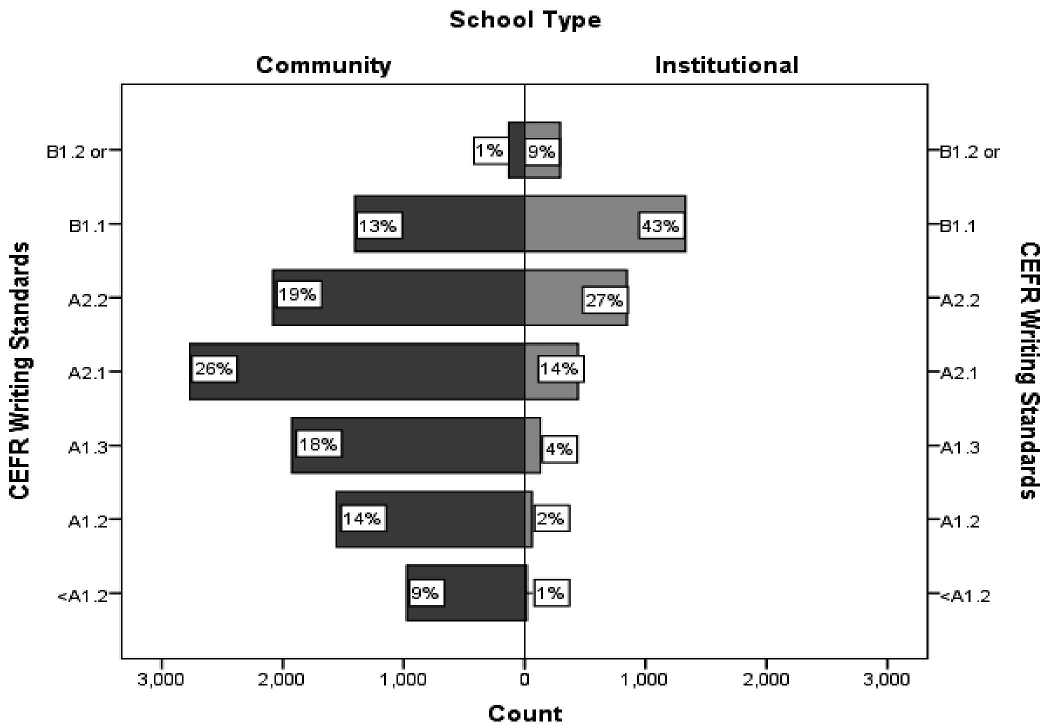

Figure 4.2.8 Student achievement level against international CEFR Writing standard

[46]At the grade 8 there were 62% of these students.

On the basis of the dataset, 67% of the whole student population has reached at least at the level A2.1 in Nepali Writing, and hence they can write simple messages on familiar everyday topics (personal letters and notes) using concrete vocabulary and conjunctions, construct sentences, with some grammatical errors (table 4.2.11).

Table 4.2.11 Percentage of 5th graders with the specific CEFR levels in Writing

CEFR	Brief description of ability	% reaching the level	% at each level
B1.2 or higher	Can **write a few paragraphs** of structured text (lecture notes, brief summaries and accounts based on a clear discussion or presentation).	3.0	3.0
B1.1	Can write a **clearly formulated cohesive text** by connecting isolated phrases to create longer sequences (letters, descriptions, stories, telephone messages).	22.6	19.6
A2.2	Can write a **very short, simple description of events, past actions and personal experiences** or everyday things in his/her living environment (brief letters, notes, applications, telephone messages).	43.6	21.0
A2.1	Can write **brief, simple messages** (personal letters, notes), which are **related to everyday needs**.	66.6	23.0
A1.3	Can manage to write in the **most familiar,easily predictable situations** related to **everyday needs and experiences**. Can write **simple messages** (simple postcards, personal details, simple dictation).	81.3	14.7
A1.2	Can **communicate immediate needs in brief sentences**. Can write a **few sentences** and phrases about him/herself and his/her immediate circle (such as answers to questions or notes).	92.9	11.6
<A1.2			7.1

Regarding the language proficiency in Nepali grade 5, two points are worth highlighting. First, usually in language learning, the receptive skills (Reading and Listening) are found to have stayed at the upper level than the productive skills (Writing and Speaking). Also, the aimed level of the receptive skills usually is set lower than for productive skills (see Table 3.4.9c; FNBE, 2004, 140; see also ERO, 2013, table 4.6.3). In Nepali grade 5 dataset, however, the Writing skills is found at the same level than those of Reading. One explanation for this is that, in general, the writing skills is found to have been preferred over the Reading skills in the current curriculum. Second, there is no objective criterion as regards what should be the language proficiency level at the end of grade 3 or grade 5. One tool for evaluating the proficiency is to use the Finnish core curriculum (FNEB, 2004) and the description of good performance as the measurement stick. In the Finnish system, the criterion is given for the end of grade six and hence it is not fully comparable in Nepalese context and, especially, it is not relevant in evaluating grade 5 proficiency. The criterion is not given for the mother language but for the foreign languages. Anyway, some clues of the required language proficiency levels can be obtained from table 4.2.12. The closest fit to the Nepali language comes the criterion for native level Finnish or in grade 5, would be from the criteria for Finnish as the first foreign language.

Table 4.2.12 Description of good performance in language at the end the sixth grade in the Finnish system

Language and level	Reading	Writing
Finnish as Native-level (bilingual) (FNBE,2004: 135)	B1.2 Fluent basic language proficiency	B1.1 Functional basic language proficiency
Finnish as the first foreign language (FNBE 2004, 130)	A2.1 Initial stage of basic language proficiency	A1.3 Functional elementary language proficiency

It is seen that the typical Nepali reader at grade 5 in the community schools (A2.1) is at the same level as the “good” reader is in the Finnish system studying Finnish as the first foreign language at grade 6. The typical Nepali writer (A2.1) is at the level higher than the “good” writer in the Finnish system for the students studying Finnish as the first foreign language at grade 6. In institutional schools, the typical Nepali reader (A2.2) is at the higher level than the “good” reader in the Finnish system for the students studying Finnish as the first foreign language at grade 6. However, the typical Nepali writer at grade 5 (B1.1) is found to have reached almost the same level as is required for the bilingual native speaker at grade 6. This is a very high level. More comparative studies are required to confirm the results.

Dataset shows that the most typical 5th grader student of Nepali in the community school is at the CEFR level of A2.1. This means that the typical student can read and understand simple everyday texts and factual information and interpret them at a slow pace. In institutional schools, the most typical 5th grader student of Nepali is at the CEFR level of B1.1. This means that the typical student can read a few pages of a wide variety of texts about familiar topics following the main points, key words and important details even without preparation.

4.2.2 Results Based on Diversity Factors

Although there may be many others diversities in our context, mainly three diversities are taken into account in this section, which include: district-wise, school type-wise (community/institutional), and school location-wise (rural/urban) diversity. These factors can be taken as equality factors as all children regardless of their sex, language, birth place, or family background should have equal opportunities to reach the same educational goal.

District variation in student achievement

Out of 75 districts, 25 were randomly selected to represent the Ecological zones and Development regions and ultimately the country. Additionally, 3 districts of the Kathmandu Valley (Kathmandu, Lalitpur, and Bhaktapur) were also selected because they present a unique stratum in the country. The district-wise differences are presented in table 4.2.13. The table shows the achievement in ascending order. The mean represents the average achievement percentage of the particular district.

Table 4.2.13 Average achievement score in selected districts

Districts	N	Mean	SD	CV	Districts	N	Mean	SD	CV
Bhaktapur	408	80.7	13.1	16.2	Dolakha	417	57.3	22.0	38.4
Kathmandu	1,558	78.1	16.0	20.5	Humla	156	57.0	20.6	36.1
Lalitpur	531	73.3	18.1	24.7	Makwanpur	603	56.9	21.1	37.1
Kaski	711	72.6	16.5	22.7	Rolpa	501	56.8	18.2	32.0
Solukhumbu	294	71.5	18.0	25.1	Bardiya	361	55.2	23.9	43.4
Darchula	372	64.0	18.8	29.3	Achham	503	53.3	23.6	44.3
Chitwan	606	63.0	19.2	30.5	Kapilbastu	599	52.9	23.8	45.0
Baglung	564	62.3	19.2	30.8	Khotang	504	52.0	23.0	44.3
Myagdi	302	59.7	18.8	31.5	Udayapur	495	50.7	23.1	45.6
Manang	9	58.4	18.6	31.9	Jumla	159	48.8	22.9	47.0
Baitadi	552	58.2	18.8	32.4	Sindhuli	562	48.4	22.1	45.7
Parsa	453	58.1	18.9	32.5	Salyan	525	45.7	21.9	47.8
Kailali	847	57.7	21.7	37.6	Mahottari	465	41.0	22.4	54.7
Dhankuta	387	57.3	21.7	37.9	Saptari	527	37.3	22.5	60.4
Total						**13,971**	**59.7**	**23.1**	**38.7**

Of the randomly selected districts in the sample, the students' performance was very low in Saptari (37%) from the Eastern region; in Mahottari (41%) and Sindhuli (48) from the Central region; and Salyan (46%) and Jumla (49%) from the Mid-Western region. Except for Kaski (73%) and Solukhumbu (71%), the outperforming districts include districts from the Valley: Bhaktapur (81%), Kathmandu (78%), and Lalitpur (73%).

The difference in achievement due to the district is statistically significant ($p < 0.001$). The variation explained in achievement due to the district is $\eta^2 = 0.235$, that is, the district explains 24% of the variation in the data which is a very high percentage. Effect size is $f = 0.55$ – which indicates that the difference between the lowest performing districts (37%) and highest performing districts (81%) is remarkably high.

The dataset strongly suggests that there is a wide difference between the districts when it comes to the equal opportunities of children to reach the preset goals in Nepali. The results are bound to the randomly selected 28 districts. The results are very high in the districts where the proportion of institutional schools is high. However, some districts (Solukhumbu and Darchula), having no institutional schools, have performed above the national average. The results in Saptari (37%), Mahottari (41%), Sindhuli (48%), Salyan (46%), and Jumla (49%) are very poor.

Ecological zone and student achievement

The Mountain, Hill, Tarai are the three Ecological regions in Nepal. The Kathmandu Valley is also regarded an additional Ecological region for its unique feature in terms of population density, more economic and employment prospects, aggressive development facilities available. The variation in achievement by Ecological zones is presented in table 4.2.14 and figure 4.2.9.

Table 4.2.14 Achievement in the Ecological zones

Ecological zone	Community schools				Institutional schools			
	N	**Mean**	**SD**	**CV**	**N**	**Mean**	**SD**	**CV**
Mountain	1,369	60.6	21.5	35.5	38	76.1	10.5	13.8
Hill	5,417	53.6	21.3	39.8	792	76.7	14.3	18.6
Tarai	3,168	48.3	22.2	46.1	690	73.5	16.7	22.7
Valley	888	71.1	18.7	26.3	1,609	81.0	13.4	16.6
Total	**10,842**	**54.4**	**22.3**	**41.0**	**3,129**	**78.2**	**14.7**	**18.8**

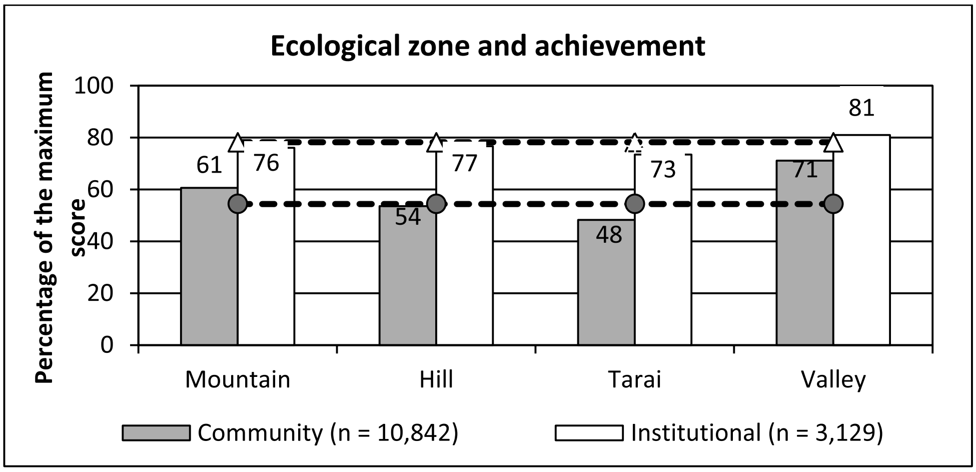

Figure 4.2.9 Differences in achievement in various Ecological zones

The data shows that, on average, the students from the Valley outperform the students from the other Ecological zones. The difference is wider in the community schools than the institutional schools. In both the community and institutional schools, the students from the Tarai area are seen to be performing the lowest (48% in the community schools and 73% in the institutional schools). It is notable that the exceptionally low value for the Coefficient on Variation for the community schools, is lower in the Valley than in other areas. The obvious reason for this is the systematic high score in the Valley compared to the other areas.

The achievement across the zones differs significantly in both the schools types ($p < 0.001$) even though the Valley is taken out of the analysis. Tukey's *post hoc* test tells that, within the community schools, there are significant differences between zones ($p < 0.001$). However, in the institutional schools, the students in the Tarai zone are at a lower level than the students in the Hill zone ($p < 0.001$). Ecological zone explains 8% of the variance in the community schools (η^2 = 0.078) and 5% in the institutional schools (η^2 = 0.045).[47] In comparison, district explains more than 24% of the variation. The effect size is f = 0.29 in the community schools and f = 0.22 in the institutional schools – showing moderate difference between the highest and lowest performing Ecological zones. The effect sizes are smaller if the Valley is taken out of the analysis (f =0.18 and f =0.11 respectively). This means that the real differences are not remarkable between the Ecological zones but the Valley differs radically from the other areas. From equality point of view, this can be taken as a good sign.

The dataset suggests that there is a moderate difference between the student performances in four Ecological zones in both community and institutional schools. Students in the Kathmandu Valley outperform the other students, and the achievement is the lowest in Tarai zone.

Development region and student achievement

Student achievement varies according to the Development regions which are divided into Eastern, Central, Western, Mid-Western and far-Western. Additionally, the Kathmandu Valley is taken as the 6th development region though administratively falls within the Central Development region. The mean achievements by the Development regions are given in table 4.2.15 and illustrated in figure 4.2.10.

Table 4.2.15 Achievement in the Development regions

Development regions	Community schools				Institutional schools			
	N	**Mean**	**SD**	**CV**	**N**	**Mean**	**SD**	**CV**
Eastern	2,077	50.3	24.1	47.9	130	74.0	16.7	22.6
Central	2,760	51.7	21.4	41.4	346	75.9	15.2	20.0
Western	1,519	57.0	20.8	36.6	666	75.8	15.3	20.1
Mid-Western	1,617	50.9	21.2	41.7	85	79.3	15.2	19.0
Far-Western	1,981	55.8	21.2	38.1	293	72.3	15.0	21.1
Valley	888	71.1	18.7	26.3	1,609	81.0	13.4	16.6
Total	**10,842**	**54.4**	**22.3**	**41.0**	**3,129**	**78.2**	**14.7**	**18.8**

[47] If the Valley is taken out of the analysis, the values for Eta squared would be 0.032 and 0.011 respectively, that is, only 3% and 1% explanation. The role of the Kathmandu Valley students in the whole national mean is remarkable.

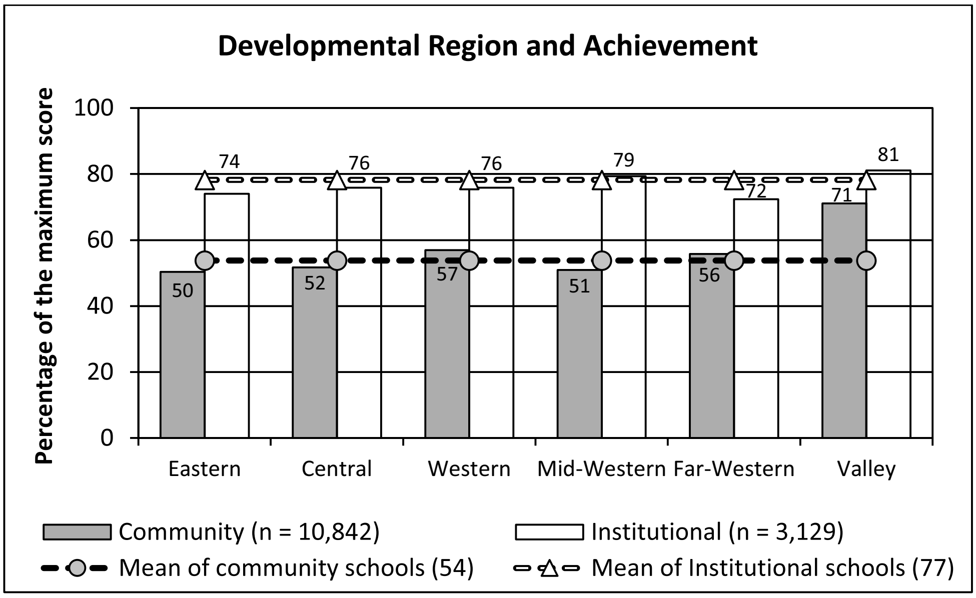

Figure 4.2.10 Comparison of student achievement in Development regions

The highest performance is found within the institutional schools in the Valley (81%) and Mid-Western region (79%). The performance is the lowest in the community schools in the Eastern (50%) and Mid-Western (51%) regions. The difference between the regions is statistically significant in both the community and institutional schools ($p < 0.001$). Tukey's *post hoc* test shows that, in the community schools, the average achievement levels in the Eastern and the Central regions are significantly lower than in the Western, Far-Western and the Valley ($p < 0.001$). The achievement level in the Western region is the highest of all the regions except for the Kathmandu Valley. The achievement level in the Valley is higher than in the other regions ($p < 0.001$). In institutional schools also the students in the Valley outperform the students in all other regions ($p < 0.001$).

Development region explains 6% of the variance in the community schools ($\eta^2 = 0.062$) and 5% in the institutional schools ($\eta^2 = 0.046$).[48] This is somehow the same proportion as found with the Ecological zone. One remembers that the district explains more than 24% of the variation which means that within the Developmental regions there are lower and higher performing districts. The effect size is $f = 0.26$ within the community schools and $f = 0.22$ within the institutional schools showing moderate difference between the highest and lowest performing regions. The effect sizes are small when the Valley is taken out from the analysis ($f = 0.12$ and $f = 0.11$ respectively).

The dataset reveals that there is wide inequality of children's opportunities to reach an adequate level of Nepali among the Development regions. Especially, the wide difference between the community schools in the Valley and in the rest part of the country (21 percent as the highest) is a strict sign of inequality of opportunities in learning Nepali. There are also wide differences between the regions in the institutional schools. The difference in the

[48] If the Valley is omitted from the analysis, the values for the *Eta* squared would be 0.014 and 0.012 respectively, that is, only 1% in both community and institutional schools. The role of the Kathmandu Valley students in the whole national mean is remarkable.

student performance in the institutional schools between the Valley and Far-Western is the highest, i.e., 9 percent.

School type and student achievement

All the schools are categorized into community and institutional (private) schools. The difference in students' achievement in Nepali in community and institutional schools is presented in table 4.2.16.

Table 4.2.16 Type of school and the average achievement

Content area	Community (N = 10,842)			Institutional (N = 3,129)			Mean difference	Cohen's *d*
	Mean	SD	CV	Mean	SD	CV		
Reading	50.4	24.3	48.1	75.5	17.5	23.1	25.0	1.09
Writing	52.8	23.8	45.0	76.3	16.0	21.0	23.6	1.06
Grammar	59.0	25.4	43.0	80.3	16.5	20.6	21.3	0.90
Vocabulary	65.6	26.0	40.6	85.4	15.9	20.2	14.9	0.81
Total	**54.4**	**22.3**	**41.0**	**78.2**	**14.7**	**18.8**	**23.8**	**1.14**

The achievement levels between the community schools and institutional schools differ from each other remarkably, i.e., with 24 percent. The average performance in total score in the institutional schools is 78% whereas it is 54% in community schools. The difference is statistically significant ($p < 0.001$) and the effect size is very high ($d = 1.14$) – showing that community schools are far below the institutional schools.[49] Difference is the highest in the content area of Reading ($d = 1.09$) and Writing ($d = 1.06$). Division of students into community and institutional schools explains 17% of the student variation in reading ($\eta^2 = 0.172$) and 16% in Writing ($\eta^2 = 0.163$). From figure 3.4.9, it is known that the deviance in the community schools is remarkable ranging from 37% to 80%. Contrarily, most private schools in the sample show very high performance. It indicates that institutional school students are comparatively good in all content areas. This can be explained partly by much higher socio-economic input into students' life in the institutional schools and strict selection of the students. The dataset reveals that, on average, the students in institutional schools outperform the students in the community schools. The difference is highest in Reading (25 percent).

School location and student achievement

One of the strata considered in sampling in NASA 2012 was the school location by dividing the schools as rural and urban. This information was given by the head teacher though some of the head teachers did not inform the school location. The achievements of the students in rural and urban schools are presented in table 4.2.17.

[49]The effect size if high ($d = 1.05$) even if the Valley is omitted in the analysis.

Table 4.2.17 Student achievement on the basis of location of school

School Location	Community				Institutional			
	N	Mean	SD	CV	N	Mean	SD	CV
Rural	9,319	53.0	21.9	41.3	947	75.4	15.3	20.3
Urban	867	61.9	20.5	33.2	1,841	78.6	13.8	17.6
Mean difference		8.9				3.2		
Cohen's d		0.33				0.26		
Total	**10,842**	**54.4**	**22.3**	**41.0**	**3,129**	**78.2**	**14.7**	**18.8**

The achievement level of the students in the urban community schools is 62% which is around 9 percent higher than that in the rural community schools (53%). The difference is statistically significant ($p < 0.001$), and the effect size is medium ($d = 0.33$). Excluding the community schools of the Valley, the score of the urban community schools lowers to 55%; the difference (3 percent) is statistically significant ($p = 0.002$) but the effect size is low ($d = 0.13$). The main difference in community schools is, hence, caused by the higher level of students in the Valley schools. The division into rural and urban schools explains 0.13% of the student variation within the community schools ($\eta^2 = 0.013$) and without the Valley schools it is only 0.1% ($\eta^2 = 0.001$). The latter is a good sign from the equality point of view. Excluding the Valley there is no difference between rural and urban community schools.

In the urban institutional schools, the achievement level of the students (79%) is 3 percent higher than that in the rural institutional schools (75%). Though the difference is statistically significant ($p < 0.001$), the effect size is, at most, moderate ($d = 0.26$). Excluding the community schools of the Valley, the difference remains the same (2.4 percent points) which is still statistically significant ($p = 0.004$), but the effect size is low ($d = 0.16$). Within the institutional schools, the effect of Kathmandu Valley is not remarkable. The division into rural and urban schools explains 1.1% of the student variation within the institutional schools ($\eta^2 = 0.011$) and 0.6% without the Valley schools ($\eta^2 = 0.006$). The latter is a good sign from equality point of view. Without the Valley there is no difference between rural and urban community schools. Within the institutional schools there is not much difference between the rural and urban areas.

The dataset indicates that the students in the urban community schools achieved 9 percent more than the students in the rural areas. Excluding the Valley schools, the difference is 3 percent. The difference as seen is not a good sign for equality though the real difference is not wide within the community schools. Within the institutional schools, there is not wide difference between the rural and urban areas.

Language at home and student achievement

In the context of Nepal, student achievement is found to be depended on the language spoken in their homes i.e., the mother tongue of the students. The mother tongue reflects, in many cases, the ethnic background and hence any difference is taken as a possible source for inequality in society.

On the basis of the total data, 36% of the 5th 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 Tharu (5.4%), Urdu (4.7%), Newari (3.1%), and Tamang (2.7%). After dividing the languages into ten groups excluding Nepali, there were still 15.4% of the students classified into the group "Other". Because the languages are very fragmented and the Nepali speakers are the majority of the students, for the purpose of the statistical analysis, all the other languages are first grouped into "Non-Nepali". The results are presented in tables 4.2.18 and 4.2.19 and illustrated in figure 4.2.11.

Table 4.2.18 Student achievement on the basis of home language

Language group	Community				Institutional			
	N	Mean	SD	CV	N	Mean	SD	CV
Nepali	6,375	57.4	21.2	37.1	2,608	78.9	14.2	18.1
Non-Nepali	4,467	50.1	23.1	46.0	521	74.7	16.6	22.1
Mean difference		7.3				4.2		
Cohen's *d*		-0.327				-0.285		

When combining all the minor language groups as "Non-Nepali", there is a notable difference between the language groups in the community schools (7 percent favoring the Nepali speakers). The difference is statistically significant ($p < 0.001$) and the effect size is, at most, medium size ($d = 0.33$). The difference of 4 percent in the institutional schools is low ($p < 0.001$, $d = 0.29$).

On the basis of the original categorization of the minor languages, the issue looks quite much interesting. It is evident that the Magar students are at much higher level in Nepali than the Nepali speaker students (67% compared with 57% in the community schools). On the other hand, the students from Tharu (38%), Rai (50%), Sherpa (50%), Newari (52%), and Other (49%) language backgrounds perform much lower than the average. Also in institutional schools Magar students are at a higher level (average of 80%) as well as the Gurung students (82%).

Table 4.2.19 Achievement of the different language groups by school types

Language[1]	Community				Institutional			
	N	Mean[2]	SD[3]	CV[4]	N	Mean[2]	SD[3]	CV[4]
Magar	61	66.9	24.6	36.7	120	80.3	14.2	17.7
Limbu	7	64.3	20.3	31.5	8	55.6	18.6	33.5
Tamang	349	58.9	18.8	31.8	12	78.5	12.2	15.6
Nepali	6,375	57.4	21.3	37.1	2,608	78.9	14.2	18.1
Maithili/Awadhi	25	57.2	18.5	32.4	13	78.9	13.4	17.0
Urdu	606	55.9	21.1	37.8	25	77.5	11.6	15.0
Newari	392	52.4	21.2	40.5	32	72.9	16.2	22.2
Gurung	17	51.0	26.7	52.2	3	81.9	15.0	18.3
Sherpa	28	50.3	20.1	40.1	1			
Rai	44	50.1	20.1	40.0	1			
Tharu	680	38.2	19.6	51.3	47	80.9	12.6	15.6

1)Language groups with less than 10 students are omitted 3) SD=Standard Deviation
2)Language groups are sorted on the basis of the community school mean 4) CV=Coefficient variation

It is interesting that the Tharu performed low in the community schools, but their performance is high (81%) in the institutional schools. This means that the language background itself is not the determining factor in achievement, rather it might be the learning environment.

The differences between the students in the highest and lowest performing language groups are statistically significant ($p < 0.001$) and notable; the effect sizes are moderately high ($f = 0.25$ in the community schools and $f = 0.18$ in the institutional schools). The division into smaller language groups explains about 6% ($\eta^2 = 0.059$) and 3% ($\eta^2 = 0.032$) of the variation in the data in community and institutional schools respectively. Though the differences are wide between the extreme groups, it is good to keep in mind that the number of students is quite small in some of the language groups. Hence, the effect size is moderate. When looking only at the minority languages and hence excluding the Nepali speakers and the group “Others”, the effect size is high ($f = 0.43$) in the community schools – indicating remarkable difference between the highest performing minority group (Magar, 67%) and the lowest performing group (Tharu, 38%).

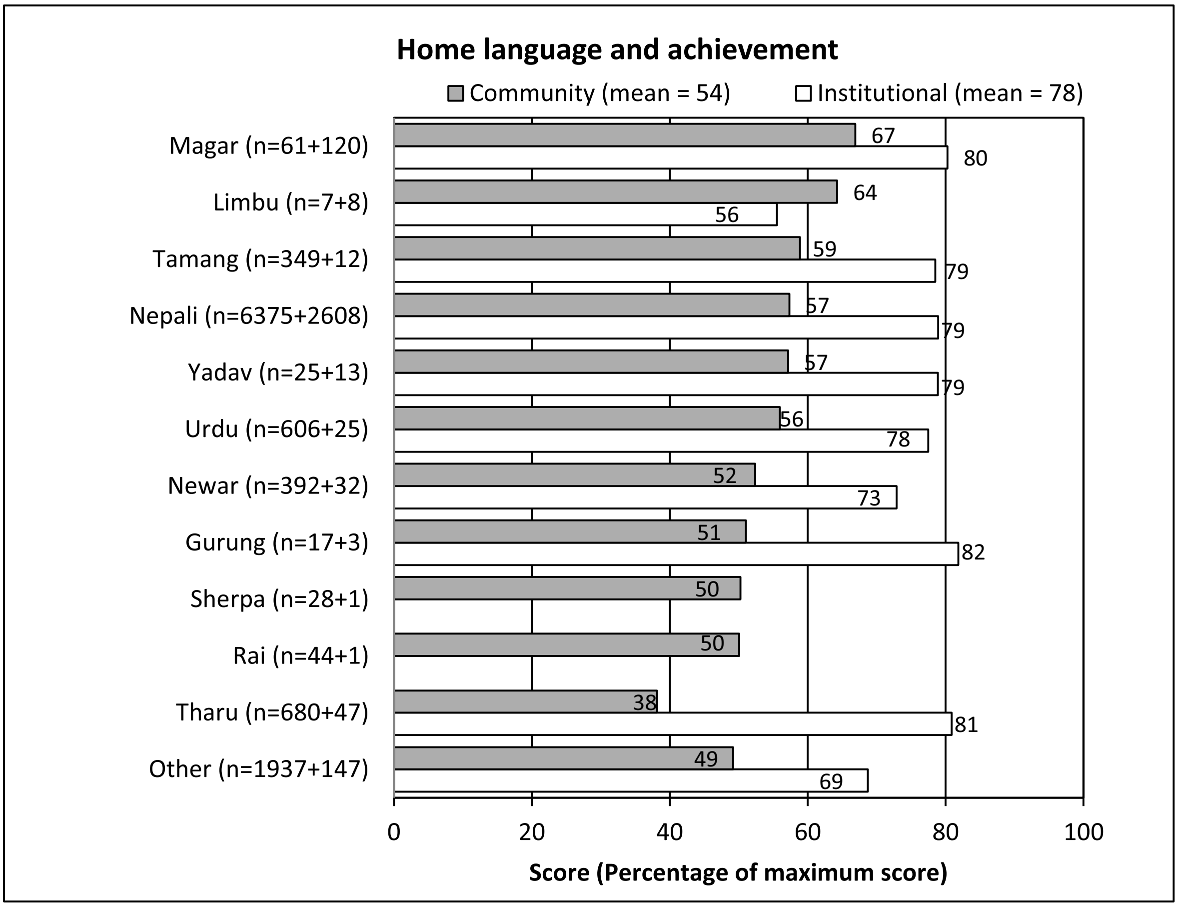

Figure 4.2.11 Relation between language at home and achievement

Language and Development region

When combining the results from the Development region and mother tongue, one notices that the achievement score of the students within a certain language group varies drastically between the different regions (table 4.2.20).

Table 4.2.20 Achievement of various language groups by Development region

Development region	Nepali[1]	Tharu	Urdu	Newari	Tamang	Magar	Rai	Maithili/ Awadhi	Sherpa	Gurung	Limbu
Eastern	56.5	40.7	61.6	39.6	46.1	33.6	52.0		50.2	66.3	
Central	58.7	39.9	55.1	53.1	60.0	50.7	48.5	62.7	30.0	25.3	48.9
Western	66.8	68.9	65.9	55.6	35.0	85.0	5.6	65.1	72.2	6.7	58.1
Mid-Western	54.1	73.7		51.2	39.4	64.4				45.0	
Far-Western	58.6	40.6	56.3	57.1		73.9					42.2
Valley	78.2	83.8	62.3	82.2	84.2	80.2		65.6	93.3	84.1	90.0
Total	**63.6**	**40.9**	**56.8**	**53.9**	**59.5**	**75.8**	**50.8**	**64.6**	**51.8**	**55.7**	**59.6**
N[3]	8,983	727	631	424	361	181	45	38	29	20	15

1) *The language groups of less than 10 students are not included in the table.*
2) *The main population is highlighted . In some un-highlighted cases there is only one student behind the mean. 3) The language groups are ordered on the basis of their frequency*

All language groups except for Urdu and Maithili/Awadhi score high in the Valley. The students from the Magar, Maithili/Awadhi, and Sherpa speaking groups performed high also in the Western region. In many cases, the lowest means of the language groups are found in the Eastern or Western regions – especially in Tharu, Newari, Magar speaking societies. The extremely poor result of among Rai and Gurung students (6 and 8) in the Western region do not affect the mean of the language group.

Language and Ecological zone

In all the specified language groups, except for Urdu, Magar and Maithili/Awadhi, the students in the Valley are found ahead compared to the other Ecological zones (table 4.2.21). The results in Hill region are lower than the mean in all the language groups. In many cases (Tharu, Urdu, Rai, Magar, Gurung, and Limbu), students are performing lower at the Tarai zone than in the other zones. Newari, Magar, and Sherpa students are seen to be performing lower in the Hill zone.

Table 4.2.21 Achievement of various language groups by Ecological zones

Ecological zone[1]	**Nepali**	**Tharu**	**Urdu**	**Newari**	**Tamang**	**Magar**	**Rai**	**Aabadhi/ Maithali**	**Sherpa**	**Gurung**	**Limbu**
Mountain	60.9	45.6	62.4	49.6	57.0	86.9	86.7	44.2		51.7	
Hill	57.9	59.8	55.1	46.6	67.3	46.8	49.8	68.0	50.3	58.6	77.8
Tarai	62.7	39.7	43.7	54.4	58.6	45.7	21.1	64.3		6.7	53.1
Valley	78.2	83.8	62.3	82.2	84.2	80.2		65.6	93.3	84.1	90.0
Total	**63.6**	**40.9**	**56.8**	**53.9**	**59.5**	**75.8**	**50.8**	**64.6**	**51.8**	**55.7**	**59.6**
N[3]	8983	727	631	424	361	181	45	38	29	20	15

1) *The language groups of less than 15 students are not included in the table.*
2) *The main population is highlighted by the gray shade. In some un-highlighted cases there is only one student behind the mean.*
3) *3) Language groups are ordered on the basis of frequency.*

The dataset shows that there is an educational inequality among the language groups in learning Nepali. As a whole, the Magar speakers (76%) are far ahead of the other students (41 – 65%) in Nepali proficiency whereas the Tarai Tharu students are far behind the others (40%). Relatively poor results are also found among the Rai (51%), Sherpa, (52%) and Newari (54%) populations. Development region and Ecological zone-wise analysis raise an important question as Why are the results in Nepali so lower in some regions or zones among the certain language speakers and what can be done to raise their standard?

Caste/Ethnicity and student achievement

Although participation rate of Hill Dalit students has increased remarkably in the lower level of education, their number in the secondary and higher education is still very small and their achievement is also low. The results concerning the ethnicity and achievement are presented in table 4.2.22 and illustrated in figure 4.2.12.

Table 4.2.22 Achievement of various ethnic groups by school type

Ethnicity/Caste	Community				Institutional			
	N	**Mean**	**SD**	**CV**	**N**	**Mean**	**SD**	**CV**
Brahman	874	62.2	20.6	33.1	794	81.3	12.3	15.1
Chhetri	2,541	58.6	20.4	34.7	693	78.1	14.7	18.8
Madhesi	758	49.4	22.4	45.3	148	74.5	17.8	23.9
Janjati	3,010	56.7	22.0	38.7	903	79.1	14.0	17.7
Dalit	1,468	53.5	20.7	38.7	122	71.4	18.2	25.6
Others	1,633	46.0	22.7	49.4	335	75.0	15.5	20.7
Total	**10,521**	**54.4**	**22.3**	**41.0**	**3,016**	**78.4**	**14.6**	**18.7**

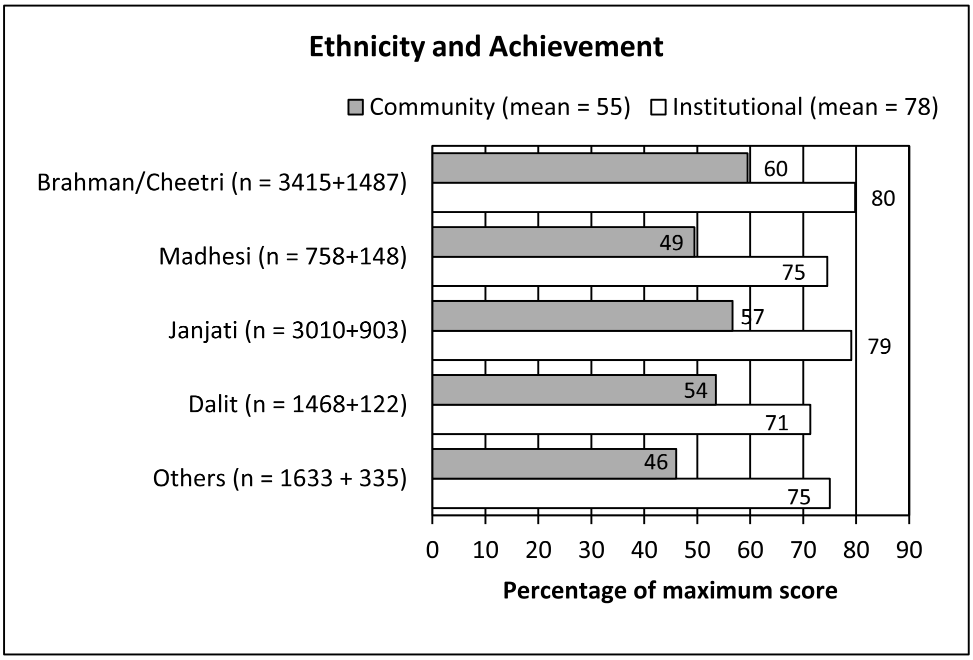

Figure 4.2.12 Relation between caste/ethnicity and achievement

In the community schools, the students from the “Other” caste/ethnicity are performing the lowest (46%) in Nepali, followed by Madhesi (49%), Dalit (54%) and Janjati students (57%). Dalit and Madhesi students perform below the average also in the institutional schools. The overall difference between the groups is statistically significant ($p < 0.001$) though the effect size is medium ($f = 0.23$) in the community schools. The division of students according to their ethnic background explains just 4.8% of the student variation ($\eta^2 = 0.048$). Within the institutional schools, the effect size is also small ($f = 0.16$). Here, the division of the students according to their caste/ethnic background explains just 2.4% of the student variation ($\eta^2 = 0.024$). From equality point of view this is a good sign though there is still a lots to do to reduce the achievement gap between the ethnic groups.

It is possible that the low performance of Madhesi students in Nepali is due to the little use of Nepali language in their daily life and greater reliance on their own mother tongue for daily communication. A positive sign from the equality point of view in the community schools is that the Madhesi students perform better than the national mean (55%) in the

Kathmandu Valley (71%), Eastern Mountain (58%) and Hill (68%) including in the Far-Western Mountain (79%) and Hill (66%). On the other hand, their performance is remarkably lower in the Tarai zone (ranging 38–51%) where their population is largest (table 3.4.20). In the institutional schools, Madhesi students are good in the Kathmandu Valley (79%), Eastern Tarai (75%), Central Tarai (80%), Western Hill (74%), and Mid-Western Tarai (85%). However, the Madhesi students from the Western and Far-Western regions still performed the lowest of all the institutional schools.

Table 4.2.23 Madhesi students' achievement in different Ecological and Development regions

School Type	Ecological zone	Development Region						
		Eastern	**Central**	**Western**	**MidWestern**	**FarWestern**	**Valley**	**Total**
Community	Mountain	58.5	40.3		67.2	78.9		55.3
	Hill	68.6	62.7	57.2	49.7	66.4		61.6
	Tarai	37.6	50.1	50.7	46.8	49.0		46.9
	Total	40.6	50.3	51.3	50.2	57.6	71.4	49.4
Institutional	Mountain		63.9					63.9
	Hill	70.6	83.3	74.4		27.8		70.0
	Tarai	75.3	80.5	54.5	84.9	67.4		74.0
	Total	**73.2**	**79.6**	**58.3**	**84.9**	**57.5**	**78.9**	**74.5**

The dataset informs that there are statistically significant though not necessarily remarkable differences between the castes/ethnicities in Nepali. Students from the Madhesi (49%) and Dalit (53%) as well as "Other" ethnic backgrounds (46%) are performing significantly lower than the students from the Brahmin, Chhetri, and Janjati groups. Madhesi students performed lower especially in Tarai zone including the Western and Far-Western regions.

Gender and student achievement

As the gender equality is important in the modern discourse, the matter is handled somehow more extensively than in the previous sections of equality. Basic results are presented in table 4.2.24 and figure 4.2.13.

Table 4.2.24 Student achievement of boys and girls by school type

Sex	Community				Institutional			
	N	Mean	SD	CV	N	Mean	SD	CV
Boys	4,884	53.4	22.2	41.5	1,565	76.4	15.5	20.3
Girls	5,524	55.5	22.3	40.2	1,453	80.4	13.3	16.5
Total	10,284	54.5	22.3	40.8	2,995	78.4	14.6	18.7

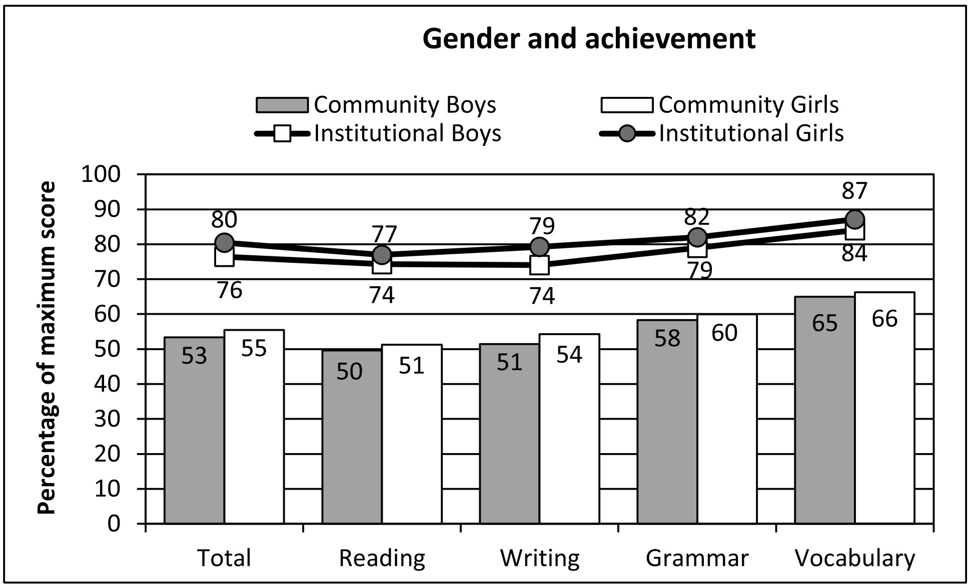

Figure 4.2.13 Comparison of boys and girls in different content areas

There is statistically significant difference between boys (53%) and girls (55%) ($p < 0.001$) in community schools. Girls are ahead in all the content areas though the widest difference is in the area of writing and girls (54%) are better writers than boys (51%) ($p < 0.001$). However, differences are seen to be slightly wider in the institutional schools which are comparatively higher in writing (5 percent) ($p < 0.001$). In all content areas, the differences are small though effect sizes are $d < 0.10$ in both types of schools except for Writing in institutional schools where the effect size is $d = 0.34$. From the equality point of view, this is a positive sign though there is still a lots to do to reduce the gap.

Looking at the Ecological zones, girls perform slightly higher compared to the boys in Hill (4 percent) and Valley (3 percent) in Nepali (figure 4.2.14). The boys are better than the girls by 6% in institutional schools in the Mountain zone. Otherwise, it is difficult to find any difference between the gender when it comes to the Ecological zones.

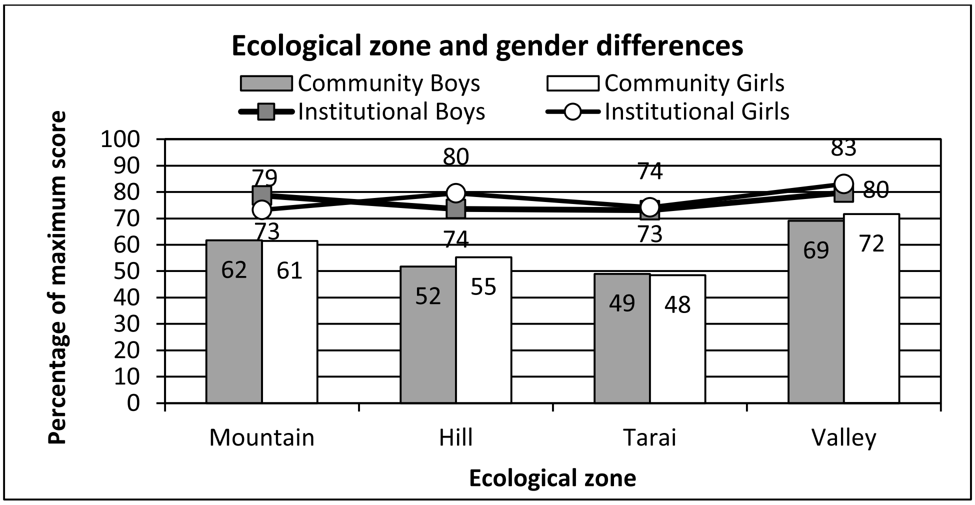

Figure 4.2.14 Ecological zone and gender-wise differences by school type

There are no notable differences between the Developmental regions when it comes to boys' and girls' equal opportunities to reach the same educational goals (figure 4.2.15). The difference noticed between boys and girls is seen somehow wider in the community schools in the Eastern, Central, and Valley regions (3–4 percent points). In institutional schools, the gap is widened in the Eastern (9 percent), Central and Western regions (4 percent in both).

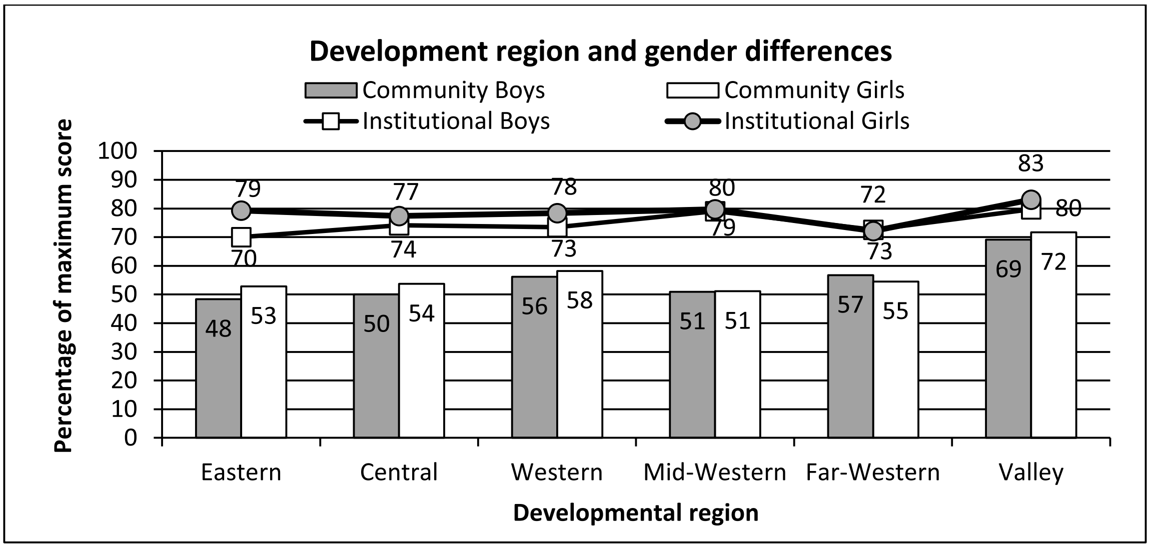

Figure 4.2.15 Development region and gender-wise differences

The dataset shows that the differences between boys and girls in Nepali proficiency are very moderate. The differences are found wider in the institutional schools compared to the community schools. In Writing, the effect sizes are small indicating that the differences are not at all remarkable. From equality point of view, this is a positive sign. A tendency is seen that the girls are slightly out-performing than the boys in both types of schools. However, the boys in the institutional schools outperform girls in Mountain area.

4.2.3 Selected Explanatory Factors and Achievement

The simplistic model in section 2.4 (figure 2.2) represents several possible factors, which have explained the differences in student achievement. Some of the factors have already been handled in section 4.2.2 including geographical factors, such as districts, Ecological zone, and Development region as well as school related technical factors, such as school type and school location. Similarly, some individual related factors were also handled, such as home language, caste and gender. In this section, several other factors are taken into consideration, which are the socio-economic status (SES) of the students' families, paid work after school, students' attitude towards Nepali as a school subject, age of the student, and support provided to the studies are mainly the family and individual related factors. Some important school and teacher related factors include: the availability of school books, homework assigned by the teacher, students' future plans, and selected activities in the school.

While discussing these factors two things are worth mentioning, particularly for grade 5. First, student had difficulty to respond meaningfully to some of the background questionnaire and therefore teachers were asked to help the students to fill the questionnaire. Second, there are missing values in the background questionnaires. For example, in the question of mother's education, some students did not answer the question. This evidently has an effect in the analysis, and therefore the readers need to be critically aware of the situation when it comes to the grade 5 students. In most cases, however, the results are considered to have been credible and comparable with grade 3 and 8 datasets.

Parents' education, occupation, home possessions and student achievement

There are several variables indicating the socio-economic status, which were categorized into parents' education, parents' occupation, home possessions (whether or not the student has his/her own space to do homework, or a dictionary, for example), home accessories (how many mobile phones, televisions and computers there are 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 economic, educational, and occupational background of the family. In this section, the education of the parents is further elaborated on so that illiteracy of the parents is analyzed in relation to the Nepali language achievement.

Several SES-related variables were analyzed by using a data mining tool of SPSS,DTA. The method is 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 most significant ways from each other in relation to student achievement. Some examples of this are handled with parents' educational background and its relation with students' achievement in Nepali.

Parents' education

In the background questionnaire, parents' education is grouped into eight categories: 1) illiterate, 2) literate, 3) grades 10 pass, 4) SLC pass, 5) IA pass, 6) BA pass, 7) MA pass, and 8) Above MA pass. The question was asked to the students and hence there may be some impurities embedded in the data; the number of (just) literate mothers and fathers in the dataset is found too high (see figure 3.2.18b). However, with the huge dataset the results are credible.

DTA classifies mother's education into four groups with statistically significant differences in students' achievement levels (see figure 4.2.16): illiterate (students' average is 58%), just literate (59%); grade 10 passed (70%); SLC pass or higher (77%). The difference between the groups is statistically significant in each case ($p < 0.001$). The results indicate that if the mother is grade 10 passed, the student performance is on average +12 percent more than the illiterate mother and she has contributed +11 percent point more over the just literate mother. If mother is SLC passed or higher, the student performance is +19 percent points higher than when the mother is illiterate. Figure 4.2.17 shows this scenario in detail.

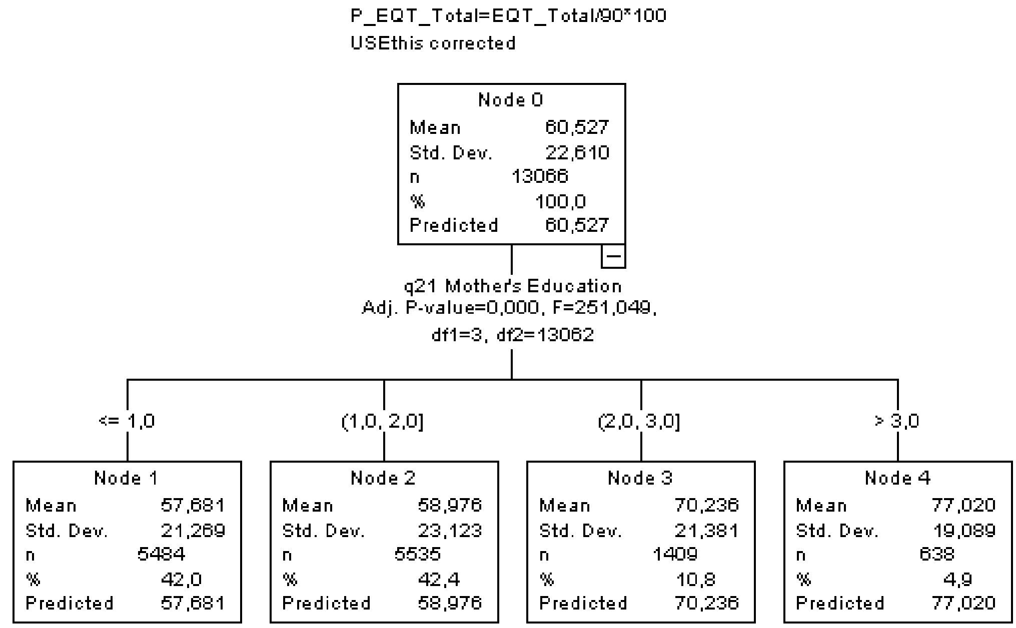

Figure 4.2.16 DTA of mother's education and students' achievement

The figure 4.2.17 shows that if the mother has higher education up to BA, the advance was +26 percent points over the illiterate mother. Mother's education explains 9% of the student variation (η^2 =0.057) which indicates a moderate effect size (f = 0.25). Obviously, the result means that the children from the highly educated mothers are mainly found in the private schools. The highest educational group ("Above MA"), which looks to be deviating from the trend, include also those students who always tend to achieve always the highest even though it may not be true (see below the discussions with the other SES indicators).

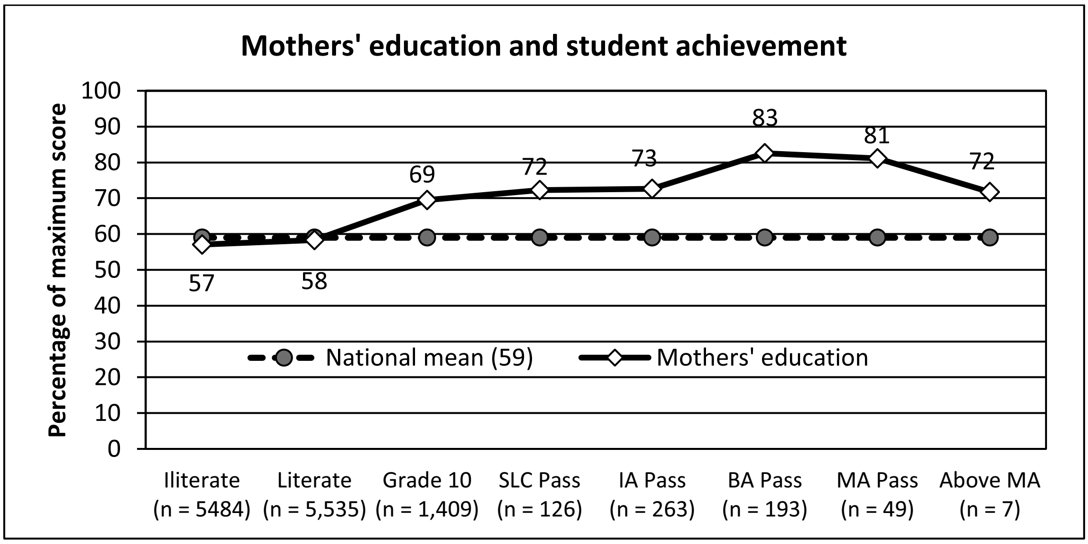

Figure 4.2.17 Mothers' education and students' achievement

In parallel, DTA divides father's education into four categories: illiterate (55%), (just) literate (58%), grade 10 passed or SLC passed (67%), and higher than SLC passed (73%) (figure 4.2.17). The difference between each group is statistically significant ($p < 0.001$). In practical words, the results mean, if father has the educational qualification equivalent to grade 10 passed or SLC passed, his children achieves on average, + 13 percent greater score compared to the illiterate father. If the father has passed SLC or above, the student's performance is +19 percent higher over the children of illiterate father. Figure 4.2.18 shows this scenario in detail.

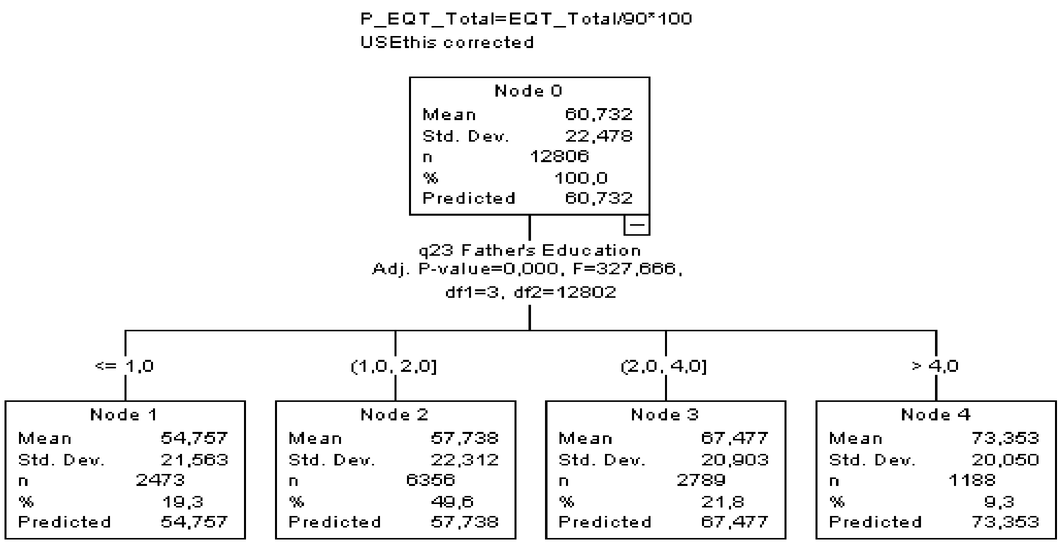

Figure 4.2.18 DTA of fathers' education and students' achievement

Figure 4.2.19 shows that if the father was MA passed or higher, the advance was + 25 percent points over the illiterate father. Obviously, the high average means that the children from the highly educated fathers (as well as mothers) are mainly found in the private schools. Father's education explains 7% of the student variation ($\eta^2 = 0.075$) which indicates a moderate effect size ($f = 0.28$). The last category includes also those students who had a highest possibility of achieving more which has not been true (see below the discussions with the other SES indicators).

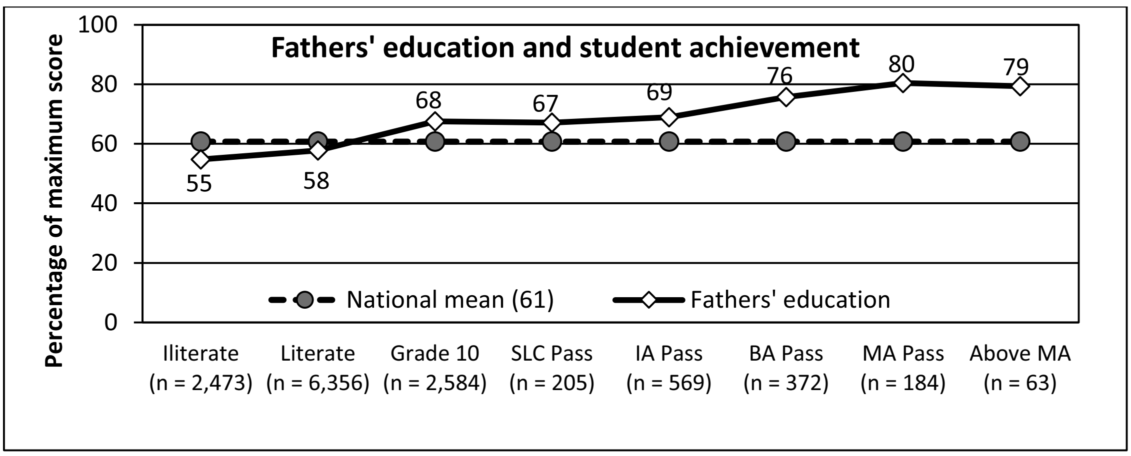

Figure 4.2.19 Fathers' education and students' achievement

It is noteworthy that, in grade 5, fathers' education has given slightly higher leverage to children's achievement in Nepali than that of mothers' (compare with Mathematics, section 3.2, and English, in section 5). While mother's literacy and education (up to grade 10) has contributed to raise the children's achievement by 12 percent, father's literacy and education (up to grade 10), has helped to raise it by 13 percent. On the other hand, mother's higher education (BA and MA passed) have raised the children's achievement by 24–26 percent whereas father's higher education have raised it by 22–26 percent. In both cases, the effect size is moderate or high (f=0.31– 0.33) showing that the difference between the highest and lowest group is remarkable.

After combining the mothers' and fathers' education, the poorest prediction in DTA for the children's future achievement in Nepali comes when the father is illiterate regardless the mothers' education (55%) and when the father is just literate (58%), or both parents are just literate (58%).[50] The highest results are found in the group where father is IA passed or higher regardless the mothers' education (73%). It is evident that the educational capacity provided by the parents can be utilized by the students which means: the higher the parents' education (especially, father), the better results will be achieved by the children.

In what follows with the final SES variable, the cut-off for parental education was set to "grade 10", that is, passing the grade 10 (or higher), the indicator of mothers (and fathers') education for SES was set to 1, and the education lower than SLC passed gave the value 0.

Parents' occupation

The occupation of parents was categorized into eight groups: 1) working abroad, 2) farming and working at home, 3) only working at home, 4) teaching, 5) services, 6) business, 7) daily wages, and 8) working at others' home. The result related to mothers' and fathers' occupation is seen respectively in figures 4.2.20, 4.2.21, 4.2.22 and 4.2.23.

While comparing the students' mean by DTA, the achievement is the lowest when mothers' occupational background comes from agriculture or the mother is working abroad (57%) (figure 4.2.20). Statistically, it is significantly higher when the mother is working only at home (63) or is a teacher (75%).

[50] The lowest results come in the groups where the father is just literate but the mother's education is unknown (40%) or father's education is unknown (48%).

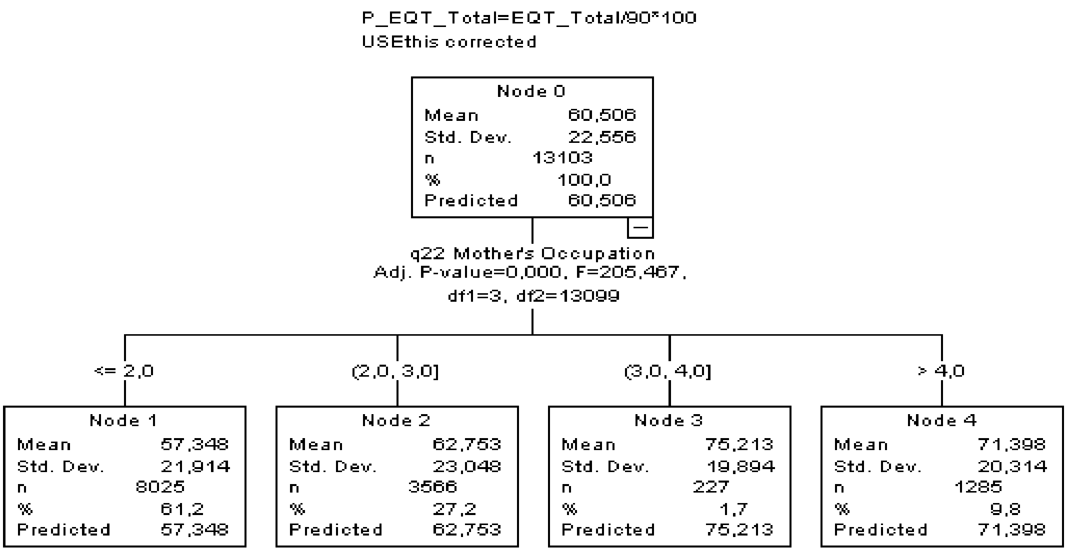

Figure 4.2.20 DTA of mother's occupation and students' achievement

Figure 4.2.21 shows that the achievement level is higher when mother is involved especially in service (76%) and teaching (75%). Mothers' occupation explains 5 percent of the student variation (η^2 = 0.049) which indicates a moderate effect size (f = 0.23). However, the result indicates that the mothers who do not work for agriculture or work abroad is seen to be very beneficial for achievement of the students.

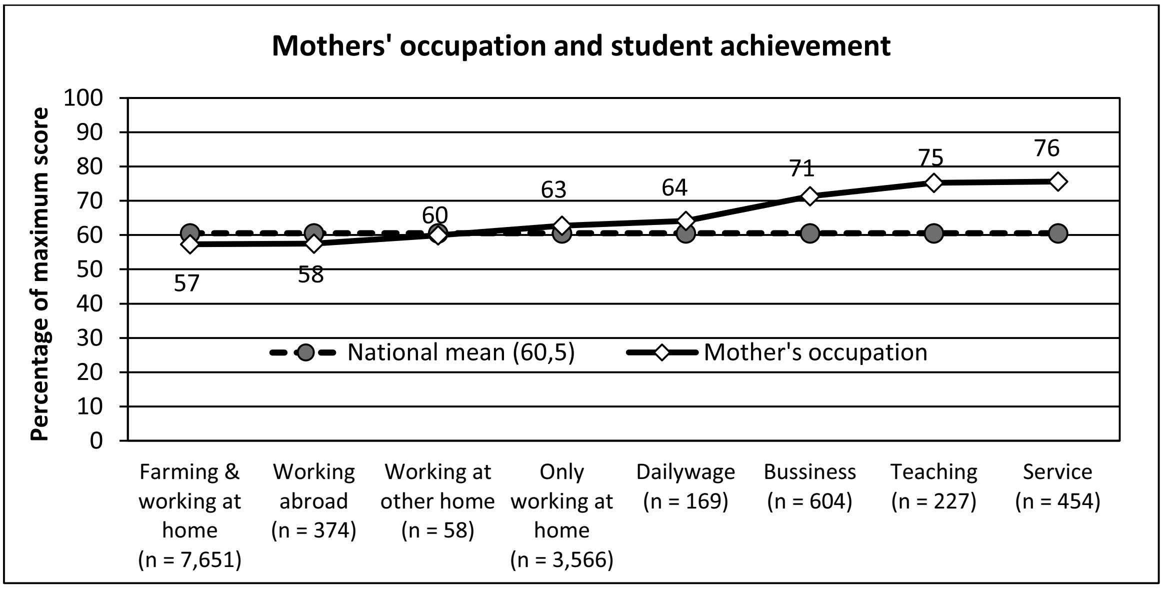

Figure 4.2.21 Mothers' occupation and students' achievement

When it comes to the fathers' occupation, on the basis of DTA (figure 4.2.22), the lowest achievement is found among the children when the father works in farming and working at home (55%). Even lower means are found for the children whose fathers are working as daily wage earners or a work at others homes (51%) (compare figure 4.2.20). Significantly higher achievement is seen for those whose fathers were in service (73%).

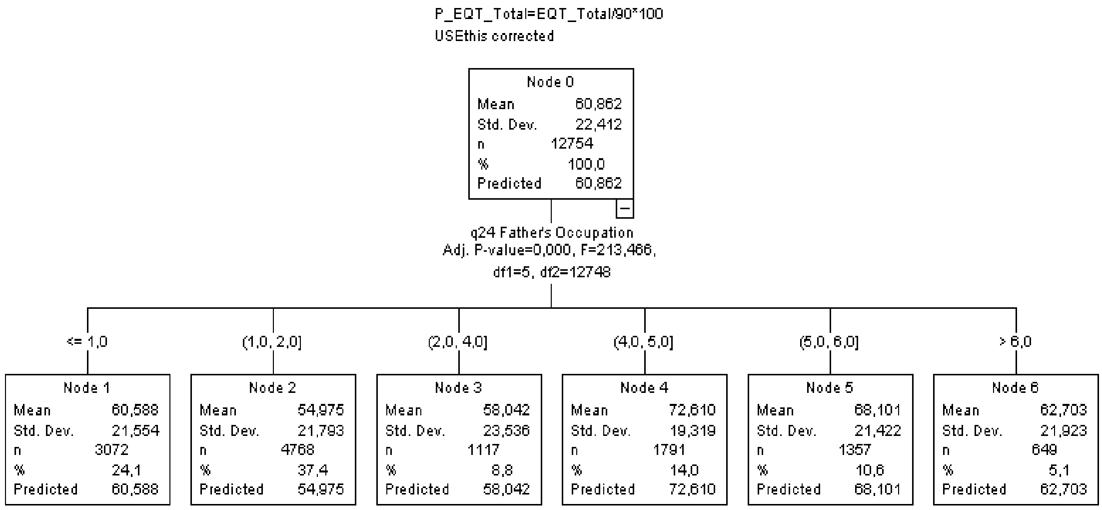

Figure 4.2.22 DTA of fathers' occupation and students' achievement

More precisely, if the father works either only at home (probably unemployed) or is a farmer, the Nepali skills of their children are remarkably lower (51 or 55%) compared to the children whose fathers are involved in more or less education requiring occupations such as business (68%), teaching (70%) or service (73%) (figure 4.2.24). The difference between the lowest and highest group is 22 percent which is a wide gap. Fathers' occupation explains 9 percent of the student variation (η^2= 0.092), which indicates moderate effect size (f = 0.32).

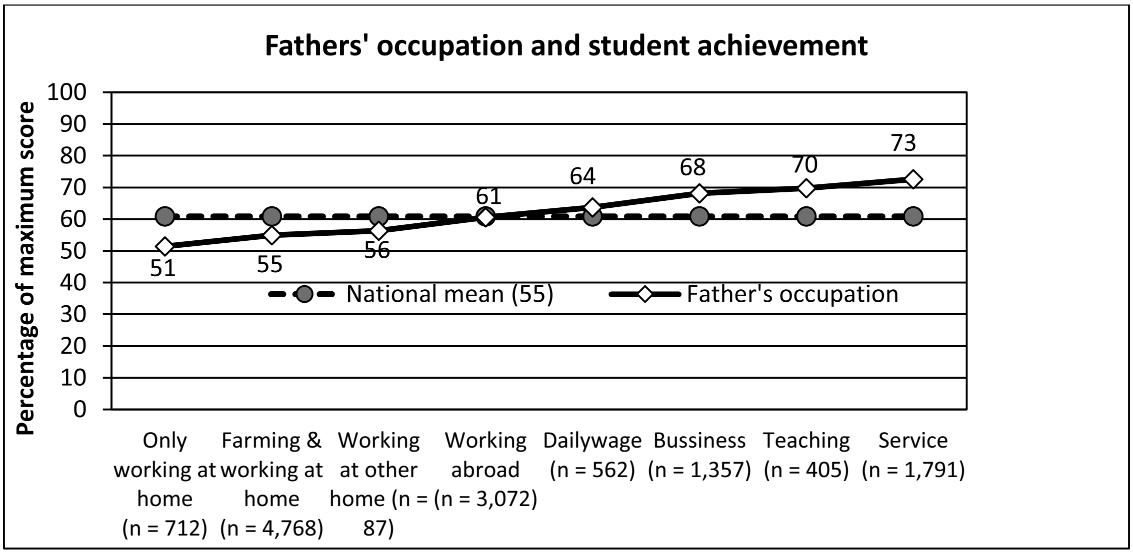

Figure 4.2.23 Fathers' occupation and students' achievement

When combining the mothers' and fathers' occupation, DTA shows that the lowest student achievement is found in the families where fathers' occupation is not known (47%), father comes from agriculture related occupations and mothers' occupation is not known (38, n = 57) or mother works abroad (51, n = 100). The highest achieving students come from the families where both father and mother are service holders or father is a teacher and mother works only at home (72–78%). It is worth noting that the latter occupations require academic degrees. For the later use as a SES indicator, the cut-off for parents' occupation was made so that being in the agriculture gives 0 and all other options give 1.

Home possessions and accessories

Facilities and resources available in home have also been considered whether they have some effects on the achievement. There were two kinds of home possessions defined in the background questionnaire for the students. One is related to the facilities that help in studying at home: 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 and poetry books, artistic things like pictures, and books that help them for study such as a dictionary. Another type of home possession includes different types of normal home accessories (and hence, in what follows these are called home *accessories* to differentiate them from home *possessions*) such as the number of mobile phones, televisions, and computers.

There are 11 questions in the student background questionnaire related to home possessions. Each was scored 1 if the student had an access to that possession (e.g. having a separate room or a table for study). Adding up these items, the maximum score was 11 indicating that the student reported that they have accessed to all possessions, explaining that the lower the score, the fewer possessions they have at home. Figure 4.2.24 shows the connection of home possessions and achievement level. Except for the highest category, the students' achievement level raises logically when there is a greater access to home possessions.[51] Pearson correlation between the achievement level and the factor (r = 0.19) is statistically significant (p < 0.001) and indicates a medium effect size (d = 0.42).

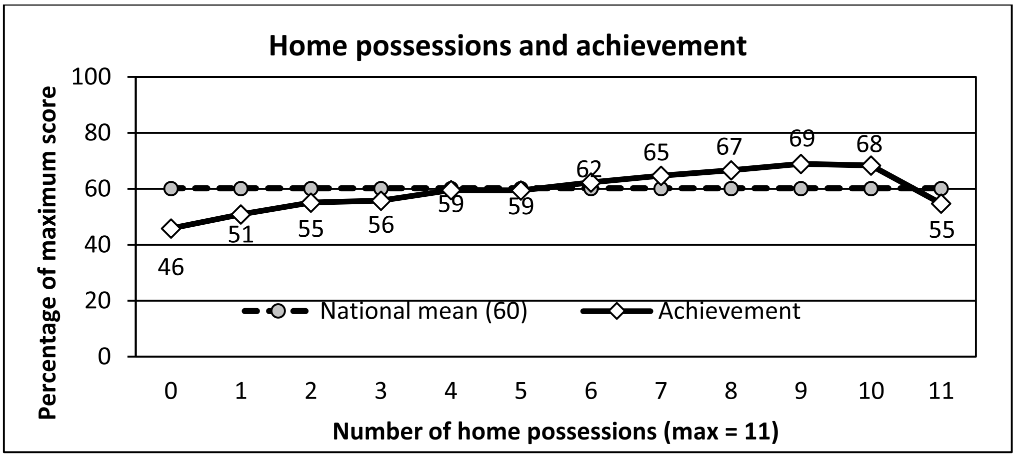

Figure 4.2.24 Relation between home possessions and achievement

[51] The same phenomenon, though not as radical as here, was seen also in 2011 datasets (see, ERO, 2013, figure 3.4.24 and 3.4.22): the students who selected all the possibilities may not have understood the question in the same way as the other students. Most probably, in any case, they actually did not have all the possessions though they claim that.

The cut-off for the SES factors was set on 6 possessions: if 6–10 items mentioned in background questionnaire were met, the student was given 1, otherwise 0.

The same pattern – the more possessions, the better results – can be seen also with home accessories, as seen in figure 4.2.25. The question in the background questionnaire was set differently compared to home possessions; with the accessories it was asked "*how many of the following accessories do you have in your family?"* with the options 0–3 (or more). For the indicator, the availability of home accessories is dichotomized in the same way as in home possessions. After dichotomizing the items individually by using meaningful cut-offs found with ANOVA and DTA (and maximizing the differences in achievement level, see table 4.2.25), all three indicators were summed up.[52] The maximum score was 3 indicating that the student possessed at home a set of the accessories.

Table 4.2.25 Dichotomizing the indicators for home accessories

Accessory	Cut-off for 1	Cut-off for 0
Mobile phone	2, 3	0, 1, missing
Television	1–3	0, missing
Computer	1–3	0, missing

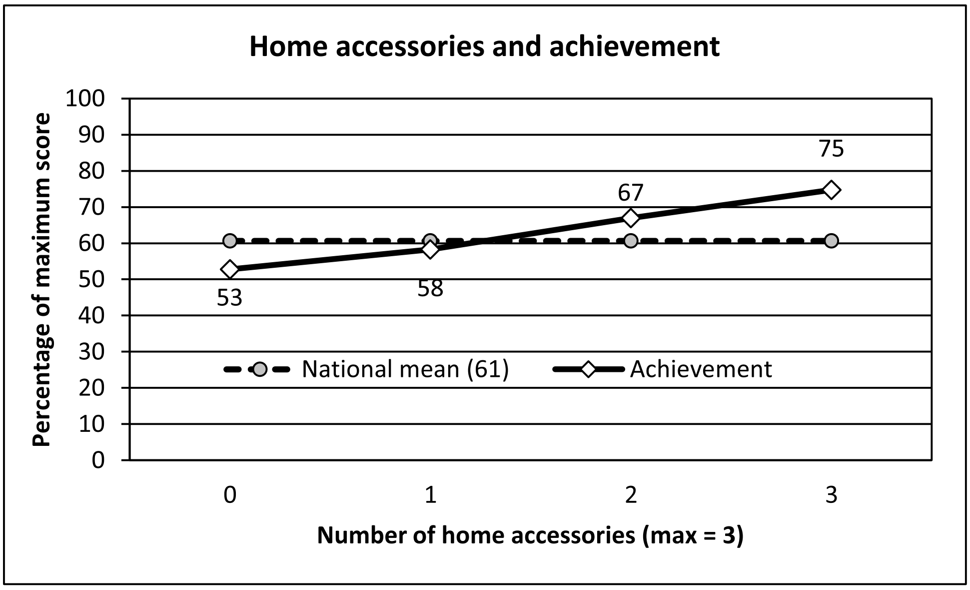

Figure 4.2.25 Relation between number of home accessories and achievement

Figure 4.2.25 shows how increase in the number of home possession or accessories increases students' achievement from 53% (if none of them are available) to 75% (if all three of them are available). Availability of all the stated facilities indicates the higher SES of the family. In comparison with the families where none of the accessories were met, the score is higher with + 12 percent units (53%) for the children meeting at least two of the

52 There was also fourth item in the questionnaire – the number of radios in home. However, this item was behaved pathologically in the analysis: the more the radios at home the less the achievement. Hence, it was not taken as an indicator of SES.

accessories. Correlation between the number of home accessories and achievement is r = 0.27 (p < 0.001), which is certainly positive and indicates moderate effect size (d = 0.62).

For the later use in SES, the cut-off for the factors was set on 2 accessories out of 3: if 2–3 items were met in the background questionnaire, the student was given 1 otherwise 0.

The dataset shows that parents' educational level strongly predicts the children's future achievement level in Nepali. The achievement level is very low where either the father or mother or both are illiterate. Around 42.0% of the students had an illiterate mother and 19.3% had an illiterate father.

The dataset also suggests that either economic or intellectual capacity or the both at home help children to improve their Nepali proficiency. If the father or mother or both are coming from an agricultural or daily wage related occupation, the students' achievement in Nepali is significantly lower than with the other occupational groups. Overall, 58.4% mothers and 37.4% fathers worked in agriculture or only at home.

It is also evident that when children have very few home possessions (zero to three out of the 11), the achievement level is remarkably lower than the national average (< 56). With nine to ten possessions, the average score is very high (up to 69) compared to the national average. The same is true of home accessories: When none or at least two accessory indicators out of three are available, the results are lower than average (40–52); and when there are more than two, the result is remarkably higher (64). Of the total students 3.5% of the students did not have any of the home possessions and 32.7% had no accessories.

SES and Achievement

The socio-economic status was formed on the basis of seven indicators which were all first dichotomized. The variables (mothers' education, fathers' education, mothers' occupation, fathers' occupation, home possessions, home accessories, and type of school where students were studying) were summed (as SES) and changed into the percentage of the maximum score (P_SES). Deeper description of the transformations is seen in section 2.5. The P_SES represents the percentage of SES of the student's family, where 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 are positive). Here, 0 refers to the lowest possible SES (that is, all the seven indicators of SES are negative). The analysis of the P_SES by using Univariate GLM (that is, the Regression modeling) shows the strong relation between SES and achievement. Figure 4.2.26 presents the relationship between SES of the students and the achievement.

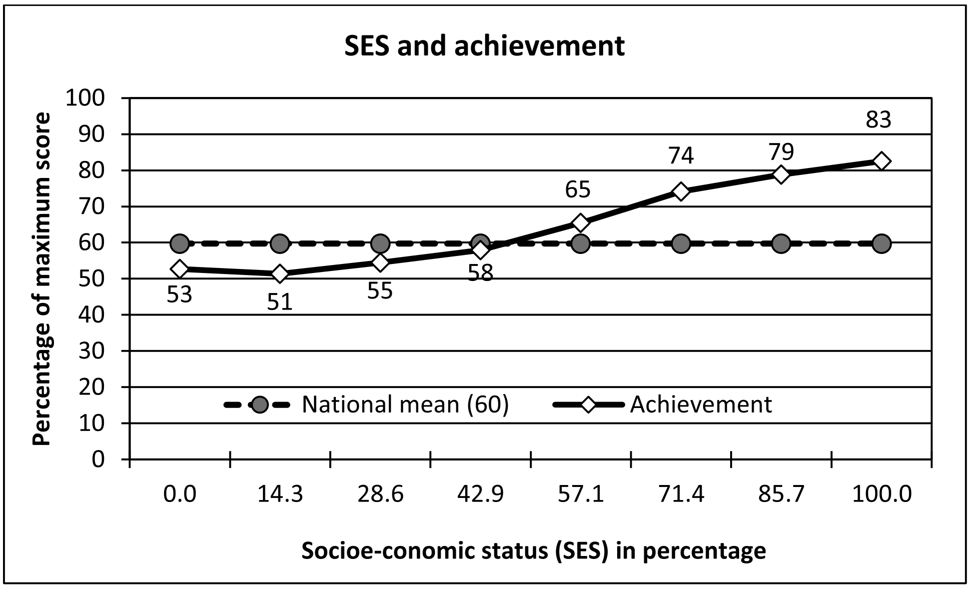

Figure 4.2.26 Relation between SES and achievement

Figure 4.2.26 shows a firm and positive relationship between SES and the Nepali achievement; the higher the SES, the higher is the achievement. Pearson correlation between the variables is $r = 0.41$ which is a high value ($p < 0.001$) and indicates high effect size ($d = 1.07$). The difference in achievement between the lowest SES groups (51–53%) and the highest one (83%) is remarkable. SES explains about 18% of the student variation ($\eta^2 = 0.18$) which also indicates a high effect size ($f = 0.47$).

From sociological viewpoint, it is interesting to know which of the individual indicators of SES are not met in those families where the children perform the lowest. Figure 4.2.27 illustrates the fact that in the families meeting less than four SES indicators, the challenge lies mainly in three factors marked in figure 4.2.27 with dark circle, triangle, and square: both mothers' and fathers' education is low and the child does not attend the private school. The adult education is possible to tackle with a good education policy.

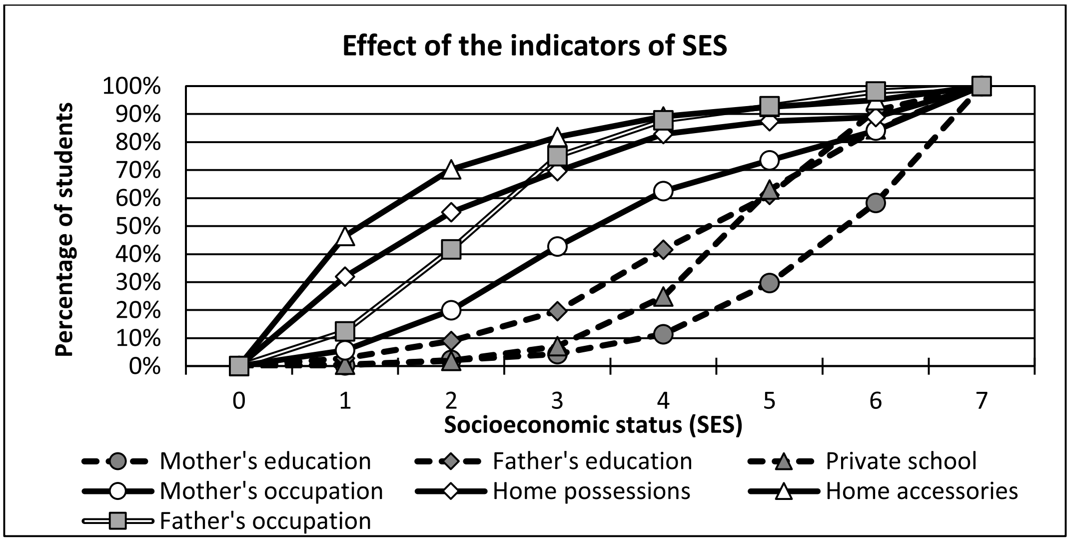

Figure 4.2.27 Effect of individual SES indicators in achievement

The dataset is evident that the socio-economic status plays a strong role in the educational processes in Nepal. The difference between the lowest and highest SES groups is remarkable (30 percent). This means that if the socio-economic standard of the lowest performing students rises up to a decent level, meeting the four out of seven indicators, the results in these groups will improve remarkably. Especially, challenging is the situation in the families where father or both parents are illiterate or they both work in agriculture or in daily wage work. The data sets show that 7.6% students are at the lowest level of SES.

Working beyond the school hours and achievement

Though several questions related to the students' activities outside school were set in the student background questionnaire, only two of them are briefly handled here: Working beyond the school time for a paid job, and involving in household work. The values of the variables are divided into five categories: 0 (no time at all), 1 (less than 1 hour per day), 2 (1–2 hours per day), 3 (2–4 hours per day), and 4 (more than 4 hours per day).

The DTA indicates that the cut-off is set on whether the students work for a paid job or not. The DTA shows that when the children do not need to involve in paid job at all, the results are statistically higher (65%) in comparison with the situation where they need to work for the paid job (55%). The difference is remarkable and the effect size is moderate (f = 0.24). The same logic is seen both in the community and institutional schools (fig. 4.2.28). If the students work for a paid jobs (even if it is less than one hour), the results are statistically significantly lower than the average. The ANOVA shows that the relationship is strict ($p < 0.001$) though effect size is moderate (f = 0.21) when students need to be engaged in paid work before and after school. It is notable, though, that most of the grade 5 children do not need to be engaged in paid work. Working after school indicates that the family is poor and the extra incomes are needed for living. It is obvious that when a student needs to work for more than 4 hours per day, there is no time or energy to study or to handle school tasks properly. Within the institutional schools, the difference between the children working over 4 hours per day (68, n = 48) is notably lower than those who do not at all need to work for paid job (80, n = 2,172) (see figure 4.2.28, institutional schools).

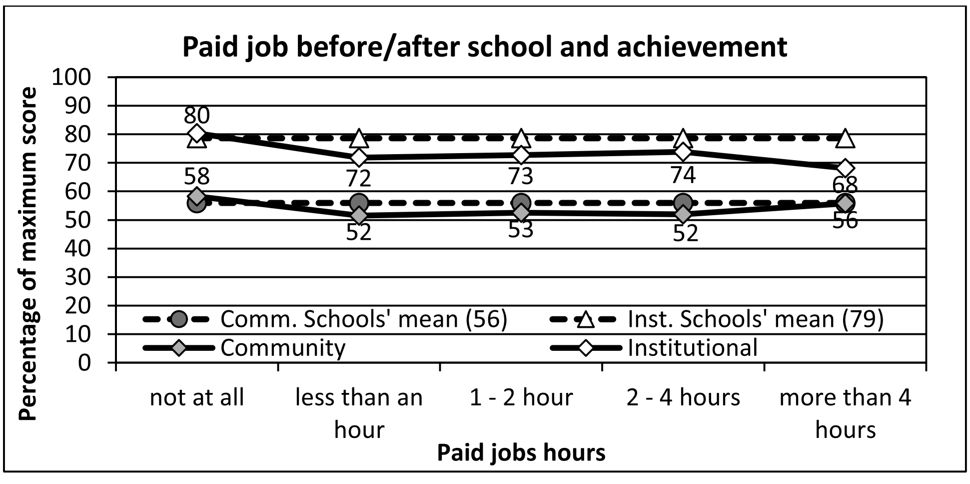

Figure 4.2.28 Relation between achievement and paid job beyond school time

When it comes to the involvement in unpaid household work, it is usual practice in families that the children take part in household chores at home as it is part of the socializing process of the children. The DTA shows that when the child spends time up to two hours for the household chores, the results are statistically higher (63%) than those who do not spend at all (56%) or more than 4 hours per day (57%). The effect of not participating in the household chores is more pronounced in the community schools than in the institutional schools (figure 4.2.29). It is seen that in the institutional schools, it does not make any difference whether the children work for two hours or less or not at all, whereas the effect is seen the same as when spending four or more hours in chores. Within the community schools, the results are significantly lower if children do not participate in the chores. Differences are significant ($p < 0.001$) though the effect size is small or moderate ($f = 0.16$ in the community schools and $f = 0.10$ in the institutional schools). It is noteworthy that more than 9% of the students (n = 1.074) reported that they spent more than 4 hours per day for household work. In the rural areas, this will be obligatory to be involved in stock raising when the cattle is far from home. It is self-understood that, in these cases, there is not much energy to decently concentrate on their study or on school work.

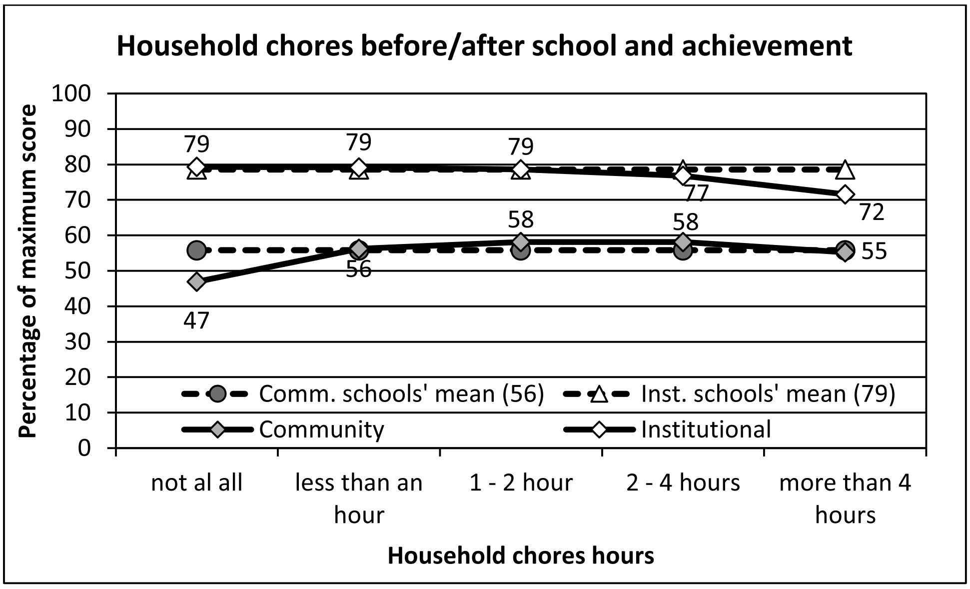

Figure 4.2.29 Hours in the household work and achievement

While looking from gender perspective, boys are found performing comparatively lower than girls even though they devote the same time in the household chores in both community and institutional schools (figure 4.2.30). The difference in the institutional schools is somehow higher than in the community schools. However, if students engage for more than two hours, girls outperform the boys by 5–6 percent in community schools and 7 points in institutional schools.

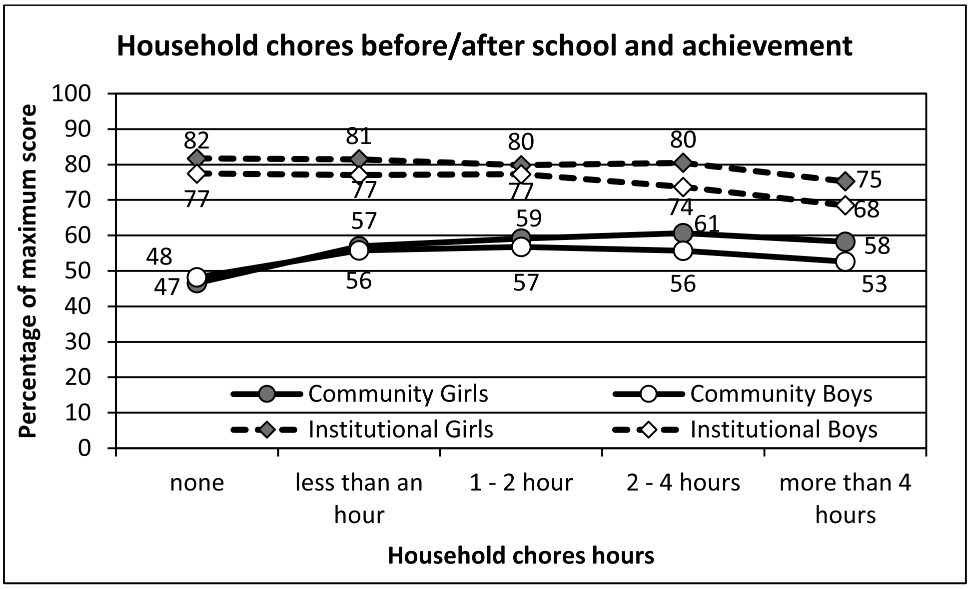

Figure 4.2.30 Household work and achievement by gender

The dataset indicates that either working for a paid job or for four hours per day in an unpaid household work outside school lowers the achievement of the student. However, a decent amount of household work up to two hours per day does not hamper the learning of the students in Nepali. The data shows that 35.7% students worked for paid job and 23.0% spent more than 2 hours in household chores.

Attitude and achievement

In the context of assessment of Nepali language achievement, the attitude tells us what the students think about Nepali and its usefulness in their daily life and future. There is a more or less firm relationship between the attitude of the students and achievement. Though the relation is not always clear, the correlation between achievement and attitude towards the subject is widely studied (see, for example Metsämuuronen 2012a; 2012b; House & Telese, 2008; Shen & Tam, 2008; Kadijevich, 2006; 2008). Some researchers have noticed remarkable differences in correlation between countries (e.g. House & Telese, 2008; Kadijevich, 2006; 2008; Wilkins, 2004; Shen, 2002; Papanastasious, 2000; 2002; Stevenson, 1998). In some countries, the correlation between attitude and achievement has been found near to zero, like in Macedonia (Kadijevich, 2008),in the Philippines (Wilkins, 2004), in Indonesia (Shen, 2002), or in Moldova (Shen, 2002) whereas in some other countries like Korea, for example, the correlation has been found as high as 0.60 (Shen, 2002). In NASA 2011, it was noticed that grade 8 students were not consistent in the attitude test and the reliability of the international test remained low (see ERO, 2013, table 2.11).

In NASA 2012, 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. The original scales included nine dimensions but in these international comparisons only three are used with four items on each dimension and two negative items on each of the first two dimensions. 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). Because of students' inconsistent manner in answering the attitude scale in NASA 2012, only the dimension of "Experiencing utility in Nepali" was taken into the measurement instrument of grade 5 students. Reliability of the score of five items is sufficient (α = 0.72). The relation between the attitude (divided into seven groups with sapproximately equal number of the students, that is, septiles[53]) and achievement score is shown in figure 4.2.31.

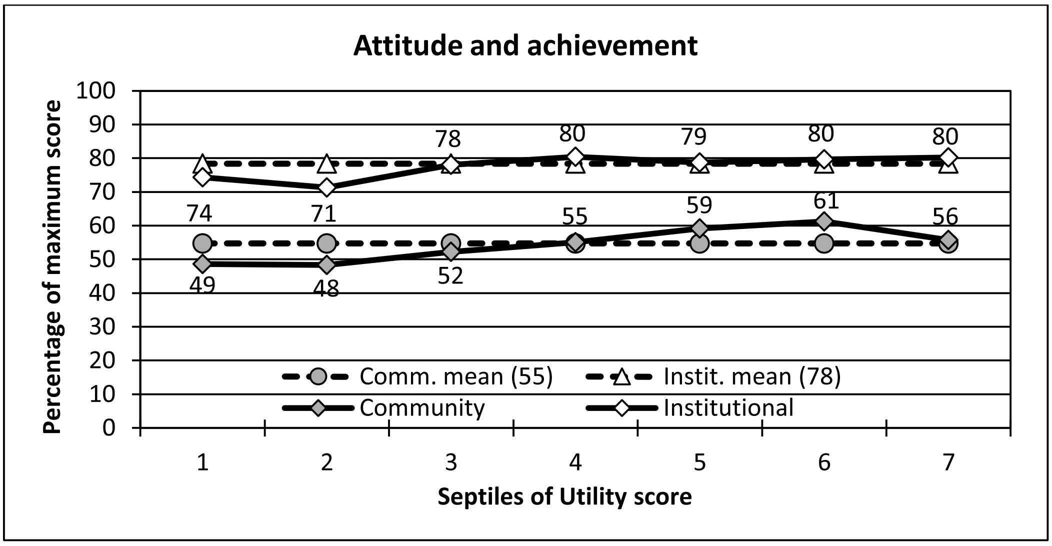

Figure 4.2.31 Relation between attitude and achievement

There is a positive, but small correlation between the attitude and Nepali achievement in the whole dataset (r = 0.12, d = 0.26). In the whole dataset, the division of attitude into seven groups explains the achievement level somehow 5% ($\eta^2 = 0.047$). This means that, according to ANOVA, the difference between the means of the lowest attitude group (51%) and highest one (66%) is remarkable (f = 0.22). The relation is moderate in both institutional and community schools. The difference between the lowest and highest attitude group is 7 percent in community schools (f = 0.20) and 8 percent points in the institutional schools (f = 0.19).

The connection of utility and achievement in Nepali is somehow clear though it is unknown whether the positive attitude is a consequence of high achievement or not. From the statistical point of view, on the basis of simple ANOVA GLM procedure, attitude explains the achievement 4.7% while achievement explain attitude 2.4%. Hence, it is more probable that, in Nepal in grade 5 Nepali subject, better achievement is a consequence of positive attitude rather than the other way round.

[53] The original score is small (maximum was 15 points) and quite many students (34%) gave the maximum score. Hence, it was not possible to form more precise classification such as deciles. Seven classes (septiles) was the most precise alternative with the given dataset.

The dataset informs that positive attitude towards the subject correlates with positive achievement in Nepali. The better achievement is more probably a consequence of more positive attitude rather than the other way round.

Student age and achievement

Though the appropriate age of children to study grade 5 is 10-12 years, in the context of Nepal, the age of the students attending to grade 5 studies varies widely. Some students have reported their age below nine years and some have above 16 years. All the ages of the students below 10 were encoded as ‘up to 9 years’, and all students above 14 were encoded as ‘15 years or above'. The descriptive statistics of the mean in each year are given in tables 4.2.26 and 4.2.27 and in figure 4.2.32.

Table 4.2.26 Students’ achievement in different age groups

Age	N	Mean	SD	CV
Up to 9 years	439	53.0	25.2	47.5
10 years	2,039	57.4	23.7	41.3
11 years	3,738	62.5	22.9	36.6
12 years	4,224	61.7	22.7	36.8
13 years	1,842	58.4	21.9	37.6
14 years	754	56.6	22.3	39.5
15 years or above	505	48.1	22.2	46.2
Total	**13,541**	**59.7**	**23.1**	**38.6**

Table 4.2.27 Student achievement in different age groups by the type of school

Age	Community				Institutional			
	N	Mean	SD	CV	N	Mean	SD	CV
Up to 9 years	391	50.4	24.9	49.3	48	74.1	16.4	22.2
10 years	1,701	53.0	22.8	43.0	338	79.1	14.2	18.0
11 years	2,635	55.1	22.0	40.0	1,103	80.3	12.9	16.0
12 years	3,116	55.7	22.0	39.5	1,108	78.5	15.0	19.1
13 years	1,518	55.2	21.5	39.0	324	73.3	17.0	23.2
14 years	670	54.6	22.3	40.8	84	72.6	15.1	20.8
15 years or above	491	47.6	22.2	46.6	14	65.6	14.6	22.3
Total	**10,842**	**54.4**	**22.3**	**41.0**	**3,129**	**78.4**	**14.6**	**18.8**

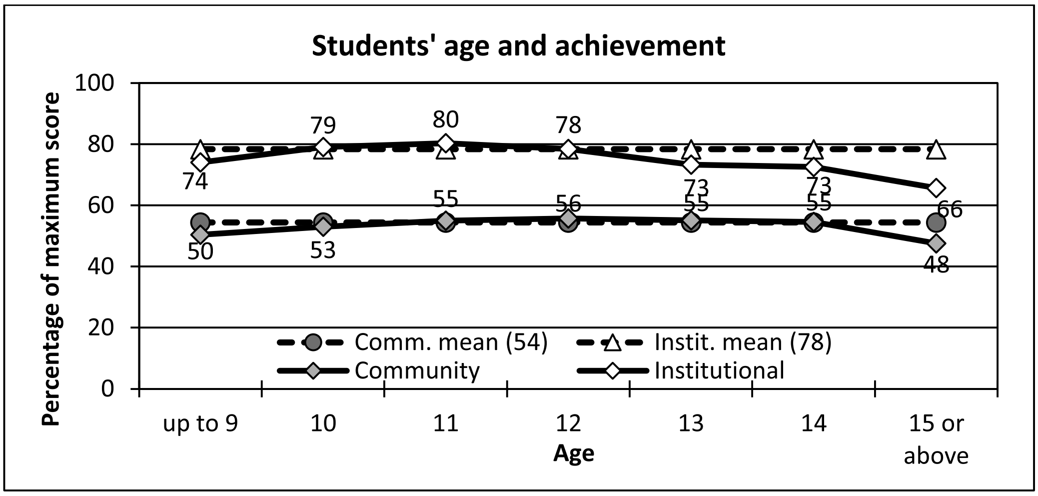

Figure 4.2.32 Relation between age and achievement

It is seen evident that the best achievers are those students who are at their proper age for grade 5 studies (10 to 12 years old) having scored 53–56% in community schools and 79–80% in institutional schools. The higher the age is – meaning that the students have either started their schooling much later than they should have, or they have repeated the earlier or same grades – the lower ability level is evident. The achievement level is remarkably lower than the average when the students are above 15 years or higher in both the community (48%) and institutional schools (66%). However, in institutional schools, this scenario is evident even from the age of 13 or above (65–72%). Otherwise, the correlation between the age and achievement in community schools is low ($r = –0.003$) with no actual connection ($d = 0.006$). In institutional schools the correlation is stronger ($r = –0.116$; $p < 0.001$), though, indicating still a small effect size ($d = 0.25$). An obvious reason for the zero correlation is the curvilinear shape of the phenomenon.

Dataset suggests that the highest performance in Nepali is found among the students studying with their proper age group, that is, at the age of 10 to 12 years. Otherwise, the achievement is found to be going lower as the age increases. As many as 26.1% of the students fell aside 10–12 years.

Support for the study and student achievement

The relation between the support received for studies and achievement was analyzed based on the responses provided to question as: *"who supports you when you do not understand what you have read?"* In the question, only one option was selected, in many cases, there would have been several support providers, which are not taken into account. The descriptive statistics of the support received are given in tables 4.2.28.

Table 4.2.28 Support for the study and achievement level

Support received	N	Mean	SD	CV
No one	338	63.3	22.2	35.1
Tution	1,220	63.3	23.0	36.4
Brother/Sister	6,061	61.7	22.1	35.8
Teacher	2,102	59.2	20.9	35.4
Mother	887	59.2	24.1	40.7
Father	2,567	56.9	24.5	43.2
Total	**13,175**	**60.4**	**22.7**	**38.6**

It is found that outperformers in Nepali from community schools are the independent learners who do not receive any kind of support (58%). The achievement of independent learners is 3 percent more compared to those who received private tuition (55%). The support received from siblings (57%) is found to have been more effective in children's learning than parents (51%) in community schools. The effect of the support received is, at most, medium. Effect size is $f = 0.2$ in community schools and $f = 0.15$ in institutional schools – indicating that the difference in scores of the extreme groups is not high.

The whole dataset modestly suggests that self-learning and tuitions raise the achievement level more than the support received from parents and teachers. Brother/sister also contributes significantly in achieving better scores. However, the difference between the highest and lowest performing groups is not notable. It is possible that the group receiving private tuition also spends more time on their homework, explaining higher score.

Availability of textbook and student achievement

The data shows that there are still some students who do not have the proper textbook up to the end of the academic session. Table 4.2.29 shows the descriptive statistics of availability of the Nepali textbook and the achievement.

Table 4.2.29 Availability of textbook of Nepali and the achievement

Availabilityof textbooks	N	Mean	SD	CV
Yes	12,574	61.1	22.3	36.5
No	480	47.4	24.9	52.6
Total	**13,054**	**60.6**	**22.6**	**37.2**

Out of 13,054 students who responded to the question, 3.7% (4.0% in community schools and 2.7% in institutional schools) did not have a textbook. The relation between the textbook and achievement is significant ($p < 0.001$) though the effect size in the whole dataset is small ($f = 0.11$). The difference in achievement is 13.7 percent (12.4 in community schools and 12.0 in institutional schools).

The dataset shows that 3.7% students lack the textbook in Nepali even up to the end of academic session. The achievement level of these students is significantly lower than those who have access to the textbook.

Homework assigned/checked and achievement

The results regarding the homework is based on the related responses obtained from the students. Although some deviant responses are also found as some reported to have received and other not having it within the same class, in the wider scope, the results are found enough to make sense. Statistics related to homework assigned and checked is presented in tables 4.2.30 and 4.2.31 and then shown in figure 4.2.33.

Table 4.2.30 Homework given/checked and the achievement

Status of homework	N	Mean	SD	CV
Given Everyday- Checked Someday	1,618	62.2	22.8	36.7
Given Someday-Checked Everytime	897	61.9	22.4	36.1
Given Everyday-Checked Everyday	9,213	60.7	22.6	37.2
Given Someday-Checked Someday	978	58.8	21.8	37.1
Given Everyday - Not checked	141	55.0	22.6	41.1
Not given	156	46.3	22.8	49.3
Given Someday - Not checked	82	43.1	23.4	54.3
Total	**13,085**	**60.5**	**22.7**	**37.4**

Table 4.2.31 Homework given/checked by the school type

Status of homework	Community				Institutional			
	N	Mean	SD	CV	N	Mean	SD	CV
Given Everyday–Checked Everyday	7,436	56.2	21.9	38.9	1,777	79.6	14.1	17.7
Given Everyday–Checked Some day	1,079	54.5	22.3	40.9	539	77.6	14.6	18.8
Given Someday–Checked Eachtime	591	54.2	21.3	39.4	306	76.9	15.7	20.5
Given Someday–Checked Someday	694	51.2	20.1	39.2	284	77.3	13.0	16.8
Given Everyday–Not checked	108	50.2	21.5	42.7	33	70.4	19.2	27.2
Not given	124	42.0	22.0	52.4	32	63.1	17.8	28.2
Given Someday–Not checked	67	37.8	20.7	54.7	15	66.9	20.4	30.5
Total	**10,099**	**55.2**	**21.9**	**39.7**	**2,986**	**78.4**	**14.6**	**18.6**

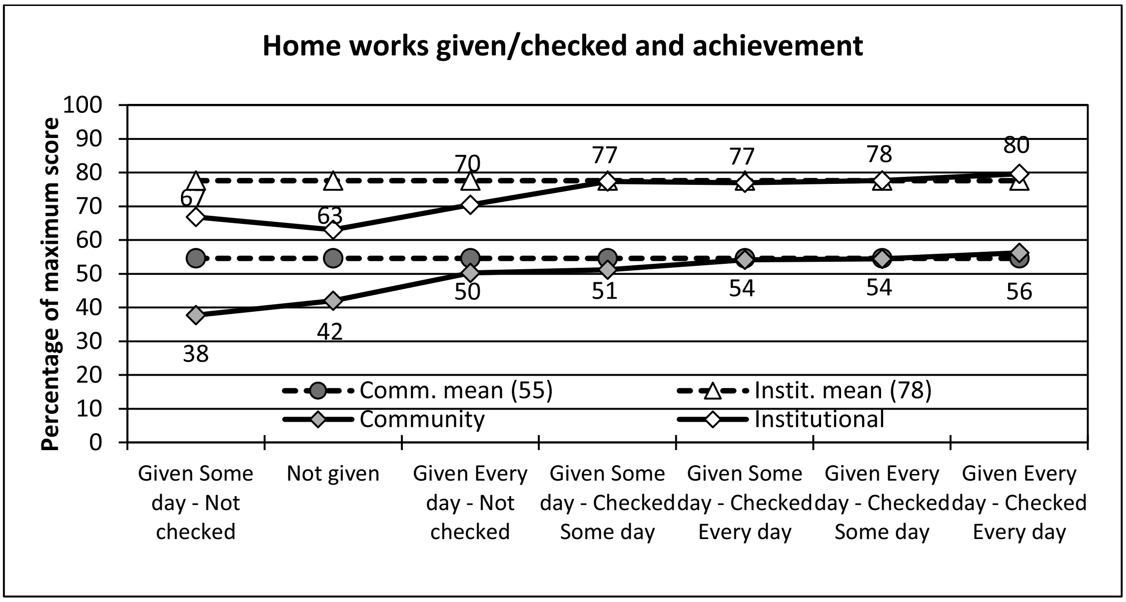

Figure 4.2.33 Relation between homework given/checked and achievement

It is evident that the achievement was notably lower (42–50% in community schools and 63–70% in institutional schools) for the students not receiving or not getting the homework checked compared to the students having received and got it checked every day (56% in community schools and 80% in institutional schools). The differences are statistically significant ($p < 0.001$). In Nepali, homework is seen to be the most regular activity. Thus, those groups not receiving and not getting it checked is found very small and hence, the effect size is also small (f = 0.11 for community schools and f = 0.16 for institutional schools). However, the grouping explains around 13% of the variance in the data (η^2= 0.013 for community schools and η^2= 0.025 for institutional schools).

The dataset strongly suggests that if the teacher assigns and checks homework systematically, the achievement becomes higher than those without having it. By assigning and checking homework daily though not every day, the students are likely to raise their scores up to 15 percent. In grade 5 Nepali, 2.9% of the student either did not get homework or not get it checked.

Future aspiration of the student and achievement

The future aspiration of by the students can encourage in studies or, in some cases, when knowing that the future aspiration does not require long studies, the motivation for hard work in school is likely to decline. The student's future aspiration was divided in eight categories: (1) farming, (2) business, (3) teaching, (4) government job, (5) going abroad, (6) engineer, (7) doctor, and (8) other. Future aspiration is found to have been connected strictly with their achievement, which is presented in tables 4.2.32 and 4.2.33 and figure 4.2.34.

Table 4.2.32 Students' future aspirations and achievement

Future aspiration	N	Mean	SD	CV
Farmer	555	41.4	22.9	55.3
Business	487	50.4	23.1	45.4
Teacher	3,976	56.2	21.7	38.6
Government Officer	1,024	58.0	22.9	39.4
Going abroad	812	59.6	23.2	38.9
Engineer	2,115	62.7	21.6	34.6
Doctor	3,825	66.9	20.9	31.4
Other	413	74.8	18.4	24.3
Total	**13,207**	**60.4**	**22.7**	**72.0**

Table 4.2.33 Relation between Students' future aspirations and achievement by type of school

Future aspiration	Community				Institutional			
	N	Mean	SD	CV	N	Mean	SD	CV
Farmer	524	40.6	22.3	55.1	31	56.2	27.6	49.1
Business	442	48.3	22.0	45.5	45	71.6	23.8	33.2
Going abroad	539	51.1	21.7	42.5	273	76.5	15.6	20.4
Teacher	853	54.2	22.5	41.4	171	76.9	14.4	18.8
Government officer	3,622	54.4	21.3	39.2	354	74.9	16.1	21.5
Engineering	1,565	57.1	21.0	36.8	550	78.6	14.1	17.9
Doctor	2,487	59.9	20.9	34.9	1,338	79.9	13.2	16.5
Other	166	64.3	21.7	33.8	247	81.9	11.3	13.8
Total	**10,198**	**55.1**	**21.9**	**39.7**	**3,009**	**78.4**	**14.6**	**18.7**

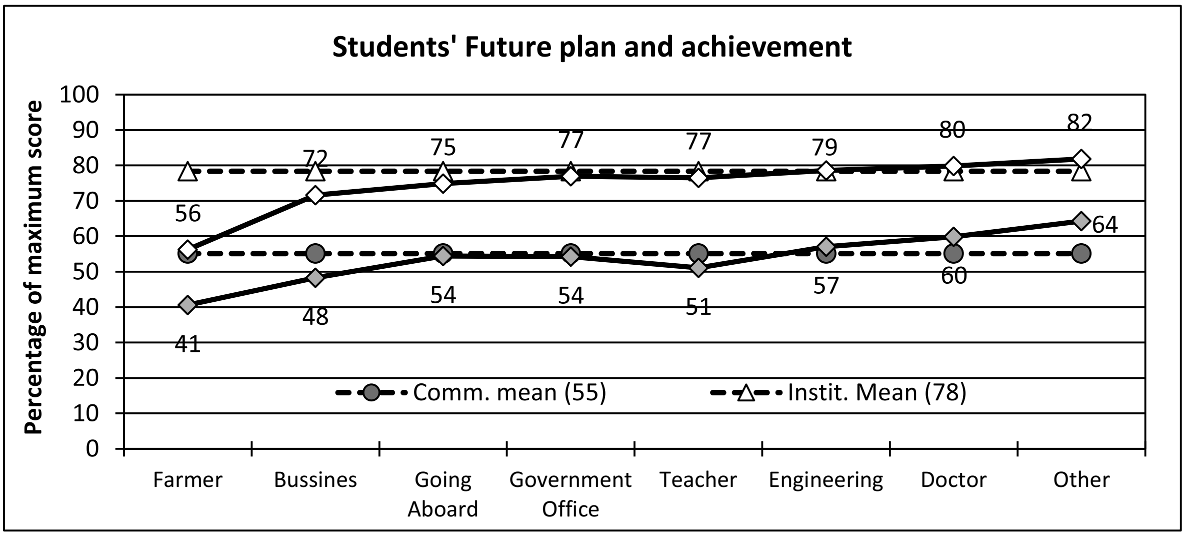

Figure 4.2.34 Relation between students' future aspirations and achievement

Some occupations, like Engineers, Medical Doctors, and Teachers, are favored much in the society – most probably because of higher economic prospects associated with. This can be seen in the fact that even the weakest students in the dataset (scoring less than 20% of the maximum score) are found to have been aspiring to these occupations. Of these, less number of students, 15% are aspiring to be the Engineer, 17% of them aspire to be the doctor, and 32% of them aspire to be the teacher. On the basis of their achievement level, this dream, most probably, will turn to reality for some of them. Another reality is that, according to the fathers' occupation, 37% students come from agricultural background, but in the whole dataset, only 4.1% are aspiring to join the agricultural occupation, farming. When the student knows that (s)he will be continuing the family occupation–farming, the learning achievement is remarkably low (40% in community school and 56% in institutional school).

Because the occupations such as Engineers, Doctors and Teacher are wanted the most, the competition for the study places is also tough. Hence, higher achievement level is desired in order to make the dream of the future occupation come true. From this point of view, students' future plans are seen logical while comparing it to the mean achievement level. Students who are aspiring to be Engineers or Doctors really score remarkably higher (63 and 67 respectively) than those aspiring to join Farming (41%) or Business (50%). The future plan explains 9% of the achievement level ($\eta^2 = 0.086$); the effect size is moderate ($f = 0.31$) indicating that the differences between the lowest and highest group is remarkable.

The dataset shows a connection between the aim of student and their achievement. As the students aspire to take a professional career other than farming or business, their achievement is higher than the average. The number of students who aspire to be a teacher, government officer, engineer, or a doctor is remarkably high.

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 activities of the students in the school that tends to affect learning. In the student background information questionnaire, several student and school related activities were asked, some of which are positive and some are negative. Here, bullying is handled as one of the negative indicators whereas students' positive impressions of school and teacher's activities are taken as the examples of positive indicators.

Negative activities - Bullying

Bullying is one of the problems in the school that tends to worsen the learning environment for students. International studies like TIMSS and PISA (background questionnaires) give a specific emphasis to study such phenomena. In NASA 2012 student questionnaire, five questions ask the varieties of bullying that tend to happen in the school. All the questions were stemmed from the phrase "*which of the following activities happened in your school in the last month?*" The students' responses are presented in table 4.2.34 and table 4.2.35 and visualized in figure 4.2.35. 'No (%)' indicates the percentage of the students' response that no such activity happened in the school and 'Yes (%)' indicates the percentage of the students who reported the particular type of bullying happened within the last month. Altogether 26% students mentioned that, during the last month, something of their own was stolen–which is an alarming sign for the system.

Table 4.2.34 Frequencies of encountered bullying

Type of Bullying	No(%)	Yes (%)
I was made fun of or called names	73.2	26.8
Something of mine was stolen	73.8	26.2
I was hit or hurt by other student(s)	78.4	21.6
Fellow students kept me outside without involving in activities	78.5	21.5
I was made to do things I didn't want to do by other students	85.4	14.6

Table 4.2.35 Bullying and the achievement by the type of school

Intensity of bullying	Community				Institutional			
	N	Mean	SD	CV	N	Mean	SD	CV
No bullying	4,703	57.9	21.2	36.5	1,310	80.2	13.4	16.7
20% bullying	2,353	56.3	21.2	37.6	842	78.8	13.4	17.0
40% bullying	1,510	54.3	21.6	39.8	459	75.4	17.0	22.5
60% bullying	864	50.2	22.0	43.9	283	76.7	15.4	20.1
80% bullying	349	46.7	22.7	48.6	74	71.9	17.6	24.5
100% bullying	350	34.9	21.7	62.3	22	65.6	25.5	38.9
Total	**10,129**	**55.1**	**21.9**	**39.7**	**2,990**	**78.4**	**14.6**	**18.6**

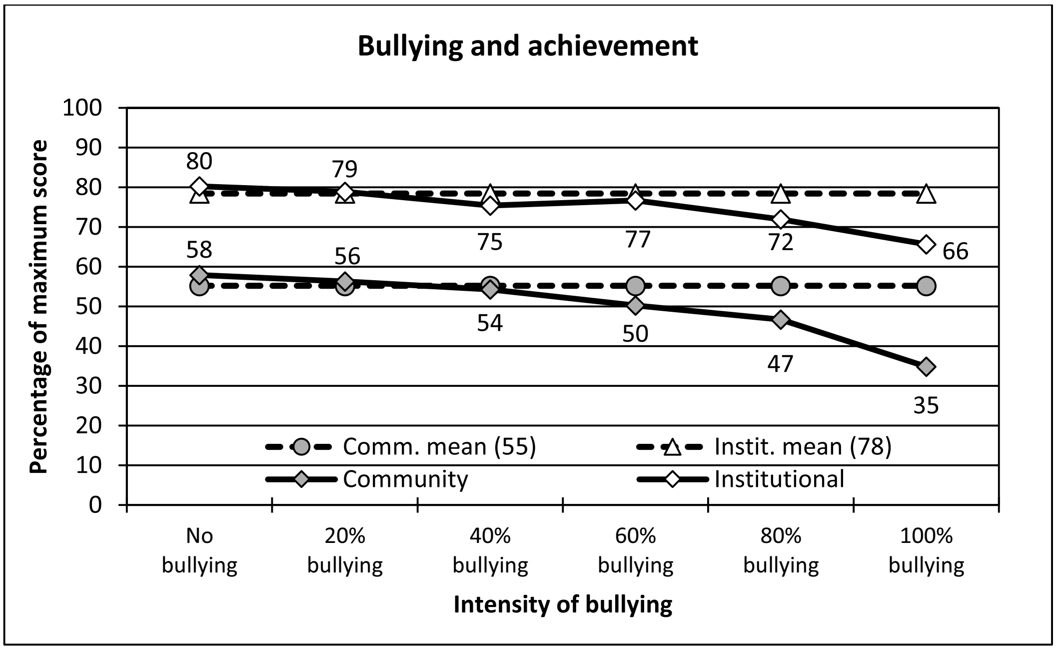

Figure 4.2.35 Relation between bullying and achievement

The sum total of all five items is taken as an indicator of bullying. Figure 4.2.35 shows the achievement of the students in each category of bullying. If only one activity of bullying was reported, it is categorized as 20% bullying, and if all five activities were reported it is categorized as 100% bullying. When knowing that 46% of the students did not encounter any bullying during the last month of an academic year, one can infer that the remaining 54% did encounter at least one type of bullying. This is a remarkable number of students. As many as 6.1% students – 6.9% in the community schools and 3.2% in the institutional schools – are experiencing a severe kind of bullying (the sum of 80% and 100% bullying). This means, in practice, that more than 44,000[54] grade 5 students in Nepal have been encountering physical, psychological, and social bullying every month. However, it is seen that learning outcomes are notably lower than the average with 15% of the students who have encountered more than two different types of bullying (34–50% in community schools and 65–76% in institutional schools). There is at most 26 percent gap in achievement between the students who do not experience bullying and those students who encountered extreme forms of bullying (four or five kinds). Though there are only 6.1% students who reported to have experienced this kind of bullying (n = 795), the difference is statistically significant (p = 0.001) and the effect size is medium or small (f = 0.22 in community schools and f = 0.16 in institutional schools). Though extreme cases of severe bullying are rare, bullying is seen to be quite common in schools. This negative phenomenon has been causing a needless harm to young children, so it has to be rooted out from schools.

The dataset shows that a high number of the students (54%) have encountered some kind of bullying in school within a month. Around 6% students – 6.9% in community schools and 3.2% in institutional schools – are experiencing severe kind of bullying. This means that more than 44,000 of grade 5 students in Nepal is encountering physical, psychological, and social bullying every month. The phenomenon is found to have been affecting the learning outcomes in almost all the groups of the students who felt bullying. Given circumstances urge to put all possible efforts to root the phenomenon out from the schools.

Positive activities in school

The activities that can boost the learning achievement of students are categorized as positive activities. Such positive activities about the school were asked to the students in two sets of questions collected. Table 4.2.36 shows the responses of the students in all four categories which are in the 4-point rating scale anchored to fully disagree (0) and fully agree (3). Generally speaking, the 5th graders expressed satisfaction with the school and student related activities in school. Interestingly, remarkably lower number of students (12.8%) expressed that the teacher is not treating them fairly. The phenomenon is also found in Mathematics of grade 5 (8%, see section 3.2. table 3.2.32), in English (9%, see section 5, table 5.32), and was also seen in 2011 datasets with grade 8 students: 11% of the Mathematics students, 12% in Nepali and 13% in Social Studies (see ERO, 2013).

[54] According to the “primary level total enrollment in all types of schools by district, Flash I_2012–2013”, there were 731,573 grade 5 students; 6.1% of these makes 44,333 students.

Table 4.2.36 Students' response towards teacher and school related activities

Teachers' and Students' activities[1]	**Respondents in % (valid percentage)**			
	Fully agree	**Partially agree**	**Partially disagree**	**Fully disagree**
q28a: I like to come and stay in school	91.6	5.0	1.4	2.0
q27b:Most teachers' are interested in student's well being	87.2	8.1	1.9	2.8
q28c: Teacher in the school care about the students	86.6	8.9	2.4	2.2
q27a: Students get along well with most teachers	85.6	10.7	1.7	2.0
q27d: If I need extra help, I will receive it from my teacher	85.6	9.5	2.3	2.6
q27c: Most of the teachers really listen to what I have to say	82.9	11.2	2.9	3.0
q28b: Students in my school like me	79.0	16.1	2.9	2.0
q27e: Most of my teachers treat me fairly	73.7	13.5	4.5	8.3
Average	**84.0**	**10.4**	**2.5**	**3.1**

1) *The activities are ordered on the basis of percentage in "Fully agree."*

Further analysis was carried out by recoding the variables into two categories (1–2 = 1, that is, agree and 3–4 = 0, that is, disagree). 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 achievement.

DTA finds three attitude groups in the indicator. These boundaries and descriptive statistics are seen in table 4.2.37 and illustrated in figure 4.2.36. 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 eight positive activities and achievement is positive (r = 0.196, p < 0.001) and effect size is moderate (d = 0.44).

Table 4.2.37 Percentage of positive response towards teacher and school related activities and the achievement

Percentage of positive actions	**N**	**Mean**	**SD**	**CV**
62.5 or lower	1,463	44.4	22.4	50.5
75.0	765	51.6	23.7	45.9
87.5	1,823	60.2	22.7	37.7
100.0	9,495	62.7	21.1	35.3
Total	**13,546**	**59.1**	**22.8**	**38.6**

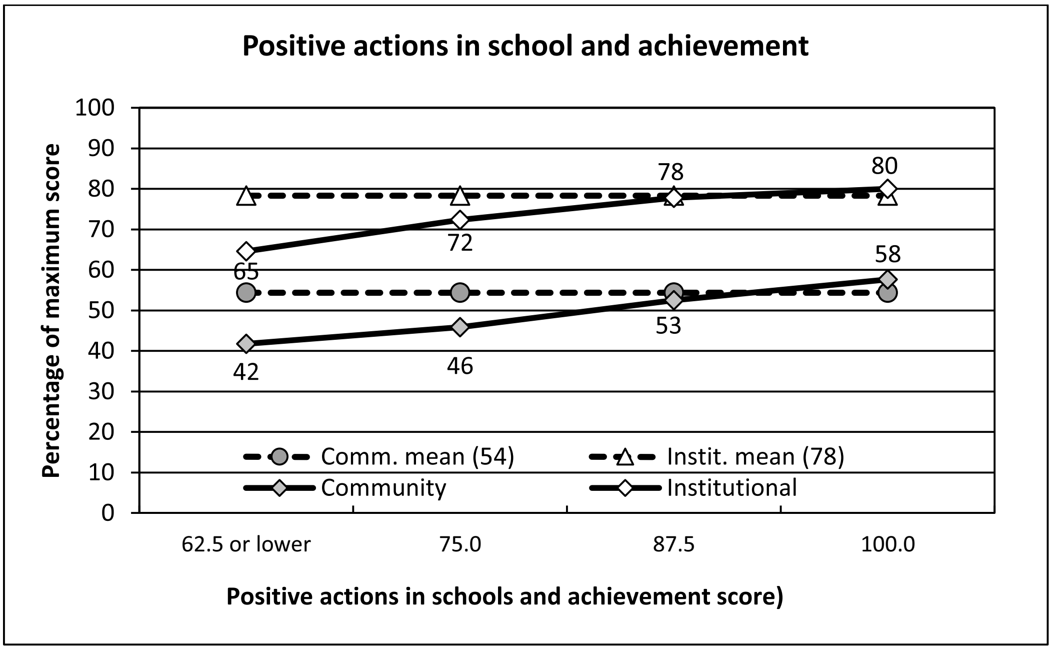

Figure 4.2.36 Relation between positive actions in school and achievement

The data shows that there is a positive relation between the students' positive feeling of the teacher and school related activities and the achievement. The increase in achievement is directly related to increase in the intensity of such activities. After dividing the indicator into four groups on the basis of DTA, the differences between the groups are statistically significant ($p < 0.001$). However, the effect size is moderate in community schools ($f = 0.26$) as well as institutional schools ($f = 0.27$); the difference between the most positive group and the most negative group is notable (16 percent). Only when the students are extremely positive towards school and teachers' behavior, the learning achievement is higher than the average. Students with a negative feeling in five or more of the eight indicators (62.5% of the attitude score) have achieved much lower (< 44) than the average in Nepali (59).

The dataset also informs that when the students feel the actions of the teachers and the schools are ultimately good, the Nepali results are better than average (58% in community schools and 80% in institutional schools). At the other extreme, in the case of feeling negatively of five out of eight actions, the results are far below the average (42% in community schools and 65% in institutional schools). As many as 12.8% students feel that their teachers do not treat them fairly.

4.2.4 Synthesis of the Analysis

Several individual student and geographically related factors have been presented above which individually explain the difference in achievement between the students. It is notable that, except for gender, all the factors showed statistically significant difference between the groups when analysed individually.

Table 4.2.38 Individual variables handled within the text and their effect in one-way ANOVA

Variable and values[1]	Leverage[2]	Eta squared[3]	Effect size[4]
Ecological zone (1 = Mountain, 2 = Hill, 3 =Tarai, 4 = Valley)	+24.71	0.139	0.40
Development Region (1= Eastern…5 = Far-Western, 6 = Valley)	+25.78	0.153	0.43
School location (0 = Rural, 1 = Urban)	+18.35	0.105	0.34
School type (0 = Community, 1 = Institutional)	+23.83	0.185	0.48
Gender (0 = girls, 1 = boys)	+1.71	0.001	0.03
Caste (1=Janjati, 2=Dalit, 3 = Madhesi,4 = Brahman, 5 = Chhetri)	+17.76	0.071	0.28
Language at home (1 = Nepali, …. 12 = Other)	+34.87	0.089	0.31
Mother's Education (1 = Illiterate, …. 8 = Above MA)	+25.74	0.057	0.25
Father's Education (1 = Illiterate, …. 8 = Above MA)	+25.61	0.075	0.28
Mother's occupation (1= working abroad,..8 = working at other home)	+18.31	0.049	0.23
Father's Occupation (1= working abroad,.. 8 = working at other home)	+21.23	0.092	0.32
Home possessions (sum; max 11)	+23.03	0.056	0.24
Home accessories (sum; max 3)	+24.43	0.076	0.29
SES (sum max 7)	+32.42	0.180	0.47
I do jobs at home (1 = not at all,..4 = more than 4 hours)	+7.67	0.017	0.13
I work on a paid job(1= not at all,…. 4 = more than 4 hours)	-10.48	0.044	0.21
Attitude Utility in Nepali (sum max 15)	+19.57	0.050	0.23
Age	+14.44	0.022	0.15
Who helps you …?(1 = Father, …. 6 =Teacher)	+6.45	0.009	0.10
Do you have textbook of Math subject (0 = no, 1 =Yes)	+13.73	0.013	0.11
Homework(0 = not given,..6 = Given everyday, checked everyday)	+19.09	0.010	0.10
Bullying (sum; max 5)	-26.09	0.046	0.22
Positive Activities in school (sum; max 8)	+20.10	0.067	0.27

1) *The order of the variables is the same as handled in the Sections above*
2) *Difference between the lowest and highest group-mean*
3) *On the basis of one-way ANOVA*
4) *4)Cohen's f*

On the basis of univariate ANOVA, school type, followed by the Development region and Ecological zone, is seen to be the most effective single factors in affecting the achievement level of the student; effect sizes are $f = 0.48$, $f = 0.43$, and $f = 0.40$ respectively. Some of these variables in table 4.2.37 show a strong relation to each other and hence they do not add value in explaining why some students are performing much better than others. In what follows, the synthesis of the analysis is done in two ways as presented below; all the variables are presented as a result of Multilevel Modelling, and statistically best factors are collected by using the Regression modelling. For the analysis, grouping factors are changed to so-called Dummy variables when needed; for example, the Ecological zone is transformed into three variables: variables indicative for Mountain, for Hill, and for Tarai.

Modelling the overall achievement by Multilevel Modelling

The datasets collected from schools are always clustered, that is, the students within the school are more similar to each other in comparison with the case that the same amount of students would have been randomly sampled totally from the population. Multilevel modeling is used to acquire the correct test values while taking into account the clustering effect of the school. Table 4.2.39 shows the corrected estimates for the variables mentioned in table 4.2.38 while modeling the phenomenon in a multivariate manner; by using the multivariate ANOVA, the hidden commonalities of the factors are revealed.

Table 4.2.39 Individual variables and their effect in Multilevel analysis

Source[1]	df[1]	df[2]	F	Sig.
Intercept	1	2083.5	572.1	< 0.001
Ecol zone Mountain Dummy (Mountain = 1, other = 0)	1	516.6	15.2	< 0.001
Ecol zone Tarai Dummy (Tarai = 1, other = 0)	1	518.1	1.6	0.209
Dev region Valley Dummy (Valley = 1, other = 0)	1	497.5	21.1	< 0.001
Dev region Center Dummy (Center = 1, other = 0)	1	506.8	0.0	0.913
Dev region Western Dummy (Western = 1, other = 0)	1	506.1	2.6	0.108
Dev region Mid-Western Dummy(Mid-Western=1, other =0)	1	501.8	3.8	0.051
Dev region FarWestern Dummy (Far-Western = 1, other = 0)	1	515.5	0.0	0.889
School location (0 = Rural, 1 = Urban)	1	493.3	1.3	0.263
School type (0 = Community, 1 = Institutional)	1	506.6	65.7	< 0.001
Gender (0 = girls, 1 = boys)	1	11465.5	43.1	< 0.001
Caste Brahman& Cheetri Dummy (Br. & Ch.=1, other= 0)	1	11485.7	137.4	< 0.001
Caste Janjati Dummy (Janjati = 1, other = 0)	1	11506.5	101.5	< 0.001
Caste Madhesi Dummy (Madhesi = 1, other = 0)	1	11607.6	76.9	< 0.001
Caste Dalit Dummy (Dalit = 1, other = 0)	1	11484.0	65.3	< 0.001
Caste Other Dummy (Other = 1, other = 0)	1	11494.2	38.9	< 0.001
Language Dummy (Nepali = 1, other = 0)	1	11852.4	27.1	< 0.001
Homework Dummy 1or2h (1 – 2 hours = 1, other = 0)	1	11671.5	9.2	< 0.001
Paid work Dummy (0 hours = 1, other = 0)	1	11826.9	93.9	< 0.001
Attitude "Utility in Nepali" (Sum, max 15)	15	11604.2	5.1	< 0.001
Age Dummy 11to12y (11 – 12 years = 1, other = 0)	1	11497.3	24.7	< 0.001
Help by Father Dummy (Father = 1, other = 0)	1	11547.2	41.8	< 0.001
Help by Mother Dummy Mother = 1, other = 0)	1	11544.6	24.5	< 0.001
Help by Brother & Sister Dummy (Br & Sis.= 1, other= 0)	1	11569.8	55.9	< 0.001
Help by Tuition Dummy (Tuition = 1, other = 0)	1	11577.5	68.8	< 0.001
Help by No One Dummy (No one = 1, other = 0)	1	11549.7	62.2	< 0.001
Help by Teacher Dummy (teacher = 1, other = 0)	1	11561.0	49.8	<0.001
Do you have a text book in Nepali (Yes = 1, No = 0)	1	11531.6	103.0	<0.001
Homeworks Not Given Dummy (Not given =1, other = 0)	1	11566.8	0.1	0.783
Bullying (Sum, max 5)	5	11587.9	55.4	<0.001
Positive Activities in school (Sum, max 8)	24	11539.1	16.3	<0.001
SES2 (Sum, max 6)	6	11527.0	3.8	<0.001

1) *In the variables Ecological zone and Development region, one of the classes needs to be omitted in the analysis because of singularity reason. Hill zone and Eastern region are omitted; these dummies showed no statistical significance in the Regression analysis.*
2) *Shortened SES; school type is taken away; this enables estimating the parameters for school type.*

When taking into account the clustered structure in the dataset and the joint effect of the factors, many of the factors do not show main effect in achievement in Nepali. Such variables are living in the Tarai zone, Centeral, Western, Mid-Western, or Far-Western region, School location, and Homework not given.[55] These factors could be omitted from the model for explaining the differences in achievement between the students.

Statistically the best factors by using Regression Modelling

Traditional linear regression analysis with stepwise regression is used to explain the total score by the same variables as are above (see table 4.2.38). Table 4.2.40 shows the results.

Table 4.2.40 Statistically the best model of linear regression analysis explaining the average of student achievement (Method: Stepwise)

Model	Coefficients				
	Unstandardized Coefficients		Standardized Coefficients		
	B	Std. Error	Beta	T	Sig.
(Constant)	0.11	2.01		0.05	0.957
School type (0 = Community, 1 = Institutional)	12.30	0.55	0.23	22.54	< 0.001
Positive Activities in school (Sum, max 8)	1.63	0.13	0.10	12.69	< 0.001
Dev region Valley Dummy (Valley =1,other = 0)	9.44	0.56	0.16	16.87	< 0.001
Bullying (Sum, max 5)	-2.23	0.13	-0.13	-17.24	< 0.001
Textbook (Yes = 1, other = 0)	8.69	0.70	0.10	12.39	< 0.001
Language at Home (Nepali =1, other = 0)	3.66	0.40	0.08	9.12	< 0.001
Eco zone Mountain Dummy (Mountain =1, other = 0)	7.53	0.58	0.10	12.89	< 0.001
Paid work Dummy (0 hours = 1, other = 0)	2.95	0.35	0.07	8.55	< 0.001
Dev region Western Dummy (Western=1,other = 0)	2.65	0.50	0.04	5.26	< 0.001
Caste Other Dummy (Other 1, other = 0)	10.51	1.39	0.16	7.54	< 0.001
SES2 (Sum, max 6)	0.94	0.14	0.06	6.50	< 0.001
Dev region Mid-Western Dummy (Mid-eastern =1, other = 0)	-3.77	0.53	-0.06	-7.06	< 0.001
CasteBrahman & Cheetri Dummy(Bra & Che = 1,other = 0)	16.28	1.37	0.35	11.92	< 0.001
Caste Janajati (Janajati =1, other = 0)	15.39	1.36	0.32	11.28	< 0.001
Attitude ”Utility in Nepali” (Sum, max 15)	0.27	0.05	0.05	5.62	< 0.001
Gender Dummy (Girl = 0, Boy = 1)	-1.76	0.33	-0.04	-5.34	< 0.001
Age Dummy (11-12 years = 1, other = 0)	1.58	0.34	0.04	4.65	< 0.001
Homework Dummy (1-2 hours =1, other = 0)	1.67	0.39	0.03	4.33	< 0.001
School location Dummy (0 = Rural,1 = Urban)	2.06	0.53	0.04	3.92	< 0.001
Help by Bro & Sis Dummy (Bro &Sis = 1,othe r = 0)	10.69	1.22	0.24	8.79	< 0.001
Help by Tuition Dummy (Tuition=1,other = 0)	11.20	1.31	0.14	8.53	< 0.001
Help by None Dummy (None = 1, other = 0)	12.73	1.57	0.09	8.11	< 0.001
Help by Teacher Dummy (teacher = 1, othe r = 0)	10.18	1.26	0.17	8.08	< 0.001
Help by Father Dummy (Father = 1, other = 0)	9.04	1.25	0.16	7.26	< 0.001
Help by Mother Dummy (Mother =1,other = 0)	8.27	1.35	0.09	6.11	< 0.001
Caste/ethnicity Dalit (Datit =1, other =0)	12.63	1.42	0.18	8.93	< 0.001
Caste/ethnicity Madhesi (Madhesi =1,other = 0)	12.76	1.47	0.14	8.68	< 0.001

[55] Hill zone and Tarai region showed also non-significant effect

The model in table 4.2.39 can be interpreted as follows: The average mean of the students in Nepali is 0.11% of the maximum score–indicating that the student is in the lowest performing group in all the factors. In the caste category, students from the Brahman and Chhetri caste groups gain +16.3%, which is far better than other caste groups. In the institutional schools (value = 1), the student's score is, on average, +12.3% percent higher (note the sign of coefficient). Additionally, the students from the Valley gain + 9.4 percent more. The students having textbook gain 8.7%. Similarly, the students having experienced positive activities in school gain additional 1.63 in each activities (or eight) resulting a rise in their achievement (8 × 1.63= 13.04 in total), and so on.

On the other hand, the student from the Mid-Western region achieved 3.8 percent lower than from other regions. Similarly, those who face bullying, with each step (of five) the achievement level drops by 2.2 percent where the difference between the lowest and highest group is 17.8 percent.

4.3 Summary of Findings

The main findings of NASA 2012 in Nepali subject conducted for the students of grades three and five are summarized into three groups: basic results, equality indicators, and selected explanatory factors which are presented under the sub-headings as follows:

4.3.1 Basic results

- Proficiency of grade 3 and 5 students in Nepali subject is not distributed normally. There are three distinctive student populations: low and high performing students from community schools and high performing students from institutional schools.
- The students in the institutional schools perform very well and in terms of students achievement the population of community schools form two groups: high-performing and low performing schools. The variations among community schools is remarkable in both grades.
- The learning outcomes are the poorer in the content areas of Writing and Vocabulary and higher in Grammar and Reading in grade 3 whereas in grade 5, the result is poor in the content areas of Reading (56%) and better in Vocabulary (70%). The differences between the content areas are wider in community schools than in institutional schools. This implies that the test was too easy for the students in institutional schools.
- Students' ability to solve complex problems is quite low as only 37% and 47% scores on tasks requiring higher ability were reached in grade 3 and grade 5 respectively. Students are much better in recalling type of questions. Of the total, 3 % and 18% of the students in grades 3 and 5 respectively were not able to solve any of the tasks requiring higher ability. The students in institutional schools are found to have been more able to solve complex problems than their peers in community schools.
- Students are performing well in recognizing the correct answer and in recalling simple facts from the texts, fundamental thinking, the basic interpretation of paragraph, table, and chart, and a few steps of logical thinking. They are much

weaker in producing fluent texts or sentences, or preparing synthesis and abstracts from a text. In many cases, the students tended to do the open ended task (like free writing, problem solving and analysis), but the skills were not high enough for achieving higher scores as expected.

- Compared with the 1995 grade 3 results, in 2012 grade 3, the dataset indicates that there are more high performing students; girls have turned to perform better than boys, the students in the Mountain and Tarai zones score higher and the difference between Hill and Tarai has reduced. In both years, the students from the Eastern region have performed the lowest. Nepali speaking groups have raised their position compared to the other language/ethnic groups. Students from Tharu, Newari, and "Other" communities are found to have lowered down their position remarkably.
- Compared to the 2008 grade 5 results, in 2012 grade 5, there is no change in Nepali language proficiency between boys and girls; the students in the Mountain zone, Western region and the students from urban schools score higher in 2012. The Far-Western region is seen to have progressed remarkably, whereas the Eastern region and Tarai zone still perform lower. The gap between rural and urban students has increased remarkably within five years; the change in the phenomenon is remarkable. The performance in the community schools has risen remarkably but still the gap between community and institutional schools is wide.
- The average reading and writing proficiency of 3rd graders in Nepal is much lower (-1.78 and -1.87 respectively) than the international average considering the PIRLS standard (0.00). The lower achievement level was expected because the linking items were suitable for more matured students of grade 4. Similarly, the average reading proficiency of grade 5 students in Nepal is much lower than the international average of grade 4 in PIRLS standards.
- The typical grade 3 reader of Nepali is at the CEFR level of A2.1. This means that the typical student can understand simple texts containing the most common vocabulary, the main points and some details of a few paragraphs of text. In the institutional schools, a typical grade 3 reader is at level B1.1, that is, (s)he can understand the main points and some detailed messages contained in a few paragraphs in fairly demanding everyday contexts, factual texts, and acquire easily predictable new information about familiar topics from a few paragraphs of clearly structured text. Similarly, the 5th grader student of Nepali in the community school is at the CEFR level of A2.1 in reading.This means that a typical student can read and understand simple everyday texts and factual information, and interpret them in slow pace. In the institutional schools, a typical 5th grader of Nepali is at the CEFR level of B1.1. This means that a typical student can read a few pages of a wide variety of texts on familiar topics following the main points, key words and important details even without preparation.

4.3.2 Diversity factors and equality

- There is a wide difference between the districts when it comes to the equal opportunities of children to reach the pre-set goals in Nepali. The results are very high in the districts where the proportion of institutional schools is high. However, there are also some districts without the representation of institutional schools performing above the national average.
- In grade 3, except for Kaski district (77%), the outperforming districts come from the Central region and specifically in the Valley area: Kathmandu (81%), Bhaktapur (80%) and Lalitpur (78%). The student performance was very low in Saptari (48%) and Khotang (51%) from the Eastern region, in Mahottari (52%) from the Central region, and Bardiya (52%) and Salyan (52%) from the Mid-Western region.
- In Grade 5, except for Kaski (73%) and Solukhumbu (72%) districts, the outperforming districts are from the Central region specifically from the Kathmandu Valley which are: Bhaktapur (81%), Kathmandu (78%), and Lalitpur (73%). In all cases, except for Solukhumbu, the number of private schools exceeds 35% of the schools. The student performance was very low in Saptari (37%) from the Eastern region, in Mahottari (41%) and Sindhuli (48%) from the Central region, and Salyan (46%) and Jumla (49%) from the Mid-Western region. In all cases, the number of private schools is low.
- There is a moderate difference in the student performances among four Ecological zones within both community and institutional schools. Students in the Kathmandu Valley outperform the other students. The achievement of community school is the lowest in Tarai zone. Similar trends can be found in both grades.
- There is inequality of children's opportunities to reach an adequate level of achievement in Nepali among the Development regions. Especially,the wide difference in the community schools between the Valley and the rest of the country (21 percent as the highest) is a distinct sign of inequality of opportunities in learning Nepali.There are also wide differences between the regions within the institutional schools; the difference in student performance within the private schools between the Valley and Eastern region is the highest:12 percent in grade 3 and 9 percent in grade 5.
- On average, the students in institutional schools outperformed the students in community schools. The difference is highest in Writing (28 percent) in grade 3 and Reading (25 percent) in grade 5. This variance can be explained partly by much higher socio-economic input into students' life and the strict selection of students in institutional schools.
- The students in the urban community schools have gained 7 percent more in grade 3 and 9 percent more in grade 5 than the students in the rural areas. Excluding the Valley schools, the difference is only 2 percent in grade 3 and 3 percent in grade 5. However, from the educational equality point of view, the difference is not a good

sign though the real difference is not wide within the community schools. In the institutional schools, there is not wide difference between the rural and urban areas.

- There is an inequality in achievement within the language and ethnic groups. From ethnicity point of view, the students from Magar (81%) comminity perform very high in Nepali while those from Newar (5%), Tharu (56%), and Gurung (47%) performed lower than the average in grade 3. Similarly, in grade 5 Magar language speakers (76%) are far ahead of other students (41 – 65%) in Nepali proficiency. The Tarai Tharu students are far below the others (40%). Quite low results are found among the Rai (51%), Sherpa (52%) and Newar (54%) populations. The differences between the lowest and highest performing language groups are remarkable.
- There are statistically significant, though not remarkable, differences between the castes in Nepali grade 3. Students form the "Others" ethnicities/castes are performing the lowest (55%) followed by Dalit, Madhesi, and Janjati students (all 58%); the highest scores are achieved by the Brahmin (67%) and Chhetri (61%) groups. In the case of grade 5, students from the Madhesi (49%) and Dalit (53%) as well as "Others" ethnicities/castes (46%) are performing significantly lower than the students from the Brahman, Chhetri and Janajati groups. In the community schools, Dalit students performed somehow lower than average in the entire Mid-Western (< 54%) and Far-Western regions (< 57%), Western Tarai (53%), Eastern Hill (54%) and Tarai (56%). They perform better than the national mean in the Eastern (64%) and Western Hill (60%), as well as in the entire Central region (> 60%) in grade 3. Madhesi students perform lower especially in Tarai zone, in Western and Far-Western regions in grade 5.
- The differences between boys and girls in Nepali proficiency are small. From equality point of view, this is a positive sign. The tendency shows that girls tend to slightly out-perform boys in Nepali in both grades. Moreover, the difference is seen wider in the institutional schools (4 percent favoring the girls) in comparison with the community schools (2 percent) in grade 5.

4.3.3 Selected explanatory factors

- Parents' educational level strongly predicts the children's future achievement level in Nepali. Especially, harmful for the achievement level is the situation where the father or mother or both are illiterate. Similar trends are found in grades 3 and 5.
- Either economic or intellectual capacity or both at home helps children to increase their Nepali proficiency. If the father or mother or both are in agriculture–related occupation, or they are in daily wages related occupation, the students' achievement in Nepali is significantly lower than those coming from the other occupational groups like business and job.
- When children have very few home possessions – zero to three out of the 11 – the achievement level is remarkably lower than the national average (< 63%). With nine to ten possessions, the average score is very high (> 72%) compared with the national average. The same is true for home accessories: When none or only one

accessory indicator out of three is met, the results are lower than average (55–64%) and when two or more are met, the results are remarkably higher (70–76%). Similar result can be seen in grade 5 as well. When children have very few home possessions – zero to three out of the 11 – the achievement level is remarkably lower than the national average (< 56%). With nine to ten possessions, the average score is very high (up to 69%) compared with the national average.

- Socio-economic status plays a strong role in the achievement in Nepali subject. The difference between the lowest and highest SES groups is remarkable (31% in grade 3 and 30% in grade 5). This means that if the SES of the lowest performing students were raised into a decent level, the results in these groups would have been raised remarkably. Especially, challenging is the situation in the families where the father or both parents are illiterate or just literate. There is no huge difference between illiterate and just literate families but the gap is huge between illiterate/literate and grade 10 or above. The situation is not very easily improved to be equal between the community and institutional schools.
- Either working for a paid job or for four hours per day in an unpaid household work outside school reduces the students' school achievement. However, a decent amount of household work up to two hours per day has been found not impeding the students' learning in Nepali. Like other aspects this, trend is also the same in grades 3 and 5.
- Positive attitude towards the subject correlates with positive achievement in Nepali. The better achievement is more probable a consequence of more positive attitude rather than other way round in both grades.
- The highest performance is found among those students who are studying with their normal age group, that is, at the age of 8 to 10 years in grade 3 and 10 to 12 years in grade 5. Otherwise, the achievement lowers down with the increase in the age years.
- The academic support provided by the mother, brother/sister also contributes significantly in achieving better scores. In the whole sample, the highest achieving group is the one who receives private tuition and teacher support. However, the difference between the highest and lowest performing groups is not notable. It is possible that the group receiving private tuition also spends more time on their homework, which is possibly a strong factor in explaining the higher score.
- Data shows that as many as 4.4% of the students in grade 3 and 3.7% in grade 5 lack textbook in Nepali up to the end of academic year. The achievement level of these students is significantly lower than of those who have access to textbook.
- If the teacher assigns the homework and checks it systematically, the achievement is higher compared to the students without having it along with its checking. It is found that the achievement scores have been raised (up to 16 percent in grade 3 and 15 percent in grade 5) by regularly assigning and checking the homework.
- An alarmingly large number of the students (55% in grade 3 and 54% in grade 5) have been encountered bullying in school within a month and 8.8% of them are experiencing a severe kind of bullying. This means that more than 75,000 grade 3

and 44,000 in grade 5 students in Nepal have encountered physical, psychological, and social bullying in every month. The phenomenon is found to have affecting the learning outcomes in almost all the groups of the students who experiences bullying, so all possible efforts have to be put to root out the phenomenon from schools.

- When the students feel that the actions of the teachers and schools are ultimately good, the Nepali results are better than average (61% in community school and 81% in institutional schools in the case of grade 3, and 58% in community schools and 80% in institutional schools in the case of grade 5). At the other extreme of feeling that such actions are ultimately negative, the results are far below the average (48% in the community schools and 71% in the institutional schools).

  There is a connection between the future aspiration of student and their achievement. As the students aspire to have a professional career other than farming or business, their achievement is higher than the average. The number of students who aspire to be a teacher, a government officer,an engineer, or a doctor is remarkably high in grade 5.

Chapter 5: Assessment Results in English

English as a school subject is the second language (L2) for the students though it is used as a medium of language for instruction in the most institutional (private) schools. English is also one of the Lingua Franca and is considered an advantage to have command in it in order to survive in global world as it is widely used for business purposes and for international job market. English is also the most common language in use among the taxi and R*ikshaw* drivers, shop keepers, restaurant and hotel workers in the tourist areas to survive in the business and to make their living. Hence the value of having command in English in the modern world is high.

English Language proficiency at grade 5 has been assessed more or less systematically and frequently in the National Assessments of Student Achievement (NASA) in Nepal. The results of the previous national assessments (see BPEP, 1994; CERID, 1993; 1998; 1999; CERSOD, 2001; Fulbright, 2008) are not fully comparable with each other because of the missing linking procedure between the tests. So the proficiency levels are not comparable in the absolute sense (as, for example, percentages of correct answers are not). However, the proportional differences between the groups and content areas are compared in what follows.

Achievement in English of the grade 5 students was assessed through student achievement tests and the related background questionnaires among 13,794 students from 563 schools of 28 randomly selected districts all over the country including three purposefully selected districts in the Kathmandu Valley. The schools represented all ecological zones and development regions, rural and urban areas as well as community and institutional schools. Basic results of assessment as well as disaggregated results based on various strata and diversity are included in the analysis. Besides, the extents to which a number of related factors have influenced in student achievement are also described.

The reports of assessment of student achievement in grade five English begins with analyzing the basic results including overall distribution of scores, results in the different content areas and goes to the analyse the effects of different diversity factors from equality point of view. It then describes the influencing factors explaining the differences in the achievement of English subject.

5.1 Assessment Results in English for Grade 5

This section analyses the assessment results of English subject at grade 5. It starts with the analysis of basic results including overall distribution of scores. Then it presents the results in the different content areas of English language in general and goes to the analysis of the effects of different diversity factors from equality point of view. It also analyses the influences of factors explaining the differences in the achievement in English languge.

5.1.1 Basic Results of Assessment in English for Grade 5

As the basic results of assessment in English, this sub-section analyses the overall distribution of scores, results in the various content areas, various levels of cognitive domains, result variations in item types, and comparisons of results with previous assessments and international assessment results.

Distribution of overall results

The English sample is big enough to form the normal distribution. However, figure 5.1 shows that the total score is not normally distributed. The final score of English is seen to be evenly distributed.

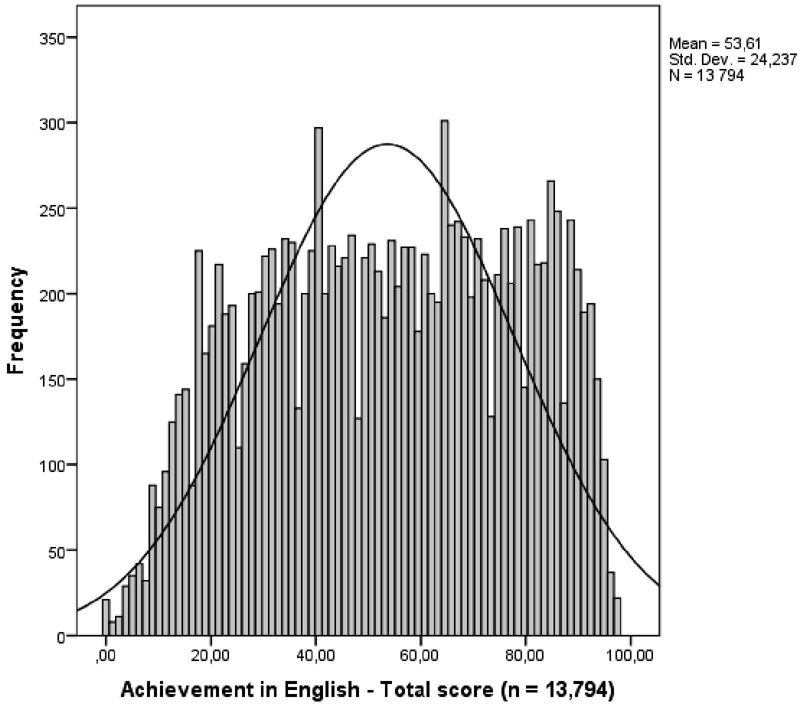

Figure 5.1 Final scores of English

There are two different, clearly distinctive, normal populations in the dataset: students from the community schools and from the institutional schools (fig. 5.2).

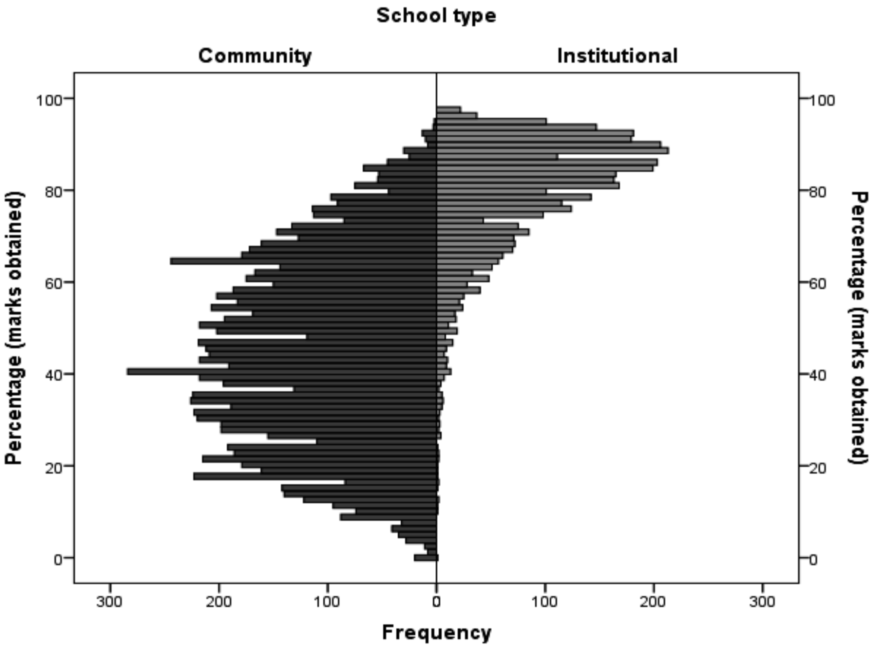

Figure 5.2 Distribution of the students' scores by school type

In figure 5.2, the left hand side distribution shows that of the community school students and the distribution on the right hand side that of the institutional school students. The main system is shifted towards the lower performing because the main population comes from the community schools. It is obvious that the main population in the institutional schools performs very well so explains a high result as the medium of language for instruction and textbooks in almost all subjects are in English. However, there are quite a number of students in the community schools obtaining equally high marks as gained by the students in institutional schools. Based on figure 5.2, it is evident that the students in the community schools are varying from the low-performers to the high performers whereas most of the students from the institutional schools are performing high or medium.

The schools – not only the students –also are clearly divided into two "populations": the high and the low performing schools. The distribution of the school mean score shows that both populations are slightly skewed: the community schools are skewed to include more low performing schools and the institutional schools are skewed to include more high-performing schools. On the basis of school mean score of the student performance, there are two categories of schools. The difference between the populations is remarkable. The difference between the school populations can be explained by the division between the private and community schools.

By analysing it further with the scatter plot and combining the socio-economic status (SES) with the average achievement in the school, figure 5.3 shows that two types of schools fall into two groups, where most of the institutional schools (triangle) are performing better and the average SES is very high. At the same time, the community schools (circle) vary from very high performing to very low-performing, showing low SES on an average.

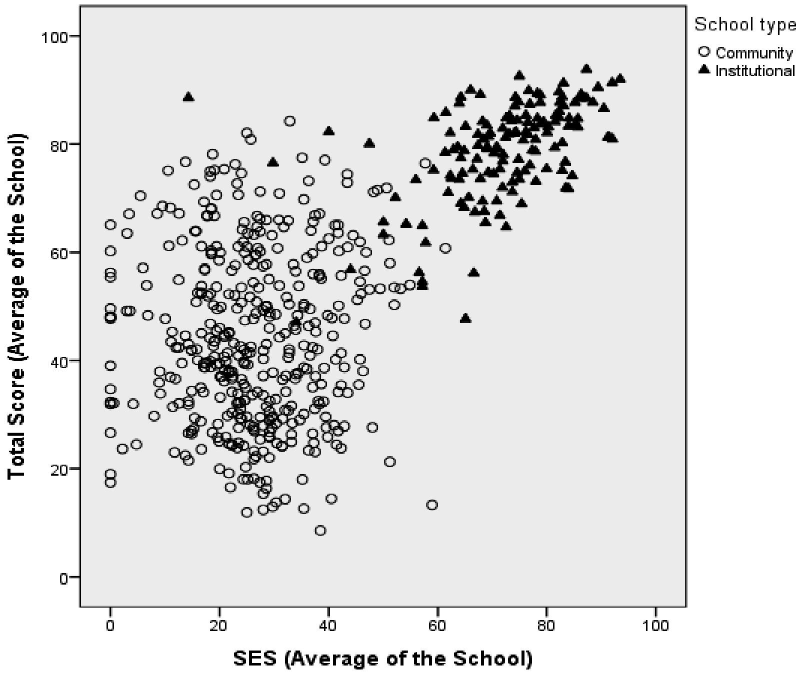

Figure 5.3 Relation between achievement and socio-economic status by types of schools

The dataset is evident for the fact that the grade 5 population in English is not normally distributed. There are two distinctive student populations: students from community schools and students from institutional schools. The variations between the community schools is remarkable.

Achievement in various content areas

The whole English test was a combination of four content areas: Reading, Writing, Grammar, and Vocabulary. The maximum marks of Reading and Writing were proportionally equal to the weightage given by the curriculum. Grammar and Vocabulary were taught as "functional skills", that is, in relation to Reading and Writing. To compare the achievement in all the topics, these sub-scores are converted into a percentage of the content area. Table 5.1 shows the students' achievement in English as a whole and the achievement level by each of the four content areas.

Table 5.1 Comparison of average score in various content areas

Content area	Mean	SD	Min	Max
Reading	50.2	25.7	0	100
Writing	49.0	25.6	0	96
Grammar	56.5	27.7	0	100
Vocabulary	57.8	25.9	0	100
English as Total	**53.6**	**24.2**	**0**	**97**

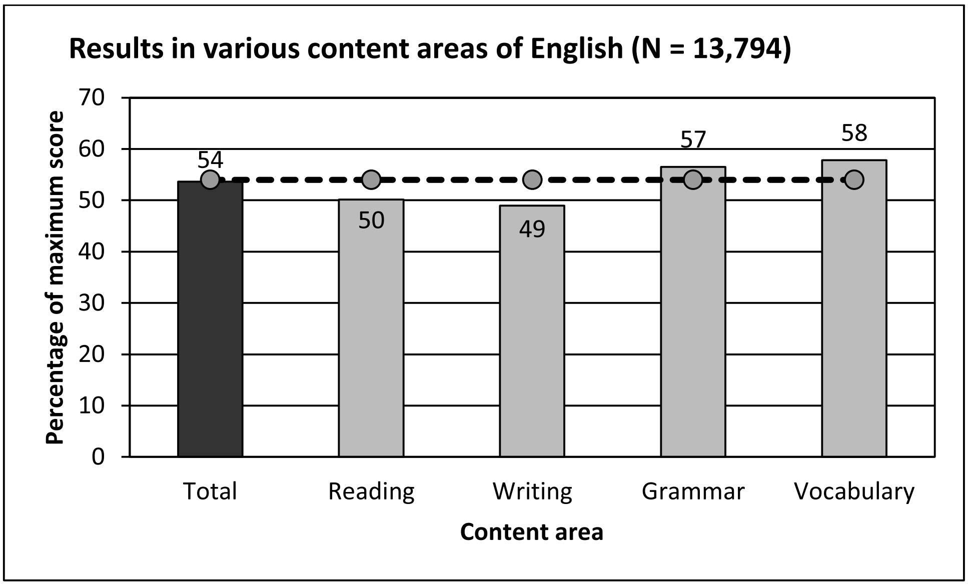

Figure 5.4 Comparison of results in various content areas of English

The percentage of achievement score shows that the national average of English is 54%. Of the different content areas, students are somehow weaker than the average in Reading (50%) and Writing (49%). They perform better than average in Grammar (57%) and Vocabulary (58%). Because of the difference in the average achievement between the community and institutional schools, it will be interesting to know whether there is proportional difference in the content areas between the students. Table 5.2 illustrates the differences.

Table 5.2 Achievement in various content areas by school type

Content area	Community schools (n =10,107)			Institutional schools (n = 3,674)		
	Mean	SD	CV	Mean	SD	CV
Reading	40.6	21.7	53.3	76.5	15.6	20.4
Writing	39.8	22.0	55.3	74.4	15.9	21.4
Grammar	48.8	26.0	53.4	78.0	19.9	25.6
Vocabulary	50.1	24.6	49.0	79.3	14.8	18.6
Total	**44.4**	**20.4**	**45.9**	**79.0**	**13.4**	**17.0**

It is evident that the differences between the content areas are wider in community schools compared to institutional schools. While the difference between Vocabulary (50%) and Writing (40%) is 10 percent in community schools whereas it is only 5 percent in institutional schools. This can be explained partly by ceiling effect in institutional schools because of the easy test for the students in the institutional schools. Hence, the best students were not able to show how high they could have been able to achieve in knowledge type of items.

The dataset indicates that the learning outcomes are the poorest in the content areas of Reading and Writing and the highest in Grammar and Vocabulary. The differences between the content areas are wider in community schools than in institutional schools. This can be caused by the ceiling effect; the test was too easy for the students in institutional schools.

Achievement in various levels of cognitive domain

The English test as whole was constructed on the basis of Bloom's taxonomy of hierarchical level of cognitive domain (Bloom *et al.,* 1956; Metfesser, Michael, & Kirsner, 1969) that is, *knowledge*, *comprehension*, *application*, and *higher ability* (reasoning/problem solving). The achievement of the students on each hierarchical level in total sample is presented in figure 5.5.

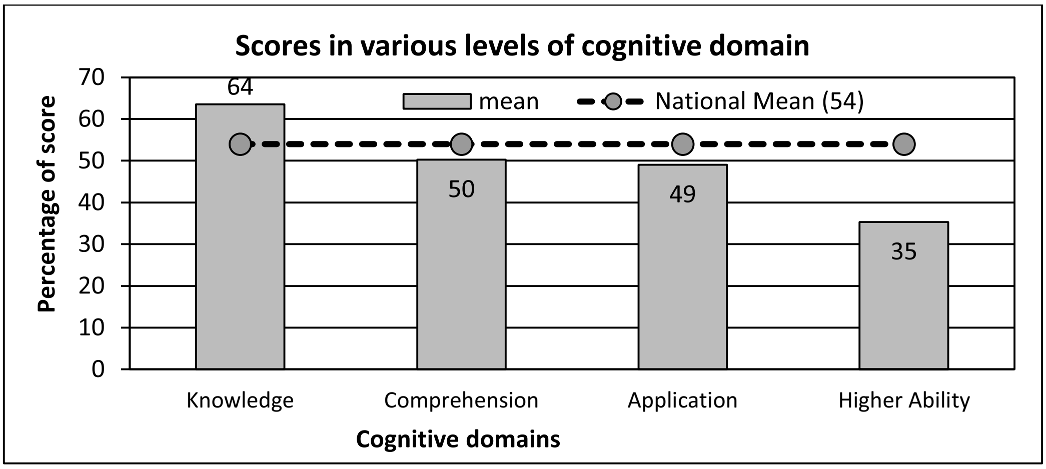

Figure 5.5 Achievement in various levels of cognitive domain

Remarkably higher numbers of students were able to solve only 15 percent or less practical problems, that is, the application type of items. About one third of the students (32%) could solve just less than 10 percent of the tasks requiring the higher cognitive abilities, and 18% students did not solve any of these tasks.

Because of the difference in the average achievement level between the community and institutional schools, it will be interesting to know whether there is proportional difference in the hierarchical level between the students. Table 5.3 illustrates the differences of achievement of community and institutional schools in various levels of cognitive domain.

Table 5.3 Achievement in various levels of cognitive domain by school type

Levels of cognitive domain	Community schools (N = 10,107)			Institutional schools (N= 3,674)		
	Mean	SD	CV	Mean	SD	CV
Knowledge	55.4	32.2	58.0	86.0	19.5	22.7
Comprehension	41.5	19.2	46.1	74.3	15.1	20.3
Application	40.1	21.5	53.6	73.7	14.4	19.5
Higher Ability	23.3	24.5	105.1	68.4	23.8	34.8
Total	**44.4**	**20.4**	**45.9**	**79.0**	**13.4**	**17.0**

The main trend is that in both the cases the students are much better in the recall type of questions than in the tasks requiring high level skills. However, there is another tendency worth noting. In institutional schools, difference between the scores of knowledge type and higher ability type questions is much smaller (18 percent) than in community schools (32 percent). Statistically, the tendency is seen in the effect size, though, the differences are remarkable in any case, the difference is notably smaller in the area of knowledge (d = 1.04) than in higher ability (d = 1.85). This means that, for one reason or another, the students in institutional schools are seen to be more able to solve complex problems in relation to simple tasks than their peers in community schools.

From the table 5.3 the exceptionally high value of Coefficient of Variation in community schools is seen in higher ability. The moderate standard deviation (24.5) with very low mean (23.3) means that, in community schools, there are also reasonably high performing students.[56]

The dataset shows that the students' ability to solve complex problems is quite low as only 35% of the maximum scores on the tasks requiring higher ability were obtained. Students are much better in the recalling type of questions (64%). Remarkable number of students (18%) was not able to solve any of the tasks requiring higher ability. The students in institutional schools are found to be more able to solve complex problems than their peers in community schools.

Types of items and achievement

There were basically two types of questions in the test: objective and subjective items. Objective items covered a wide range of content areas and were very specific to judge because there was only one correct answer for each question, or one explicit piece of information was needed to get a correct answer. There were some subjective items on each test version requiring a longer procedure to get the full marks. Both the objective and subjective types of items were presented according to various cognitive levels (knowledge, comprehension, application, and higher ability) and a wide range of difficulty levels, though the subjectively scored items tend to be more demanding because of the demand of higher cognitive skills. Table 5.4 comprises the basics statistics of the item type-wise achievement levels.

Table 5.4 Item type-wise mean scores

Type of items	Mean	SD	Minimum	Maximum
Objective	59.3	23.1	0	100
Subjective	43.6	30.5	0	100

It is obvious that the subjectively scored tasks – usually those with more demanding requirements for the correct answer – are solved much lower (44%) than the objective items (59%). Most of the objective items were knowledge, comprehension and application type whereas subjective items were application and higher ability type.

The dataset reveals that the students are performing well in recognizing the correct answer and in recalling simple facts from the texts, fundamental thinking, the basic interpretation of paragraph, table and chart, and a few steps of logical thinking. They are much weaker in producing fluent texts or letters, or preparing synthesis and abstracts from a text. In many cases, the students tended to attempt the open-ended task (like free writing, problem solving and analysis), but the skills were not high enough for achieving the highest marks.

[56] In the community schools in the Kathmandu Valley, the mean of Higher Ability items is 46.9 (CV = 59.3). Hence, the CV is not exceptionally high compared with the CVs of the other cognitive levels in the total sample.

Comparison of NASA 2012 results to previous datasets

The national assessment carried out in various years aim at assessing the changes in the achievement level and the progress over a period of the years. The datasets of previous English assessment are, however, somehow fragmented and achieved by using various strategies for sampling which makes the comparison difficult.[57] The previous datasets also carry two other challenges hindering the comparison with the present dataset. First, the National Assessment of grade 5 students carried out by the Basic and Primary Education Project showed that the national average of the students was 39%(BPEP,1994:9) . Later, the study conducted in 1998 by the Research Centre for Education Innovation and Development (CERID) reporteds it to be around 33%(1998:13). In NASA 2012, the national average of English in grade 5 is found to be 54%. These figures are coming from Classical Test Theory which may not be comparable with each other because of lack of a proper linking procedure. The differences between the scores can easily be explained by the different difficulty levels of the tests. Second, the previous datasets of grade 5 are not available and, hence, any IRT modelling-based procedures for comparison could not be made.

Though the comparison cannot be made in absolute sense, proportional comparisons can be made, with caution, on the basis of the previous results. The proportional differences are collected in tables 5.5 to 5.8.

Table 5.5 Comparison achievement by gender in 1998 and 2012

Indicators	1998 (CERID, 1998:13)		NASA 2012	
	Boys	Girls	Boys	Girls
Mean	34.3	33.0	54.8	53.0
SD	13.9	15.2	24.2	24.4
CV	40.6	46.0	44.1	46.0
N	727	635	6,604	6,519

Compared with the 1998 dataset, the difference between the boys and girls has not changed radically; boys still outperform girls mildly (less than 2 percent point). The Coefficient of Variation (CV) shows that the distributions of girls have stayed very stable over the 15 years (CV = 46) but the variance of the boys has raised mildly during years (CV has risen from 41 to 44). In both years, the difference between boys and girls is not remarkable (Cohen's $d < 0.10$).

Table 5.6 Comparison of score of 1998 and 2012 by Ecological zones

Indicators	1998 (CERID, 1998:14)			NASA 2012		
	Mountain	Hill	Tarai	Mountain	Hill	Tarai
Mean	37.4	37.2	35.5	50.8	48.9	46.3
SD	17.5	15.8	14.8	21.0	23.5	21.3
CV	46.8	42.5	41.7	41.3	48.1	46.0
N	105	718	539	1,371	6,291	3,715

[57] For example, CERID 1998 sample included five districts covering all developmental regions (one district from each developmental region) and ecological zones. This deviance in samplings makes the comparison somehow difficult.

Compared to the 1998 dataset, the difference between the Ecological zones is seen to have widened mildly. The difference between the students in the Mountain and Hill zones has increased from 0.2 to 1.9 percent, and between Mountain and Tarai it has been further wider from 1.9 to 4.5 percent. The Coefficient of Variation (CV) shows that the distribution of students from the Mountain area has been narrowed down by 4 percent (CV has changed from 47 to 41) while it has widened in the Hill (from 42 to 48) and Tarai (from 42 to 46) zones. Both years, the differences between the zones are not remarkable (the pairwise values for Cohen's $d < 0.21$).

Table 5.7 Comparison of scores of 1998 and 2012 by location of schools

Indicators	1998 (CERID, 1998)		NASA 2012	
	Rural	Urban	Rural	Urban
Mean	39.5	32.2	45.1	67.2
SD	15.2	15.1	21.2	20.3
CV	38.5	47.0	47.0	30.2
N	821	541	8,989	1,558

Compared to the 1998 dataset by the location of schools, it is evident that the scenario has changed radically during the 15 years. In the 1998 dataset, the difference between the rural and urban schools was moderate (d = 0.48) , favouring the rural schools;whereas in the 2012 dataset the difference is remarkable (d = 1.05), favouring the urban schools more even though the Valley schools are not included in the comparison.[58] The change in the phenomenon is remarkable. It most probably tells the rise of institutional schools in the urban areas. Table 5.8 presents the comparison of previous datasets with the results of NASA 2012 in English.

Table 5.8 Summary of comparison of achievement of 2012 with the previous datasets

Selected background variables					
	Gender	**Ecological zone**	**Development region**	**School location**	**School type**
Main finding	No change in difference; girls still under-perform boys mildly	Differences have increased moderately. Students in the Mountain zone scored higher while in Tarai they scored lower	Students in the Far-Western region have scored higher while in the Eastern region they have scored lower	Remarkable rise in performance in the urban schools	No remarkable change in difference between community- and institutional schools

The dataset suggests that compared to the 1998 and 1999 results; there is no change in difference between boys and girls. The students in the Mountain zone, Mid and Far-Western regions scored higher and the students in urban schools score remarkably higher in 2012. Dataset from 20 years back indicates that the performance in the community schools in the Kathmandu Valley has risen remarkably. On the other hand, the students in the Eastern region and Tarai are performing lower in 2012 than 15 years ago.

[58]If the Valley schools are included in the analysis, Cohen's d were d = 1.20.

Comparison NASA 2012 results to international mean

The NASA 2012 test was made comparable with the international PIRLS reading assessment. Four of the released PIRLS items were used as linking items from the international item bank. 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 2012 is zero, it corresponds to the average level of the international students of grade 4 in their own native language. As the text and the related items were targeted to the native speakers, it was expected to be somehow difficult for the speakers of English as a second language.

Figure 5.6 shows the comparison of the students' achievement with the international standard. In the figure, the x-axis shows the content areas of English and y-axis shows the ability shown by the students. The middle horizontal line indicates the international average. When the ability is below the average, the bars go down, whereas the bars will go upwards when the ability is above the international average.

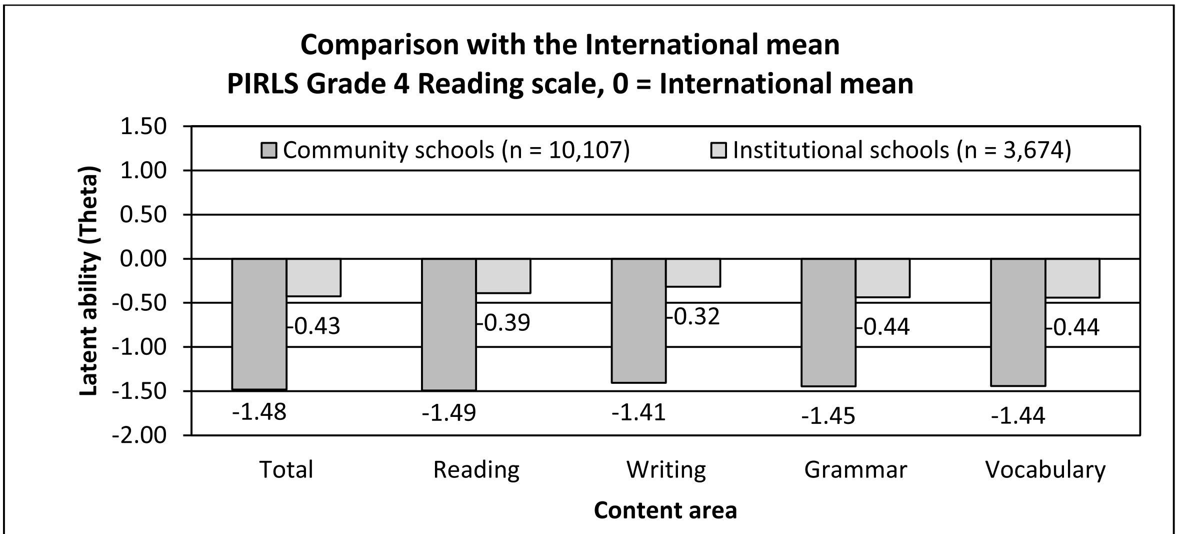

Figure 5.6 Student achievements of English in the international PIRLS reading scale

Figure 5.6 shows that the average ability shown by the students of Nepal in English in Reading is generally below the international average. This indicates that the students in Nepal score remarkably lower than their international peers. The achievement level of an average grade 5 student in the community schools ($\theta = -1.49$) is very low compared with an average international student of grade 4. The achievement level of an average student in the private schools ($\theta = -0.39$) is much higher than his/her peer in the community school. It is good to remember two things. First, all the linking items came from the content area of Reading and hence there is actually no real equating in the other areas. However, they are modelled on the basis of proficiency in the reading test. Second, the difficulty level of the text (complex English text suitable for native English speakers of 4th graders) was not best suiting to the 5th graders' English proficiency of Nepali native speakers. Hence, the lower proficiency is expected.

The dataset is evident that from the international comparison point of view, the average reading proficiency of Nepalese students is much lower than the international average in PIRLS standard.

Comparison with the objective standards – CEFR levels

Another type of international comparison was made on the basis of the standards of Common European Framework in Reference for Languages (CEFR). CEFR classification with the standard setting procedure (3TTW) (Metsämuuronen, 2013; ERO, 2013) was applied to assess the criterion based proficiency in the English language. The main results of English Reading and Writing proficiency levels in Nepal are depicted in figures 5.7 and 5.8 and in tables 5.9 and 5.10.

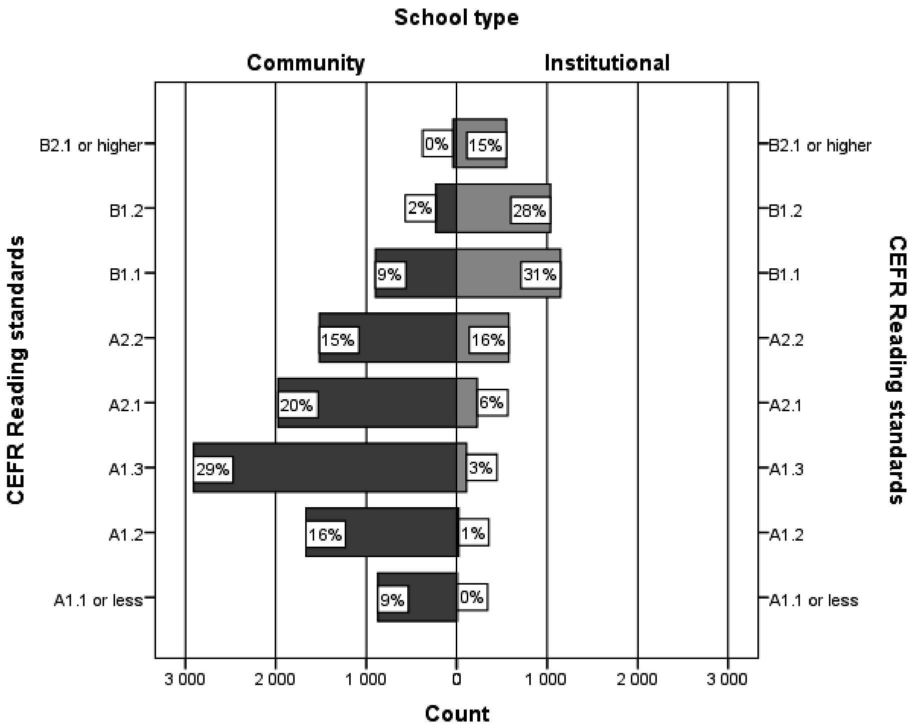

Figure 5.7 Comparing student achievements of English with international CEFR Reading standard

In the community schools, the most typical 5th grader of English is at the level of A1.3. This means that the typical student "*can read familiar and some unfamiliar words, understand very short messages dealing with everyday life and routine events or giving simple instructions and* can *locate specific information required in a short text (postcards, weather forecasts)*." (FNBE, 2004.) In the institutional schools, the most typical 5th grader student of English is at the level of B1.1. This means that the typical student "*can read a few pages of a wide variety of texts about familiar topics (tables, calendars, course programmes, cookery books), follows the main points, key words and important details even without preparation; and* can *follow the main points, key words and important details of a few pages of text dealing with a familiar topic*." (FNBE, 2004.)

In figure 5.7, it is seen that 54% of the 5th graders in the community schools are at the level A1.3 or lower which means that they can read only very short notices and postcard type of texts and only pick up some facts out of the text. This kind of elementary reading skill was tested, for example, in the reading task called "Pokhara letter" of level A1.3, a four-sentence postcard was given. The first sentence in the postcard was: "*I have been in Pokhara for 4 days now*". The first question of multiple choice type was: "*How many days had Sita been in Pokhara when she wrote the letter?*" The alternatives were: three, four, six and thirteen days. As many as 23% students did answer nothing (missing value), 12% of them selected three days, and 3% selected thirteen days. The distracting numbers were mentioned in the text, but they had nothing to do with the question.

On the basis of the dataset, 59% of the whole student population has reached at least the level A2.1 in English reading and hence they can understand (only) simple texts containing the most common vocabulary (table 5.9).

Table 5.9 Percentages of 5th graders who reached the specific CEFR levels in Reading

CEFR level	Brief description of ability	**% reaching the level**	**% at each level**
B1.2 or higher	Can read a **few pages of** text independently (newspaper articles, short stories, popular fiction and non-fiction, reports and detailed instructions) about **his/her own field or general topics**.	4.3	4.3
B1. 2	Can read **a few paragraphs of text about many *different* topics** (newspaper articles, brochures, user instructions, simple literature).	13.5	9.2
B1. 1	Can read **a few pages of a wide variety of texts about *familiar* topics** (tables, calendars, course programmes, cookery books).	28.3	14.8
A2. 2	Can understand the **main points and some details of** messages consisting of a few paragraphs in **fairly demanding everyday contexts** (advertisements, letters, menus, timetables) and **factual texts** (user instructions, brief news items).	43.5	15.2
A2. 1	Can understand **simple texts containing the most common vocabulary**.	59.4	15.9
A1. 3	Can understand **very short messages dealing with everyday life** and routine events or giving simple instructions.	81.3	21.9
A1. 2	Can understand names, signs and other very short and **simple texts related to immediate needs.**	93.6	12.3
<A1.2	Can read some **familiar and unfamiliar words**, understand very **short messages** dealing with everyday life and **routine events** or giving **simple instructions.**		6.4

Regarding the English writing proficiency in the community schools, the typical 5th grader students of English vary widely ranging from A1.2 to A2.2 though the mode is at A1.3 (figure 5.8). This means that compared with the reading proficiencies, there are quite many good writers in the 5th graders in community schools.

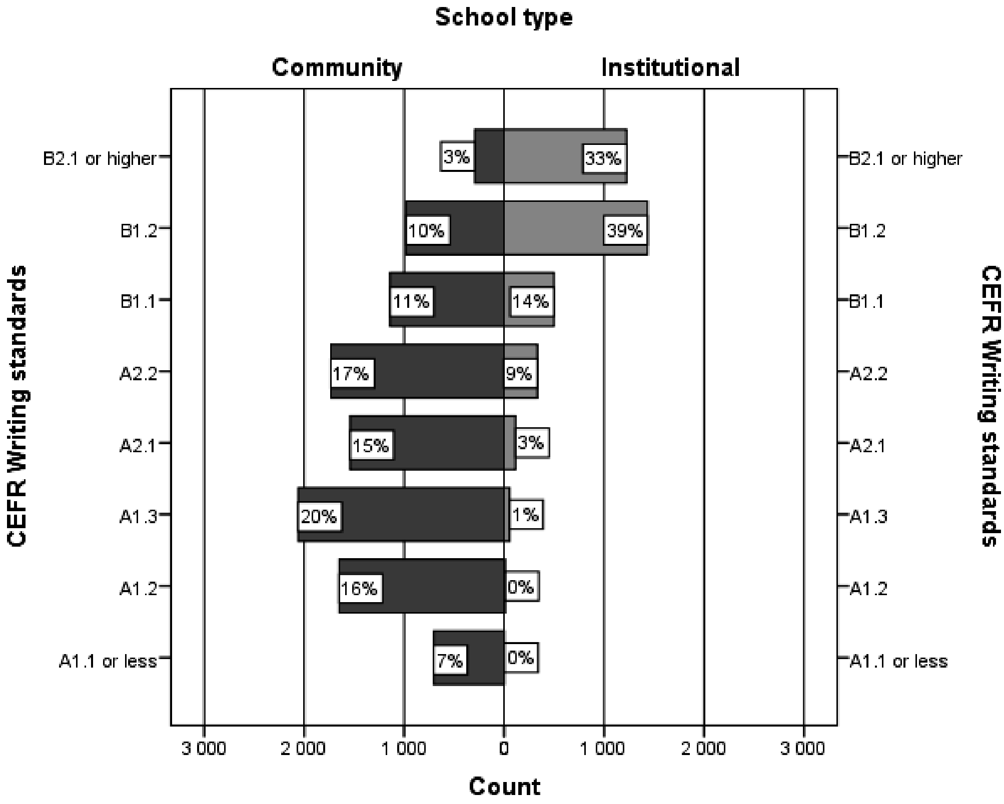

Figure 5.8 Student achievement in the international CEFR Writing standard

The typical student of A1.3 "*can manage to write in the most familiar, easily predictable situations related to everyday needs and experiences,* and can *write simple messages (simple postcards, personal details and simple dictation).*"(FNBE 2004.) The typical student of A2.2 "*can manage in routine everyday situations in writing and can write a very short, simple description of events, past actions and personal experiences or everyday things in his/her living environment (brief letters, notes, applications, telephone messages).*" (FNBE 2004.) In institutional schools, the most typical 5th grader writer of English is at the level of B1.2 (39% of students) though many students are at a level even higher than this (33%). This means that the typical student can, at least, "*write personal and even more public messages, describing news and expressing his/her thoughts about familiar abstract and cultural topics, such as music or films and* can also *write a few paragraphs of structured text (lecture notes, brief summaries and accounts based on a clear discussion or presentation).*" (FNBE 2004.)

On the basis of the dataset, it is seen that 55% of the students in the whole student population have reached at least the level A2.1 in English Writing and hence they can write a **very short, simple description of events, past actions and personal experiences** or **everyday things in his/her living environment** (brief letters, notes, applications, telephone messages) (table 5.10).

Table 5.10 Percentages of 5th graders who reached the specific CEFR levels in Writing

CEFR	Brief description of ability	% reaching the level	% at each level
>B2.1	Can **express information and views effectively in writing** and comment on those of others. Can combine or summarise information from different sources in his/her own texts.	11.0	11.0
B1.2	Can **write a few paragraphs** of structured text (lecture notes, brief summaries and accounts based on a clear discussion or presentation).	28.5	17.5
B1.1	Can write a **clearly formulated cohesive text** by connecting isolated phrases to create longer sequences (letters, descriptions, stories, telephone messages).	40.4	11.9
A2.2	Can write a **very short, simple description of events, past actions and personal experiences** or everyday things in his/her living environment (brief letters, notes, applications, telephone messages).	55.4	15.0
A2.1	Can write **brief, simple messages** (personal letters, notes), which are **related to everyday needs**.	67.4	12.0
A1.3	Can manage to write in the **most familiar, easily predictable situations** related to **everyday needs and experiences**. Can write **simple messages** (simple postcards, personal details, simple dictation).	82.7	15.3
A1.2	Can **communicate immediate needs in brief sentences**. Can write a **few sentences** and phrases about him/herself and his/her immediate circle (such as answers to questions or notes).	94.8	12.1
<A1.2			5.2

Two points regarding the English language proficiency of students in grade 5 are worth of highlighting. First, usually in the language learning, the receptive skills (Reading and Listening) stay at the lower level than the productive skills (Writing and Speaking). Also, the desired levels of the receptive skills are set lower than for productive skills (see Table 3.3.9c; FNBE, 2004, 140; see also ERO, 2013, table 4.6.3). In English grade 5 dataset, however, the Writing skills are seen to be at the higher level than those of Reading. One explanation for this is that, in general, the writing skills are seen to be preferred over the Reading skills in the current curriculum. Also in Nepali grade 5, the overall Reading proficiency is lower than the Writing proficiency (see chapter 4).

Second, there is no objective criterion on what they should be for the English language proficiency level at the end of grade 3 or grade 5. One aid for evaluating the proficiency is to use the Finnish core curriculum (FNEB, 2004) and the descriptions of good performance as the measurement stick. In the Finnish system the criterion is given for the end of grade six and hence it is not fully comparable in Nepalese context. Anyway, some clues of the required language proficiency levels can be obtained from table 5.11.

Table 5.11 Description of good performance in English at the end the sixth grade in the Finnish system

Language and level	Reading	Writing
English as the first foreign language (FNBE, 2004:140)	A2.1 First stage of basic proficiency	**A1**.3 Functional elementary proficiency
Finnish as Native-level(bilingual) (FNBE, 2004:135)	B1.2 Fluent basic language proficiency	B1.1 Functional basic language proficiency

In the Finnish system, the level A2.1 is taken as a good level for reading comprehension and A1.3 for Writing. Compared with these references, it is seen that the typical English reader in the community schools (A1.3) in Nepal is at the lower level than the "good" reader in the Finnish system and the typical English writer is at the "good" level (A1.3). It is notable though that there are quite many students with higher writing skills than A1.3. In the institutional schools, the typical English reader and writer is seen to have reached at the same level as is required for the bilingual native speaker at grade 6. This is a very high level. More comparative studies and standard settings are required to further confirm the results.

Dataset shows that the most typical 5th grader student of English in the community school is at the CEFR level of A1.3. This means that the typical student can read familiar and some unfamiliar words, can understand very short messages dealing with everyday life and routine events or giving simple instructions. In the institutional schools, the typical 5th grader student of English is at the CEFR level of B1.1. This means that the typical student can read a few pages of a wide variety of texts about familiar topics, following the main points, key words and important details even without preparation.

5.1.2 Results Based on Diversity Factors

Diversity is a relative and contextual term. Although many more diversities can be found in Nepalese context, the background information questionnaire in test included altogether six of the diversities. Among them only three are handled in this section which are district-wise, school type-wise (community/institutional), and school location-wise (rural/urban) diversity. These factors can be taken as equality factors since all children regardless of their sex, language, birth place, or family background should have equal opportunities to reach the same educational goals.

District variations in student achievement

The variations in the achievement of grade 5 students from the 28 sample districts in English are presented in table 5.12. The table presents the districts in ascending order according to the achievement. The mean represents the average achievement percentage of the particular district.

Table 5.12 Average achievement score in sample districts

Districts	N	Mean	SD	CV	Districts	N	Mean	SD	CV
Kathmandu	1546	80.2	13.0	16.2	Humla	153	47.9	14.3	29.9
Lalitpur	495	76.8	15.1	19.7	Kailali	813	47.3	20.7	43.8
Bhaktapur	376	75.8	14.4	19.0	Myagdi	296	46.7	22.2	47.5
Kaski	670	72.4	17.5	24.2	Sindhuli	561	46.5	21.6	46.5
Baglung	614	60.0	21.0	35.0	Salyan	550	44.6	21.7	48.7
Solukhumbu	254	56.8	18.0	31.7	Achham	524	43.7	21.2	48.5
Baitadi	605	56.3	20.3	36.1	Rolpa	474	40.1	19.7	49.1
Dolakha	421	56.1	20.5	36.5	Udayapur	521	39.4	21.3	54.1
Parsa	422	54.5	18.2	33.4	Dhankuta	348	39.4	20.9	53.0
Makwanpur	671	50.9	22.3	43.8	Manang	12	38.5	17.7	46.0
Chitwan	622	50.8	22.3	43.9	Mahottari	453	36.6	18.7	51.1
Bardiya	348	49.5	23.9	48.3	Saptari	526	36.5	18.8	51.5
Darchula	401	48.7	22.3	45.8	Jumla	130	32.9	18.8	57.1
Kapilbastu	531	48.7	19.4	39.8	Khotang	457	29.2	17.3	59.2
Total						**13,794**	**53.6**	**24.2**	**45.1**

Of the randomly selected districts in the sample, the students' performance was very low in Khotang (29%), Saptari (37%), Dhankuta (39%) and Udayapur (39%) from the Eastern region, in Mahottari (37%) from the Central region, in Manang (37%) from the Western region, and Jumla (33) and Rolpa (40) from the Mid-Western region. Except for Kaski district (72%), the outperforming districts come from the Central region specifically from the Valley area: Kathmandu (80%), Lalitpur (77%), and Bhaktapur (76%). Comparison will be unfair because 74% of the schools in the Kathmandu Valley are institutional ones and in Kaski 70% are institutional, while in the other districts in the sample, on average, only 10% were institutional ones. Out of the eight lowest performing districts, five had no institutional school. The result means that when the schools' language of instruction is English, it certainly has a positive effect on children's proficiency in English. From this perspective, interesting results are found in the districts where the number of English medium schools is low but the results are higher than the national average. Some examples of these districts are Parsa (no private schools, mean score 54%), Solukhumbu (one private school, mean score 57%), and Baitadi (one private school, mean score 56%).

The difference in achievement due to the district is statistically significant ($p < 0.001$). The variation explained in achievement due to the district is $\eta^2 = 0.358$, that is, the district explains 36% of the variation in the data which is very high percentage. Effect size is $f = 0.75$– indicating that the difference between the lowest performing district (29%) and highest performing district (80%) is remarkably high.

The dataset strongly suggests 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 English. Though the results are bound to the randomly selected 28 sample districts; even lower-performing districts could also be found if other districts would have been selected. The results are very high in the districts where the proportion of institutional schools is high. It is obvious that when the medium of instruction in institutional schools is English, it has remarkable positive effect in English language proficiency for the students.

Ecological zone and student achievement

The Mountain, Hill and Tarai are three geographical features in Nepal though the Valley is taken as a special geographical feature because of its unique feature in terms of population composition, availability of modern development facilities, economic and job opportunities. The variation in achievement among the Ecological zones is presented in table 5.13.

Table 5.13 Achievement in the Ecological zones

Ecological zone	**Community schools**				**Institutional schools**			
	N	**Mean**	**SD**	**CV**	**N**	**Mean**	**SD**	**CV**
Mountain	1,206	47.6	19.6	41,3	165	74.4	14.8	19.9
Hill	5,159	42.8	20.5	48,0	1,132	76.8	14.2	18.5
Tarai	3,174	42.1	18.9	44,9	528	71.8	16.0	22.3
Valley	568	65.6	15.3	23,3	1,849	82.8	10.4	12.5
Total	**10,107**	**44.4**	**20.4**	**45,9**	**3,674**	**79.0**	**13.4**	**17.0**

The data shows first that, on average, the students from the Valley outperform the students from other ecological zones. The difference is wider in community schools (66% compared to 42–48% in the other zones) than in institutional schools (83% compared to 72–77%). In both the community and private schools, the students from the Tarai area are performing the lowest (42% in community schools and 72% in private schools) though the difference is not notable compared with the Hill zone (43% and 77%). It is also notable that the exceptionally low value for the Coefficient on Variation in the Valley which, in community schools, is about half of that in the other areas. The obvious reason for this is the systematically high score in Valley compared with the other areas.

The achievement between the zones differs significantly in both the schools types ($p < 0.001$) even if the Valley is excluded from the analysis. Tukey's *post hoc* test explains that, in community schools, there is no difference between Hill and Tarai but the students from the Mountain zone differ from the students of both Hill zone ($p < 0.001$) and Tarai zone ($p < 0.001$). In institutional schools, there is no difference between Mountain and Tarai but the students from Hill differs from the students of Tarai ($p < 0.001$). Ecological zone explains 7% of the variance in community schools ($\eta^2 = 0.071$) and 10% in institutional schools ($\eta^2 = 0.096$).[59] In comparison, the district explains more than 36% of the variation. The effect size is $f = 0.28$ in community schools and $f = 0.33$ in institutional schools showing moderate difference between the highest and lowest performing Ecological zones. The effect sizes are small if the Valley is excluded from the analysis ($f = 0.08$ and $f = 0.15$ respectively). This means that the real differences are not remarkable between the Ecological zones but the Valley differs distinctly from the other areas. From equality point of view among the zones, this can be taken as a possible good sign.

Development region and student achievement

The students' achievement varies according to the Development regions, which are divided into Eastern, Central, Western, Mid-Western, and Far-Western. Additionally, the Kathmandu Valley is taken as the 6th development region though adiministratively it falls under the Central Development region. The mean achievements within the Development regions are given in table 5.14 and illustrated in figure 5.9.

Table 5.14 Achievement in the Development regions

Development Region	Community schools				Institutional schools			
	N	Mean	SD	CV	N	Mean	SD	CV
Eastern	2,011	37.3	19.9	53.4	82	73.7	12.5	17.0
Central	2,742	45.2	19.7	43.6	408	76.2	14.5	19.0
Western	1,296	46.8	17.9	38.3	827	78.5	12.6	16.0
Mid-Western	1,505	40.9	19.8	48.5	150	71.8	14.6	20.4
Far-Western	1,985	45.7	20.2	44.3	358	67.9	18.2	26.8
Valley	568	65.6	15.3	23.3	1,849	82.8	10.4	12.5
Total	**10,107**	**44.4**	**20.4**	**45.9**	**3,674**	**79.0**	**13.4**	**17.0**

59 If the Valley is excluded from the analysis, the values for Eta squared would be 0.007 and 0.022 respectively, that is, only 1% and 2% explanation. The role of the Kathmandu Valley students in the whole national mean is remarkable.

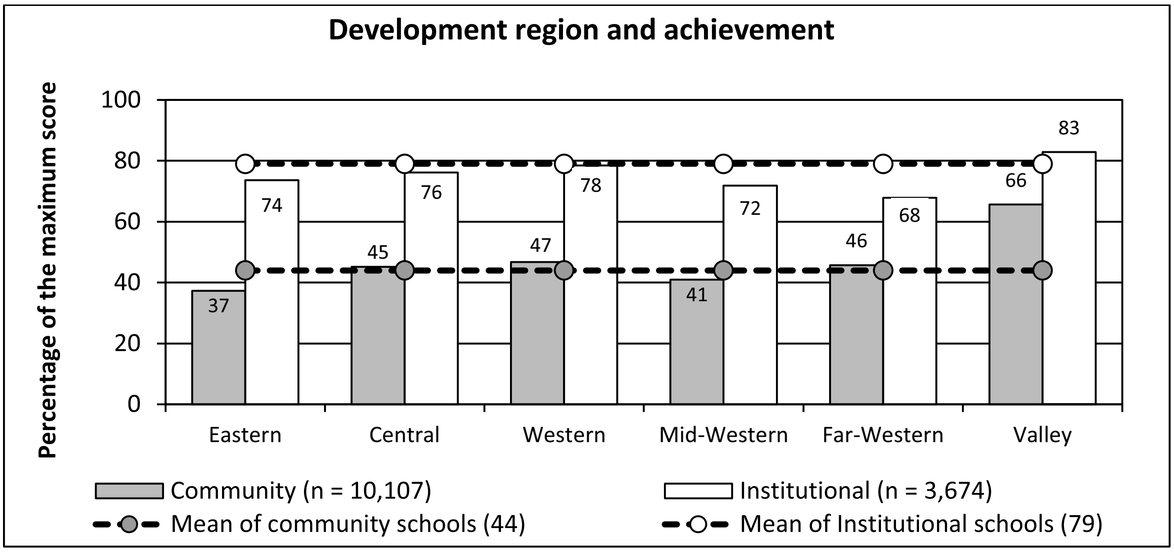

Figure 5.9 Comparison of student achievement of the Development regions

The highest performance is found in institutional schools in the Kathmandu Valley (83%) and in the Western region (78%). The performance is the lowest in community schools in the Eastern (37%) and Mid-Western (41%) regions. The difference between the regions is statistically significant both in community and institutional schools ($p < 0.001$). Tukey's *post hoc* test shows that, in community schools, the average achievement level in the Eastern region is significantly lower than the other regions ($p < 0.001$). In the Valley the achievement is higher than in all other regions ($p < 0.001$). Also, in the Mid-Western region the average achievement level is lower than any other region except for the Eastern region ($p < 0.001$). There is no difference between Central, Western, and Far-Western regions when it comes to the achievement level of community schools in English. In institutional schools, Tukey's *post hoc* test shows that the students in the Valley outperform the students of all other regions ($p< 0.001$). In the Far-Western region the students perform lower than in other region ($p < 0.01$).

Development region explains 9% of the variance in community schools ($\eta^2 = 0.092$) and 13% in institutional schools ($\eta^2 = 0.128$).[60] This is somehow the same proportion as found with the Ecological zone. It is notable that the district explains more than 36% of the variation which means that in the Development regions there are lower and higher performing districts. The effect size is $f = 0.32$ in community schools and $f = 0.39$ in institutional schools–showing moderate or wide difference between the highest and lowest performing regions. The effect sizes are moderate if the Valley is taken out of the analysis ($f =0.18$ and $f = 0.2$ respectively). Compared with the Ecological zones, the differences are wider between the regions than between the zones.

The dataset reveals that there is wide inequality among the development regions regarding children's opportunities to reach an expected level in English. Particularly, the wide

60 If the Valley is taken out of the analysis, the values for Eta squared would be 0.031 and 0.074 respectively, that is, only 3% and 7% explanation – one third and half of those with the Valley included in the analysis. The role of the Kathmandu Valley students in the whole national mean is remarkable.

difference between the community schools in the Valley and in the rest of the country (29 percent as the highest) is a clear sign of inequality of opportunities in learning English. There are also wide differences between the regions in institutional schools; the difference in student performance in private schools between the Valley and Far-Western is the highest, that is 15 percent.

School type and student achievement

All the schools are categorized into community and institutional (that is, private schools). The differences in the English achievement have been handled within the sections above. Here the main differences are presented in table 5.15.

Table 5.15 Type of school and the average achievement

Content area	Community (n =10,107)			Institutional (n = 3,674)			Mean difference	Cohen's *d*
	Mean	SD	CV	Mean	SD	CV		
Reading	40.6	21.7	53.3	76.5	15.6	20.4	35.9	1.84
Writing	39.8	22.0	55.3	74.4	15.9	21.4	34.6	1.78
Grammar	48.8	26.0	53.4	78.0	19.9	25.6	29.2	1.69
Vocabulary	50.1	24.6	49.0	79.3	14.8	18.6	29.2	1.19
Total	44.4	20.4	45.9	79.0	13.4	17.0	34.6	1.30

The achievement levels in community schools and institutional schools differ from each other remarkably as noted above. The average performance in total score in institutional schools is 79% whereas in community schools it is 44%; thus, 35% percent difference is remarkable. The difference is statistically significant ($p < 0.001$) and the effect size is very high (d = 1.30) showing that community schools are far below the institutional schools. Difference is the highest in the content areas of Reading (d = 1.84) and very high in Writing (d = 1.78). Division of the students to the community and institutional schools explains 38% of the student variation in reading (η^2 = 0.382) and 36% in Writing (η^2 = 0.358). Most private schools in the sample show very high performance. One of the reasons for this is the medium of language for instruction in the schools as most of the private schools use English as the medium of language for instruction.

The dataset is evident that, on average, the students in institutional schools outperform the students in community schools. This deviance may be explained by the English as the medium of language for instruction.

School location and student achievement

The schools were divided into rural and urban schools. This information was obtained from the head teacher though some of the head teachers did not inform about school location. The achievements of the students in rural and urban schools are presented in table 5.16.

Table 5.16 Student achievement on the basis of location of school by school type

Location of school	**Community**				**Institutional**			
	N	**Mean**	**SD**	**CV**	**N**	**Mean**	**SD**	**CV**
Rural	8,463	43.3	20.1	46.4	1,225	76.5	14.8	22.5
Urban	873	55.8	18.4	37.3	2,212	80.2	12.6	18.7
Mean difference		12.4				3.7		
Cohen's *d*		-0.52				-0.33		
Total	**10,107**	**44.4**	**20.4**		**3,674**	**79.0**	**13.4**	

The achievement level of the students in the urban community schools (56) is 12 percent higher than that of rural community schools (43). The difference is statistically significant ($p < 0.001$) and the effect size is moderate ($d = 0.52$). Excluding the community schools of the Valley, the score of the urban community schools lowers to 49. The difference (6 percent) is still statistically significant ($p = 0.013$) but the effect size is low ($d = 0.22$). The main difference in community schools is, hence, caused by the higher level of the students in the Valley schools. The division into rural and urban schools explains 3.3% of the student variation in community schools ($\eta^2 = 0.033$), and excluding the Valley schools it is only 0.6% ($\eta^2 = 0.006$). The latter is a good sign from equality point of view. Except for the schools in the Kathmandu Valley, there is no difference between rural and urban community schools.

The achievement level of the students in the urban institutional schools (80%) is 4 percent higher than that of rural institutional schools (76%). Though the difference is statistically significant ($p< 0.001$), the effect size is moderate ($d = 0.33$). If community schools are excluded from the schools in the Kathmandu Valley, the difference remains the same (3.5 percent) which is still statistically significant ($p = 0.001$) but the effect size is low ($d = 0.23$). In institutional schools, the effect of the Kathmandu Valley is not remarkable. The division into rural and urban schools explains 2.2% of the student variation in institutional schools ($\eta^2 = 0.022$) and 1.3% ($\eta^2 = 0.013$) when the school from the Kathmandu Valley are excluded.

Data shows that the students in the urban community schools have achieved 12 percent more than the students in the rural areas. Excluding the Valley schools, the difference is 6 percent. However, from educational equality point of view, the difference is not a good sign though the real difference is not wide within community schools. In institutional schools, there is not wide difference between the rural and urban areas.

Language at home and student achievement

In the context of Nepal, students' achievement tends to depend on the language spoken at their homes i.e., the mother tongue of the students. The mother tongue reflects, in many cases, the ethnic background, and hence ethnic difference can be taken as a possible source of inequality in educational achievement.

On the basis of the data, 36.5% of the 5th graders speak a language other than Nepali as their first language. These "other" languages are quite fragmented; the largest groups in the English dataset are Tharu (4.6%), Newari is (4.4%), and Urdu (3.8%). After dividing the languages into ten groups excluding Nepali, there were still 18.3% students who were classified into the group "Other". Because the languages are very fragmented and the Nepali speakers are the majority of the students, for the purpose of the statistical analysis, all the other languages are first grouped into "Non-Nepali". The results are presented in tables 5.17(a) and 5.17(b).

Table 5.17(a) Student achievement on the basis of home language

Language group	Community				Institutional			
	N	Mean	SD	CV	N	Mean	SD	CV
Nepali	5,403	45.4	20.3	44.7	2,967	79.9	12.6	15.7
Non-Nepali	4,142	43.1	20.7	47.9	657	74.5	16.1	21.7
Mean difference		2.3				5.4		
Cohen's *d*		-0.112				-0.409		

When all the minor language groups are combined as "Non-Nepali", no notable difference is found between the language groups in community schools (2 percent favouring the Nepali speakers). Though the difference is statistically significant ($p < 0.001$), the effect size is low ($d = 0.11$). The difference of 5 percent in institutional schools is moderately high ($p < 0.001$, $d = 0.41$).

On the basis of the original categorization of the minority languages, the issue looks quite much interesting as it is evident that the Magar and Tamang students are at quite much higher level in English than the Nepali speaking students (57% and 56% compared to 45% in community schools). On the other hand, the students from Sherpa (24%) and Gurung (25) background perform much lower than the average. Magar students (with the average achievement of 82%) slightly outperform the other language groups (less than 80%) in institutional schools.

Table 5.17(b) Achievement variation in different language groups

Language	Community				Institutional			
	N	Mean	SD	CV	N	Mean	SD	CV
Tamang	315	57.3	17.4	30.3	12	79.0	11.6	14.7
Magar	44	55.8	19.4	34.8	175	82.2	9.0	10.9
Urdu	451	46.5	20.5	44.2	48	79.1	9.8	12.3
Nepali	5,403	45.4	20.3	44.7	2,967	79.9	12.6	15.7
Newari	513	40.8	17.4	42.5	65	74.1	12.3	16.6
Awadhi/Maithili	27	40.1	13.9	34.7	6	78.9	10.1	12.8
Rai	84	39.0	22.8	58.3	3	65.0	36.4	56.0
Tharu	570	36.8	19.1	51.8	30	77.3	12.5	16.2
Gurung	26	25.3	18.3	72.2	9	77.5	8.6	11.2
Sherpa	16	23.9	15.0	62.9	-	-	-	
Other	2,095	42.8	21.1	49.2	307	68.9	19.0	27.6

1)Those language groups in which number of the students was less than 10 are omitted.

The differences between the highest and lowest performing groups of students are statistically significant ($p < 0.001$) and notable; the effect sizes are moderately high ($f = 0.18$ in community schools and $f = 0.25$ in institutional schools). The division into smaller language group explains about 3% and 6% of the variation in the data ($\eta^2 = 0.032$ within community schools and $\eta^2 = 0.058$ within institutional schools). Though the differences are wide between the extreme groups, it is good to keep in mind that the number of students is quite small in some of the language groups indicating a moderate effect size. When analysing only the minority languages and hence, excluding the Nepali speakers and the group "Other", the effect size is high (f = 0.40) in community schools indicating remarkable difference between the highest performing minority group (Tamang, 57%) and the lowest performing group (Sherpa, 24%).

Table 3.5.18 Achievement in the different language groups in different regions

Development Region	Nepali1	Tharu	Newar	Urdu	Tamang	Magar	Rai	Gurung	Maithili	Sherpa
Eastern	41,3[2]	35,0	34,8	39,5	41,1	44,1	39,4	34,0	31,4	21,0
Central	53,0	39,5	48,5	43,6	57,8	49,6	38,3	39,5	40,5	
Western	61,6	67,1	38,8	53,2	25,3	69,4		60,0	50,9	67,1
Mid-Western	44,8	79,7	47,4			62,0		13,9	31,6	
Far-Western	49,5	56,5	46,3	45,6	50,6		24,1	10,9		
Valley	79,1	81,4	67,7	72,8	83,7	80,2	80,4	87,3	65,8	
Total	57,6	38,6	44,6	49,6	58,1	76,9	39,9	38,7	47,2	23,9
N[3]	8370	606	578	499	327	219	87	35	33	16

1) The language groups of less than 10 students are not included in the table.
2) The main population is highlighted by the gray shade. In some un-highlighted cases there is only one student behind the mean.
3) The language groups are ordered on the basis of their frequency.

When combining the results from the Development region and mother tongue, one notices that the achievement score of the students within a certain language group varies sharply among the different regions. All language groups except for Newari and Maithili have a high score in the Valley. Almost all language groups perform lowest in the Eastern Development region. Especially, Sherpa students' perform low in the Eastern region.

Dataset reveals that there are statistically significant though not necessarily remarkable differences between the castes in English. Dalit (43%) and Janjati students (44%) as well as "Other" ethnic (40%) are performing significantly lower than Brahmin, Chhetri, and Madhesi groups. Dalit students perform lower especially in the Far-Western and Eastern Development regions.

The dataset shows that there is an educational inequality in language groups in possibilities of learning English. In community schools, the students from Magar (59%) and Tamang (55%) backgrounds perform very high in English while those from Sherpa (24%) and Gurung (25%) background perform very low. The differences between the language groups are remarkable.

Ethnicity/Caste and student achievement

Historically, it is seen that Dalits were deprieved of or had their low level of educational attainment. Hence, modern society has made lots of efforts to make education possible and accessible for all children. So their participation rate at lower level of schooling has increased remarkably.The latest household survey (CBS, 2012) shows that their number in the secondary and higher education is still very small despite the lots of efforts have been put towards making education accessible for them. The results concerning the ethnicity/castes and achievement are depicted in figure 5.10.

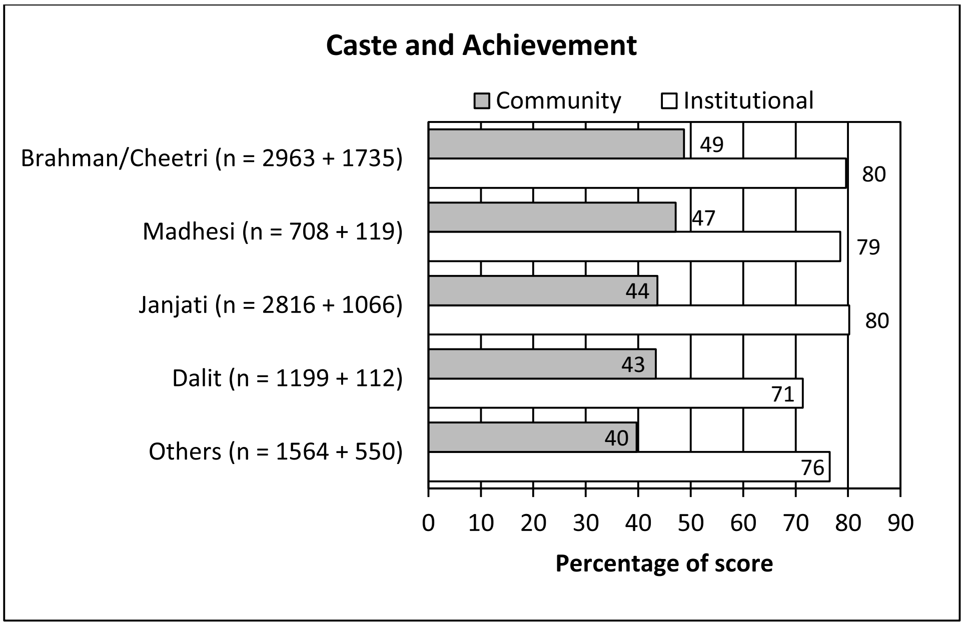

Figure 5.10 Relation between ethnicity/caste and achievement

On the basis of NASA 2012, with community schools, the students from "Other" ethnic/castes are performing the lowest (40%) in English, followed by Dalit (43%) and Janjati students (44%). Dalit students perform below the average also in institutional schools. The overall difference between the groups is statistically significant (p < 0.001) though the effect size is small or medium (f = 0.16) in community schools. Division of students according to their caste background explains just 2.5% of the student variation (η^2= 0.025). Within the institutional schools, the effect size is also small (f = 0.15). Similarly, division of students according to their ethnic background explains just 2.1% of the student variation (η^2= 0.021). From equality point of view, this is a good sign though there is still a lot to do to reduce the gap between the ethnicity.

Dalit students have been followed up because they have historically been deprived from education. A positive sign from equality point of view is that Dalit students perform better than the national mean (45%) in the Eastern (58%), Central (51%), and Mid-Western (48%) Mountain areas as well as in Western Tarai (51%) (Table 5.19). However, it is evident that results are much lower than average in the Eastern Tarai (35%), Mid-Western Hill (38%), Far-Western Mountain (36%) and Far-Western Tarai (39%). Generally speaking, the few Dalit students in institutional schools (n = 112) perform always lower than the average.

Table 5.19 Dalit students' achievement in different Ecological zones and Development regions

Type of school	Ecological zone	Development Region					
		Eastern	Central	Western	Mid-Western	Far-Western	Total
Community	Mountain	57.9	50.6		48.4	35.9	47.2
	Hill	32.2	44.7	44.6	38.0	46.7	42.8
	Tarai	35.0	40.4	51.2	44.1	38.8	41.6
	Total	36.9	42.8	46.2	41.0	43.6	43.4
Institutional	**Ecological zone**	**Development Region**					
		Eastern	**Central**	**Western**	**Mid-Western**	**Far-Western**	**Total**
	Mountain	79.7	71.4			42.2	64.3
	Hill	66.5	74.3	70.3	73.1	59.8	69.2
	Tarai		70.7		81.3	51.1	69.1
	Total	**73.1**	**71.6**	**70.1**	**77.2**	**54.2**	**71.3**

Dataset reveals that there are statistically significant though not necessarily remarkable differences between the ethnicities/castes in English. Dalit (43%) and Janjati students (44%) as well as "Other" castes (40%) are performing significantly lower than Brahmin, Chhetri, and Madhesi castes. Dalit students perform lower especially in the Far-Western and Eastern development regions.

Gender and student achievement

Lots of effort has been put globally into reducing the difference between boys' and girls' school achievement because of modern discourse in favour of gender equality. The matter is handled somehow more extensively than in the previous sections of equality. Basic results are presented in table 5.20 and figure 5.11.

Table 5.20 Student achievement of boys and girls by school type

Gender	Community				Institutional			
	N	Mean	SD	CV	N	Mean	SD	CV
Boys	4,604	44.8	20.6	45.9	1,994	78.2	13.5	17.2
Girls	4,901	44.2	20.4	46.1	1,613	79.8	13.5	16.9
Total	**10,107**	**44.4**	**20.4**	**46.0**	**3,674**	**79.0**	**13.4**	**17.0**

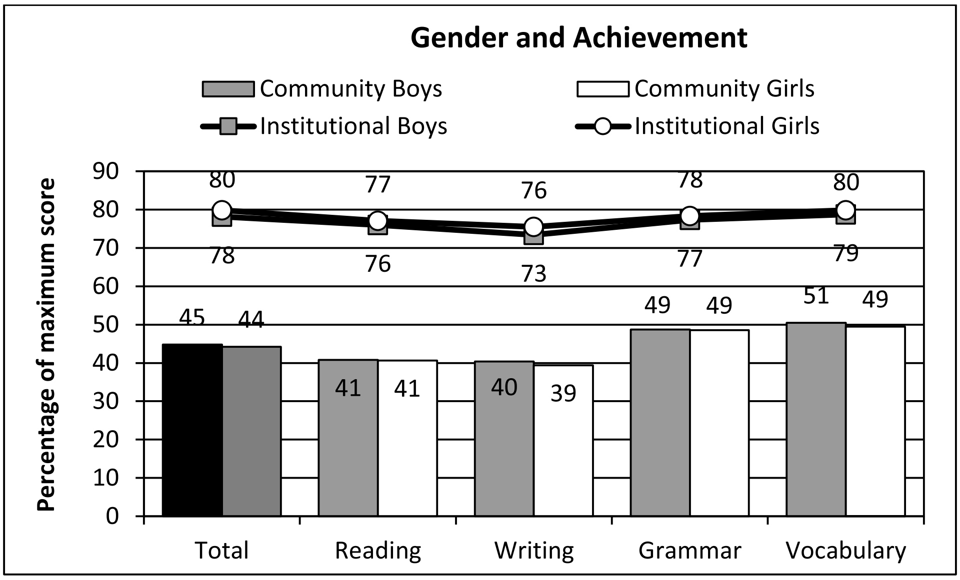

Figure 5.11 Comparison of the achievement of boys and girls in various content areas

There are no statistically significant differences between boys and girls in community schools in any of the content areas of English. The differences in institutional schools are also very small though the girls are seen to be performing slightly higher than boys in writing ($p < 0.001$) and in the total score ($p < 0.001$). The effect sizes are, however, small ($d = 0.13$ and $d = 0.12$ respectively). This is a positive indication towards gender equality

In community schools, there is no difference between genders in the ecological belts (figure 5.12). In institutional schools, girls are performing slightly better in Mountain and the Kathmandu Valley. When it comes to the Ecological zones, the differences between boys and girls are very small.

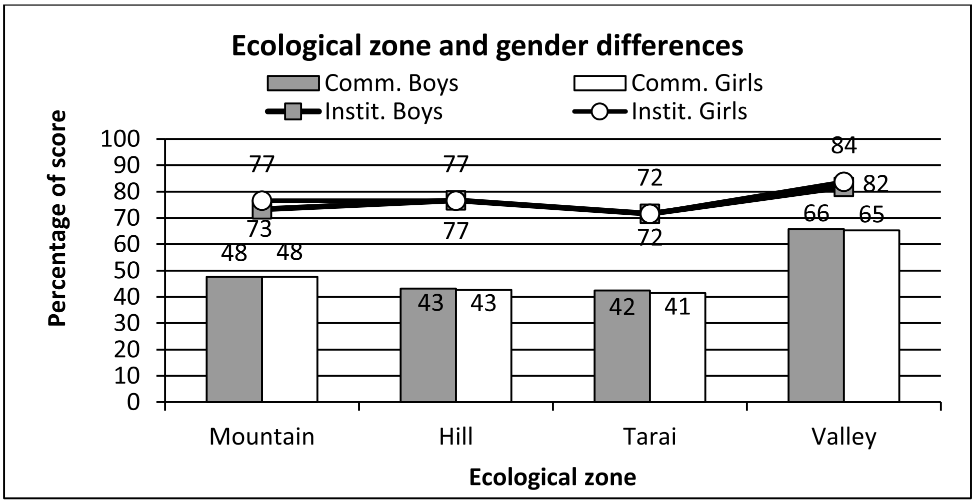

Figure 5.12 Ecological zone and gender-wise differences in achievement by school type

There are no notable differences between the Development regions which indicates that boys' and girls' have equal opportunities to reach the same educational goals (figure 5.13). The difference between boys and girls is seen to be somehow wider in community schools in the Far-Western region (3 percent) and in institutional schools of the Kathmandu Valley (2 percent).

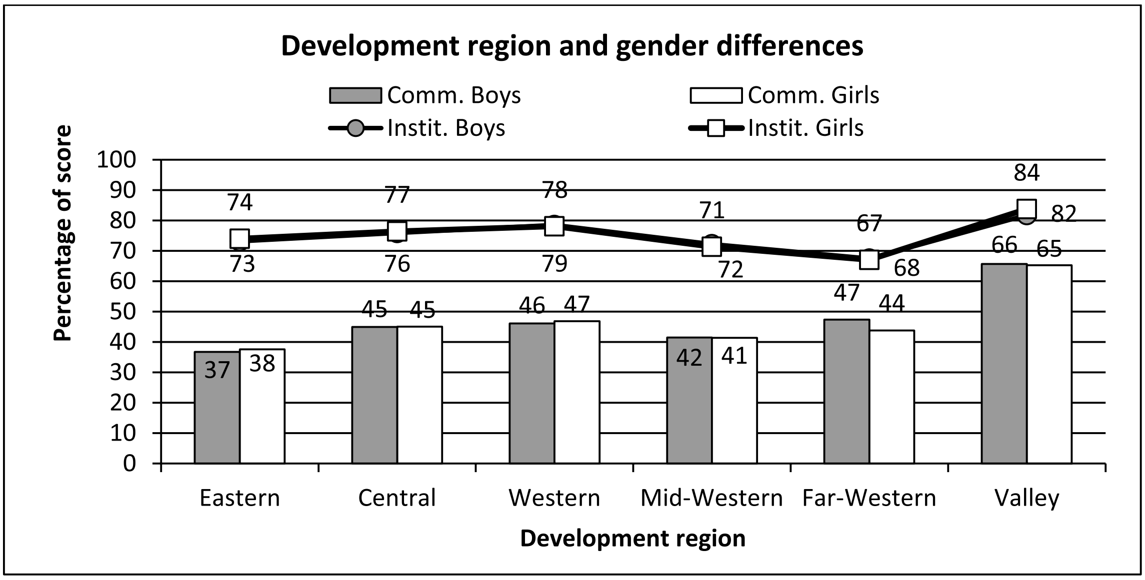

Figure 5.13 Development region and gender-wise differences

The dataset shows that the difference in achievement between boys and girls in English proficiency are very moderate. Though the differences in institutional schools are statistically significant in total score and writing, the effect sizes are very small–indicating that the differences are not at all remarkable, which is a positive sign towards the gender equality. A tendency is also that girls are slightly out-performing boys in institutional schools and boys are slightly outperforming girls in community schools.

5.1.3 Selected Explanatory Factors and Achievement

The factors like geographical factors such as districts, the Ecological zone, and Development region, as well as school related technical factors, such as school type and school location have already been discussed above. Some individual related factors also were handled, such as home language, caste and gender.

In this section, some other factors such as family, student's individual, school and teacher related factors are taken into consideration. These include the socio-economic status (SES) of the students' families, paid work after school, students' attitude towards English as a school subject, age of the student, and support provided to the studies. As the school and teacher-related factors, the availability of school books, homework given by the teacher, and selected activities in the school are also handled.

Parents' education, occupation, home possessions and student achievement

There are several variables indicating the socio-economic status which were categorized into parents' education, parents' occupation, home possessions, home accessories, and whether the student attends a private school or not. Finally, the SES is estimated on the basis of seven indicators related to the economic, educational, and occupational background of the family. In this section, education of the parents is further elaborated, so that the parents' illiteracy is taken into account in relation to the English language achievement.

Several SES-related variables were analysed by using a data mining tool of SPSS and DTA. The method is very effective in finding the cut-offs of the predicting variable, such as mothers' education, and classifying the factor into several groups, which differ statistically in the most significant way from each other in relation to student achievement. Some examples of this are handled with parents' educational background and its relation with students' achievement in English.

Parents' education

Parents' education is divided into eight categories: 1) illiterate, 2) literate, 3) grade 10 pass, 4) SLC pass, 5) IA pass, 6) BA pass, 7) MA pass, and 8) Above MA pass. As the information was obtained from the students, there is likely to have been some impurities embedded in the data since the number of (just) literate mothers in the dataset is seen to be too high. However, with the huge dataset the result is seen to be credible.

DTA classifies mothers' education into three groups with statistically significant differences in students' achievement levels (figure 5.14), which are illiterate (students' average is 49%), just literate (53%), and grade 10 passed or higher (71%). The difference between each group is statistically significant ($p < 0.001$). In practical words, the results mean that when the mother is at least grade 10 passed, her children achieve, on average, + 22 percent advance in the national test compared with illiterate mother. Figure 5.16 shows that when the mother is MA passed or higher, the advance was + 37 percent over the illiterate mother. Mothers' education explains 10% of the student variation ($\eta^2 = 0.096$), which indicates a moderate effect size ($f = 0.33$). Obviously, the result means that the children of the highly educated mothers are mainly found in private schools.

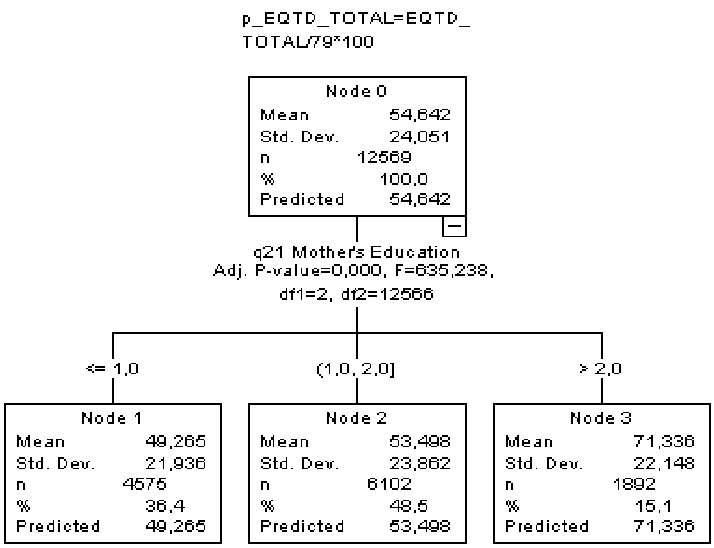

Figure 5.14 DTA of mothers' education and students' achievement

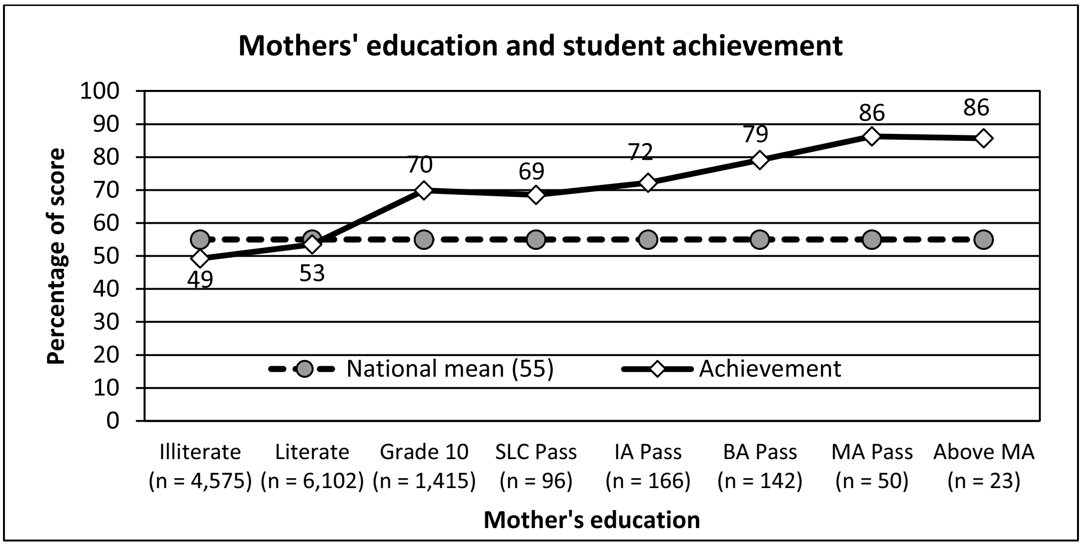

Figure 5.15 Mothers' education and students' achievement

In parallel, DTA divides fathers' education into four categories: illiterate (48%), (just) literate (51%), grade 10 passed (65%), and SLC passed or higher (71%). The difference between each group is statistically significant ($p < 0.001$). In practical words, the results mean that when the father has passed grade 10, his child achieves, on average, + 17 percent point advance in the national test compared with illiterate father. Figure 5.17 shows that when the father was MA passed or higher, the advance was + 34 percent over the illiterate father. Obviously, the high average means that the children from the highly educated fathers (as well as of mothers) are mainly found in private schools. Fathers' education explains 11% of the student variation ($\eta^2 = 0.110$) which indicates a moderate or high effect size ($f = 0.35$).

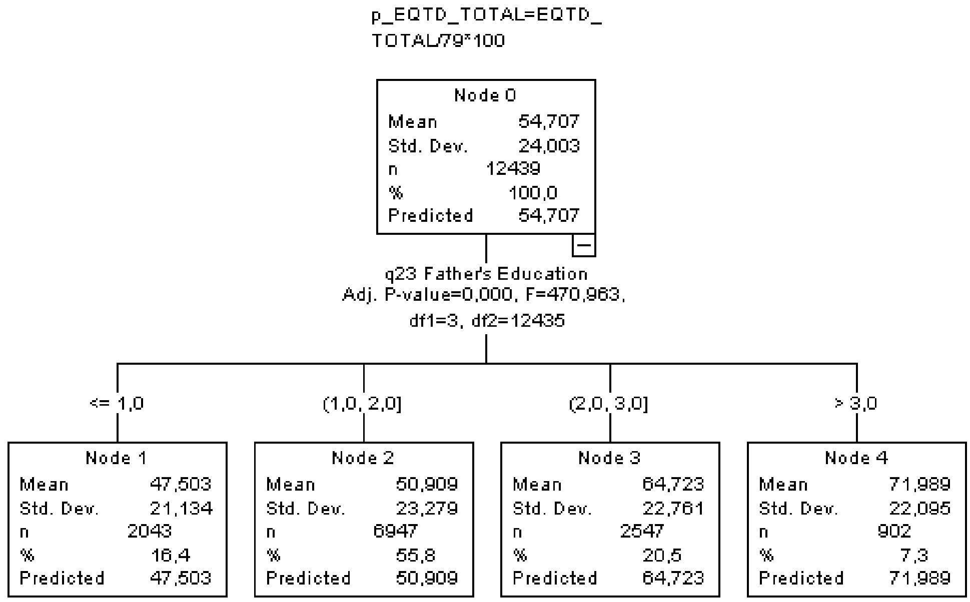

Figure 5.16 DTA of fathers' education and students' achievement

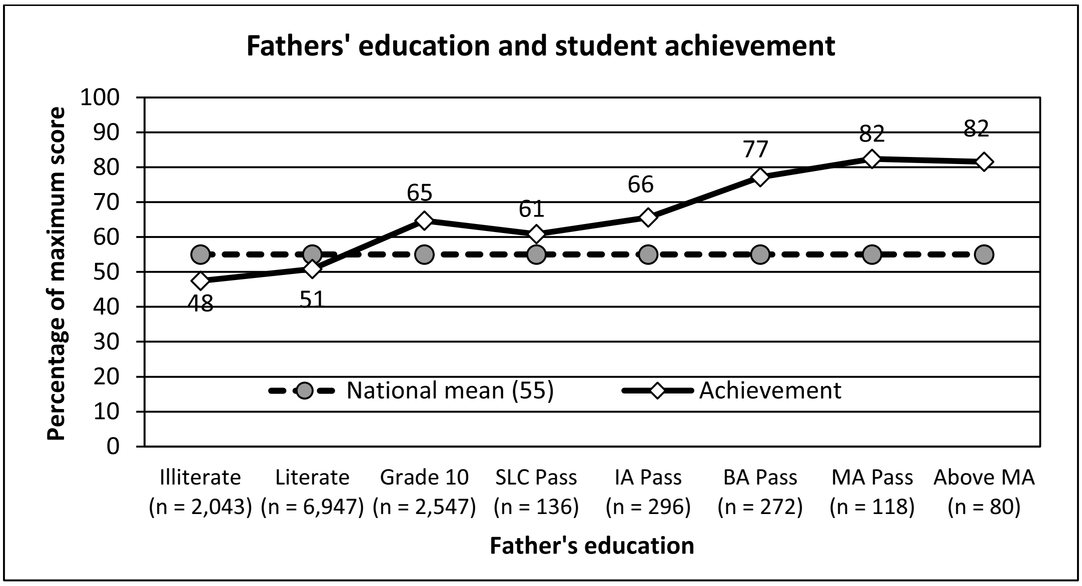

Figure 5.17 Fathers' education and students' achievement

It is noteworthy that mothers' education tends to give slightly higher leverage to children's achievement in English than that of fathers'. While father's literacy is found to have raised the child's achievement by 17 percent, mother's literacy has raised it by 22. In parallel, while father's high education (MA or higher) contributes to raise the child's achievement by 34 percent, mother's high education is contributing to raise it by 37%. In both cases, the effect size is moderate or high (f = 0.32–0.35) showing that the difference between the highest and lowest group is remarkable.

After combining the mothers' and fathers' education, the poorest prediction in DTA for the children's future achievement in English comes when the father is (just) literate but the literacy of the mother is not known (37%) or the literacy of the father is not known but it is known that the mother is illiterate (34%). The highest results are found among the groups whose both the father and mother have passed (at least) the grade 10 (74%), or whose father has passed SLC and mother is (at least) grade 10 passed (77%). It is evident that the educational support provided by the parents can be utilized by the students. The higher the parents' education, the better is the results gained by the children.

In what follows with the final SES variable, the cut-off for parental education was set to "grade 10", that is, for grade 10 (or higher),the indicator for mothers' (and fathers') education for SES was set to 1, and the lower education than grade 10 passed gave the value 0.

Parents' occupation

The occupation of parents was categorized into eight groups: 1) working abroad, 2) farming and working at home 3) only working at home 4) teaching 5) services 6) business 7) daily wages and 8) working at other's home. The result related to mother's occupation is seen in figures 5.18 and 5.19 and that related to father's occupation in figures 5.20 and 5.21.

While comparing the students' means by DTA, the achievement is the lowest when the mothers' occupational background is agriculture (49%). It is statistically significantly lower than when the mother works abroad (55%) or only at home (58%), nothing to say when mother works as a teacher (74%) or is in services (76%). Mothers' occupation explains 8.5 percent of the student variation ($\eta 2$ = 0.085) which indicates a moderate effect size (f = 0.30).

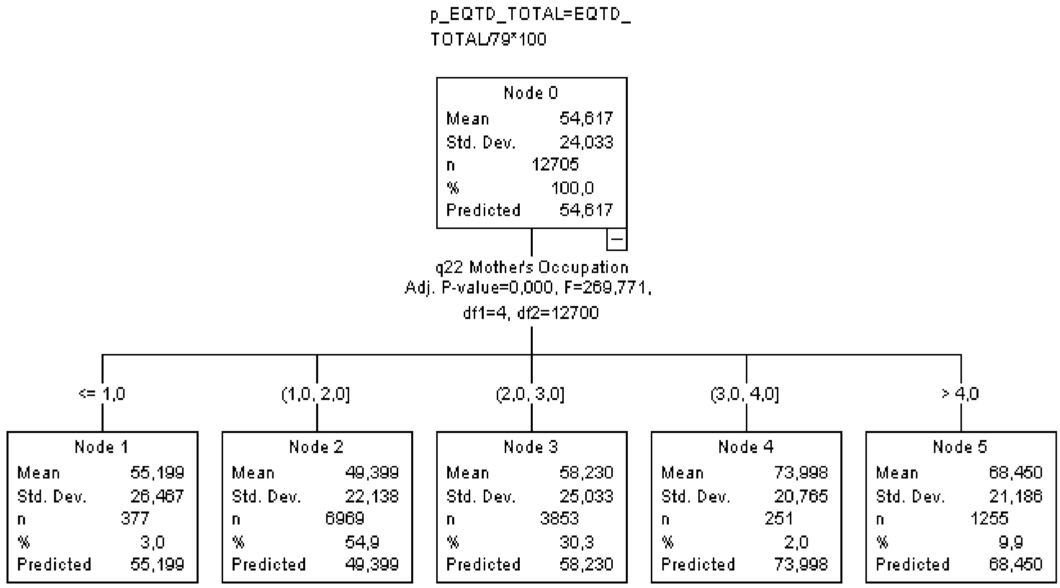

Figure 5.18 DTA of mothers' occupation and students' achievement

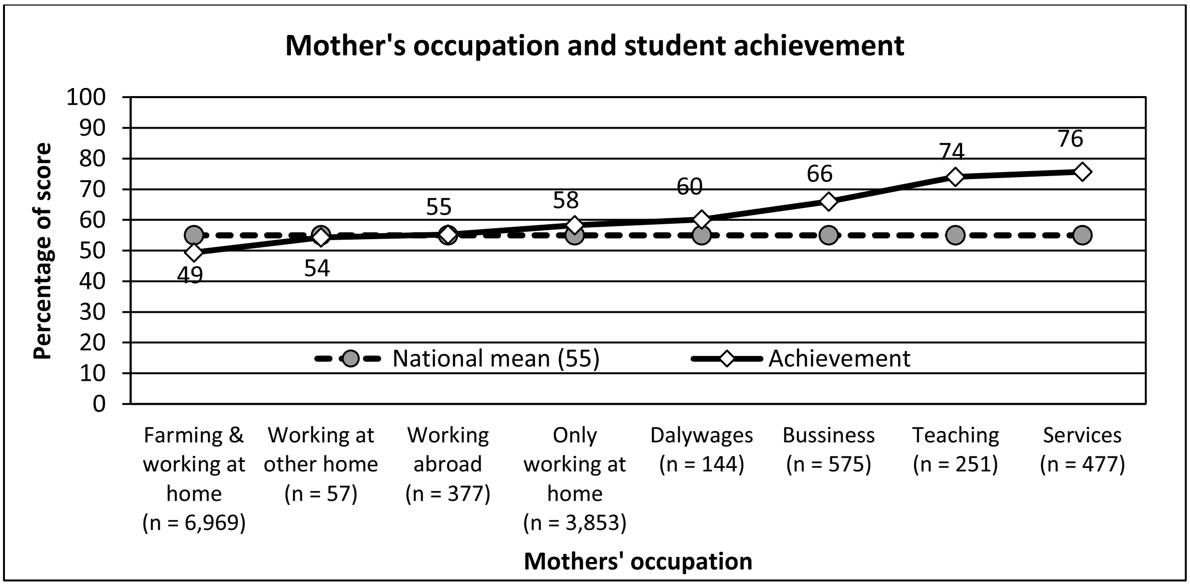

Figure 5.19 Mothers' occupation and students' achievement

On the basis of DTA (figure 5.21), when it comes to the fathers' occupation, the main division is whether the father works for agriculture (46%) or not (> 53%). More precisely, if the father either works for agriculture or works only at home (that is, probably unemployed), the English language skills are remarkably lower (46%) compared with the possibility that the father is in business (65%), teaching (68%) or service profession (72%).

The difference between the lowest and highest group is 27 percent, which is a wide gap. Fathers' occupation explains 17 percent of the student variation (η^2 = 0.166), which indicates a high effect size (f = 0.45).

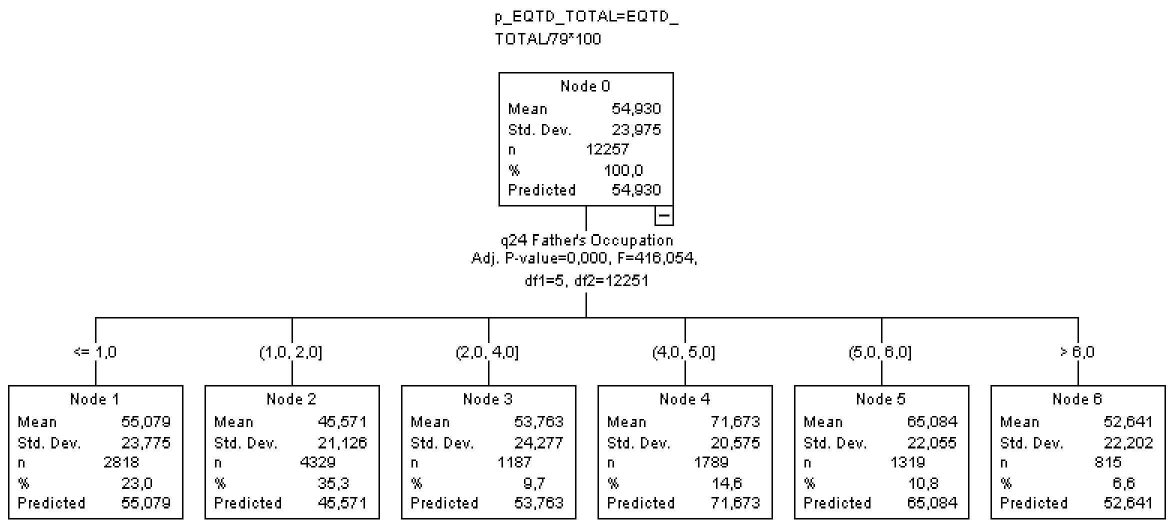

Figure 5.20 DTA of fathers' occupation and students' achievement

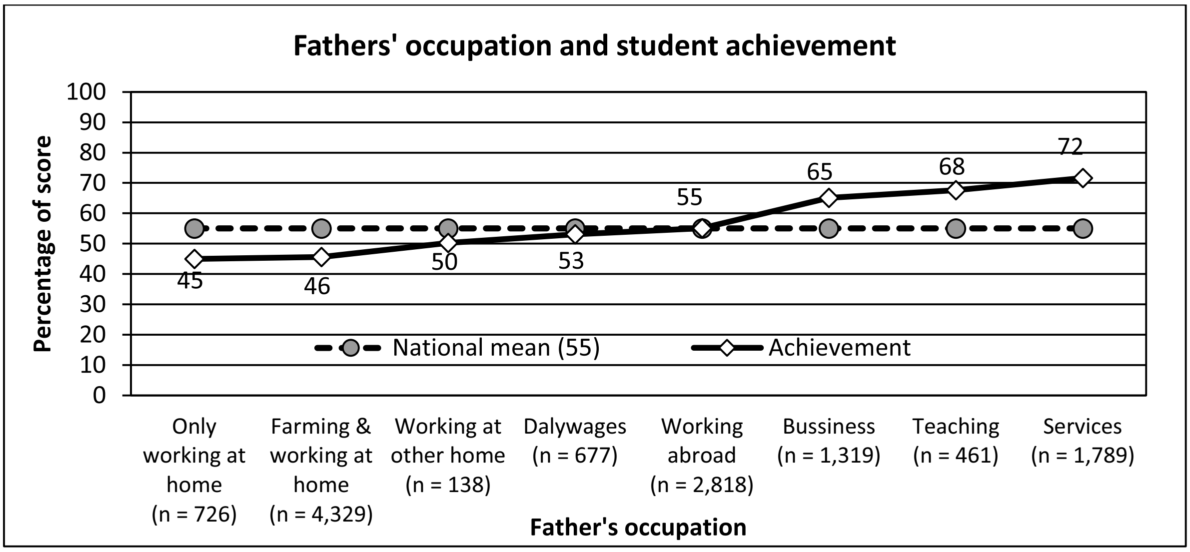

Figure 5.21 Fathers' occupation and students' achievement

When combining the mothers' and fathers' occupation, DTA shows that the lowest achievement is found in the families where either the father or both parents come from an agricultural background (45, n = 4729) or in those few cases where the father is in agriculture related occupation but mothers' occupation is unknown (36, n = 63). The highest achieving students come from the families where both the father and mother are teachers (77, n = 75) or where the father is in services regardless the mothers' occupation (72, n = 1,789). It is worth noting that the services and business occupations are more probably urban than rural .

For the later use as a SES-indicator, the cut-off for the parents' occupation was made so that being in agriculture gives 0 and the rest options give 1.

Home possessions and accessories

Facilities and resources available at home tend to 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 help in studying at home: whether they have a table for study, a separate room for them to study, 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 and poetry books, or artistic things like pictures, and books that help them for study such as a dictionary. Another type of home possession includes different types of normal *home accessories* (and hence, in what follows these are called *home accessories* to differentiate them from *home possessions*) such as the number of mobile phones, televisions, and computers.

There are 11 questions in the student background questionnaire related to home possessions. Each was scored 1 for the access to each of possession (e.g. having a separate room or a table to study). Adding these items up, the maximum score was 11 indicating that the student reported to have access to all of the possessions, and the lower the score the fewer possessions they have at home. Figure 5.22 shows the connection of home possessions and achievement level. Except for the highest category, the achievement level of the students' rises logically the more when there is access to all of these home possessions.[61] Pearson correlation between the achievement level and the factor (r = 0.30) is statistically significant (p < 0.001) and indicates high effect size (d = 0.72).

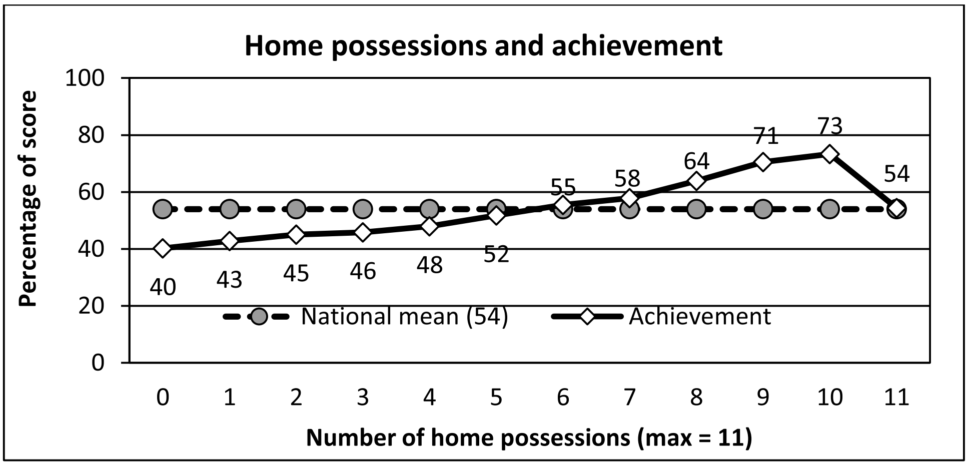

Figure 5.22 Relation between home possessions and achievement

61 The same phenomenon, though not as radical as here, was seen also in 2011 datasets (see ERO, 2013, figures 3.1.24 and 3.2.22): the students who selected all the possibilities may not have understood the question in the same way as the other students. Most probably, in any case, they actually did *not* have all the possessions though they claimed that.

For the later use in SES, the cut-off for the factor was set on 6 possessions: if 6–10 items were met as mentioned in the background questionnaire, the student was given 1 otherwise 0.

The same pattern – the more possession, the better results – can be seen also with home accessories, as seen in figure 5.23. The question in the background questionnaire was set differently compared with home possessions; with the accessories it was asked “how many of the following accessories do you have in your family?” with the options 0 – 3 (or more). For the indicator, the availability of home accessories is dichotomized in the same way as the home possessions. After dichotomizing the items individually by using meaningful cut-offs found with ANOVA and DTA (and maximizing the differences in achievement level), all three indicators were summed up.[62] The maximum score was 3 indicating that the student possessed the set number of all of the accessories.[63]

Table 5.21 Dichotomizing the indicators for home accessories

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

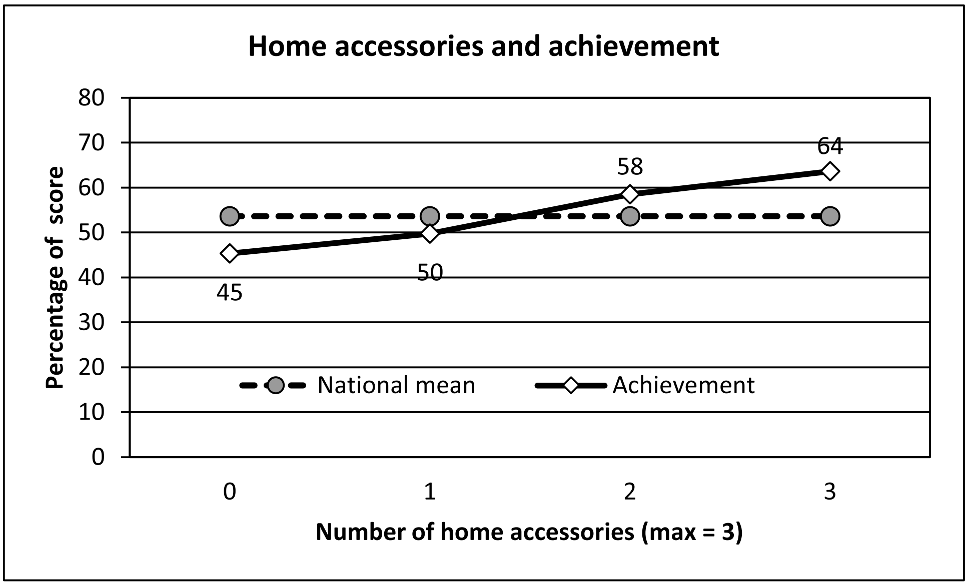

Figure 5.23 Connection between the number of home accessories and achievement

Figure 5.23 clarifies how the increase in the number of home possessions or accessories is associated with the increase in students' achievement. It is seen that the achievement has increased from 45% (if none of them are available) to 64% (if all three of them are

[62] There was also the fourth item in the questionnaire – the number of radios in home. However, this item behaved pathologically in the analysis: the more there were radios in home the less achievement. Hence it was not taken as an indicator for SES.

[63] The analysis is bound to the fact that the values were given by the students – they are, in many cases, credible. However, as with the home possessions (see figure 5.23), here also is a doubt that some of those students who marked all the possessions and accessories either did not understand the question or were just willing to fool with the questionnaire. When it comes to accessories, the effect is not noticeable (compare figures 5.23 and 5.24).

available) when the number of home accessories or home possession increases. Availability of all the stated facilities indicates the higher SES of the family. Correlation between the number of home accessories and achievement is $r = 0.28$ ($p < 0.001$) which is certainly positive and indicates moderate or high effect size ($d = 0.67$).

For the later use in SES, the cut-off for the factor was set on 2 accessories out of 3: if 2–3 items were met in the background questionnaire, the student was given 1 otherwise 0.

The dataset shows that the parents' educational level strongly predicts the children's future achievement level in English. Especially the achievement level is very low when the father or mother or both are illiterate. As many as 36.4% of the students had an illiterate mother and 16.4% had an illiterate father.

The dataset also suggests that either sound economic or intellectual capacity or both at home helps children to increase their English proficiency. If the father or mother or both are coming from an agricultural or related occupation, the students' achievement in English is significantly lower than with other occupational groups. It is reported that 54.9% of the mothers and 35.3% of the fathers worked in agriculture or only at home.

The dataset shows that when children have very few home possessions – zero to three out of the 11 – the achievement level is remarkably lower than the national average (< 47). With nine to ten possessions, the average score is very high (> 70) compared with the national average. The same is true for home accessories: When none or only one accessory indicator out of three is met, the results are lower than average (45–50%); and when two or more are met, the results are remarkably higher (58–64%). It is found that 2.3% students did not have any of the home possessions and 26.4% had no accessories.

SES and Achievement

The socio-economic status of the family was formed on the basis of seven indicators which were all first dichotomized. The variables (mothers' education, fathers' education, mothers' occupation, fathers' occupation, home possessions, home accessories, and type of school the students attend) were summed up (as SES) and changed into the percentage (P-SES). Deeper description of the transformations is seen in chapter 2. The P-SES represents the percentage of SES of the student's family, where 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 are positive) and 0 refers to the lowest possible SES (that is, all the seven indicators of SES are negative). The analysis of the P-SES by using Univariate GLM (that is, the Regression modelling) shows the strong relation between SES and achievement. Figure 5.24 presents the relationship between SES of the students and their achievement.

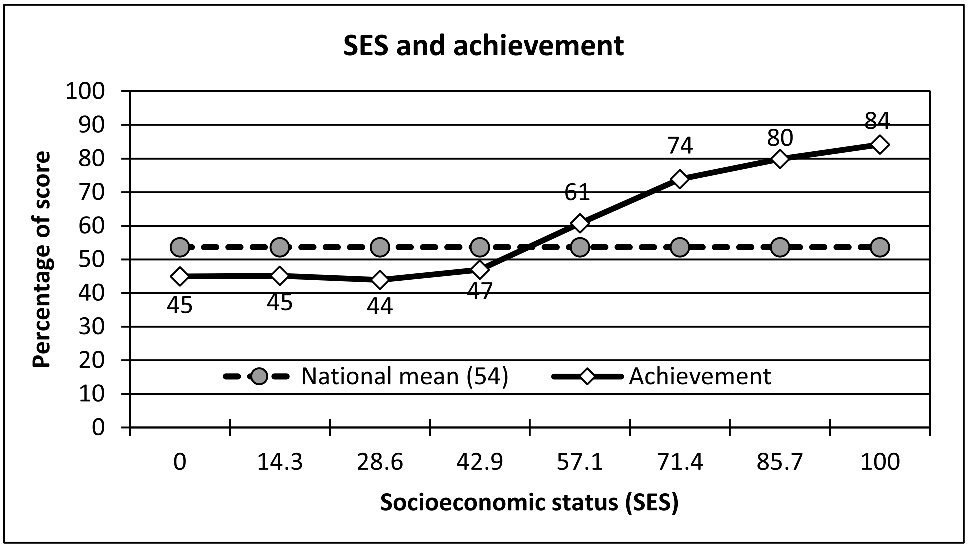

Figure 5.24 Relation between SES and achievement

Figure 5.24 shows a positive relationship between SES and the English achievement indicating the higher the SES, the higher the achievement. Pearson correlation between the variables is $r = 0.50$ which is a high value ($p < 0.001$) and indicates very high effect size ($d = 1.43$). The difference in achievement between the lowest SES groups (44% to 45%) and the highest one (84%) is remarkable. SES explains about 31% of the student variation ($\eta^2 = 0.31$) which is a very high percentage, but it is seen to be at the same level as was found in Nepali language in the grade 8 dataset (0.28, see ERO, 2013, p. 152). It is notable that in grade 5 Nepali dataset SES explains only 18% of the student variation ($\eta^2 = 0.18$) which, though indicates a high effect size ($f = 0.47$), is much lower than that in English. A simple explanation for this difference is that very high English proficiency is found in private schools which indicates higher SES and hence SES explains more English proficiency than Nepali proficiency.

It is worth noting that SES as a variable is more school related than student related factor. The correlation of SES and achievement is $r = 0.50$ in the student dataset and $r = 0.66$ in the school-wise dataset.[64] It is also worth noting that even though the SES is controlled in the student-wise dataset[65], there are still statistically significant differences between the community and institutional schools ($p < 0.001$). However, the effect size is reduced from $f = 0.81$ to $f = 61$, that is, from very high to high.[66]

From sociological point of view, it is interesting to know which of the individual indicators of SES are not met in those families where the children perform the lowest. Figure 5.26 illustrates the fact that in the families meeting less than four SES indicators, the challenge

64 Note that in grade 3 Mathematics (see chapter 3), the corresponding correlations are 0.35 and 0.46; the difference is not that wide as in the English dataset.

65 Because attending an institutional school is imbedded in the SES, the school type does not explain the achievement in ANCOVA when controlling the SES. For the ANCOVA, another SES – without considering the school type – was created.

66 Note that in the grade 3 Mathematics, the effect size is reduced from $f = 0.36$ to $f = 19$, that is, from high to moderate.

lies mainly in three factors marked in figure 5.25 with dotted line: both mothers' and fathers' education is low and the child does not attend institutional school (with English as the language of instruction). The last is difficult to change in the community schools but the low level of parents' education (especially the illiteracy) will be possible to address with an appropriate educational policy.

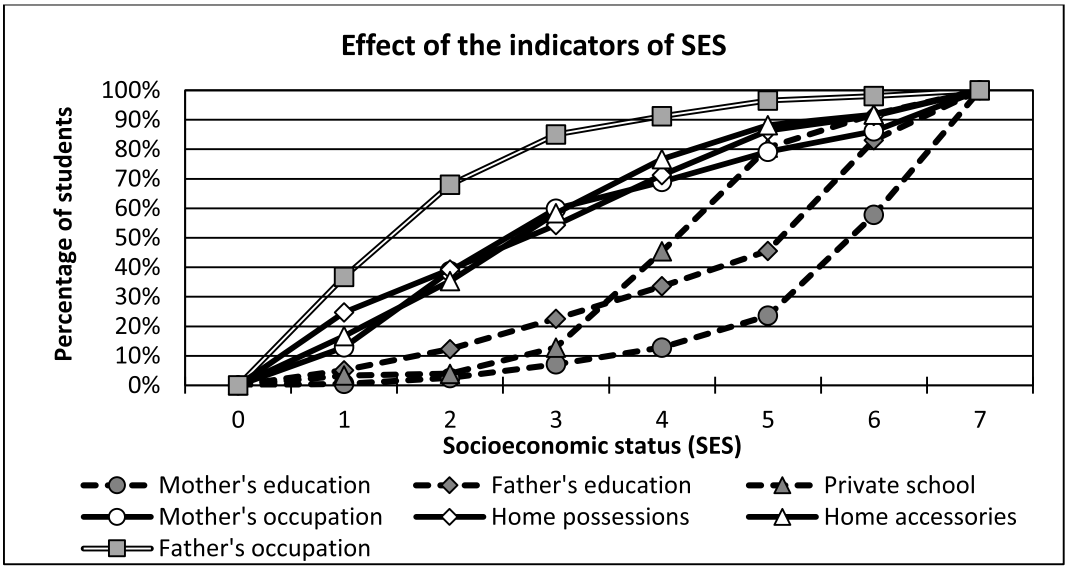

Figure 5.25 Effect of SES indicators on achievement

The dataset is evident that socio-economic status plays a strong role in the English achievement in Nepal. The difference between the lowest and highest SES groups is remarkable (40 percent). This means that if the SES of the lowest performing students is raised into a decent level, that is, in practice, that the problems in parents' low educational level is solved, the results in these groups also raises remarkably. Especially challenging is the situation in the families where the father or mother or both parents are illiterate or they both work in agriculture. It is also found that the use of English as a medium of instruction at school has determined practically the English language proficiency level, which happens more probably in institutional schools. The situation is not very easily changed to be equal between the community and institutional schools. Around 13.9% of the students are at the lowest level of SES.

Working beyond the school hours and achievement

Several questions were set in the student background questionnaire that were related to the students' activities outside the school. Two of them are briefly handled here: 1) Working after the school for a paid job, and 2) Participating in household work/chores. The values of the variables are divided into five categories: 0 (no time at all), 1 (less than 1 hour per day), 2 (1–2 hours per day), 3 (2–4 hours per day), and 4 (more than 4 hours per day).

The DTA indicates that, when it comes to the involvement in works after school, the cut-off is on whether the students work for a paid job or not. The DTA shows that when the children have no paid work at all, the results are above the national average in both community and institutional schools (fig. 5.26). If the students are working for a paid job – even less than one hour, the results are statistically significantly lower than the average.

The ANOVA shows that the relationship is firmly ($p < 0.001$) though slightly ($f = 0.22$) negative when students need to be engaged in paid job before and after school. It is notable, though, that most of the grade 5 children do not need to be engaged in paid job. Working after school implies that the family is poor and extra income is needed for subsistence for pocket money. It is obvious that when the student needs to work more than 4 hours per day, there is no time or energy to study or handle school homework. In institutional schools, the achievement of the children working for more than 4 hours per day (68, n = 80) is notably lower than that of those who do not need to work at all for paid job (81, n = 2,618) (see figure 5.26 institutional schools).

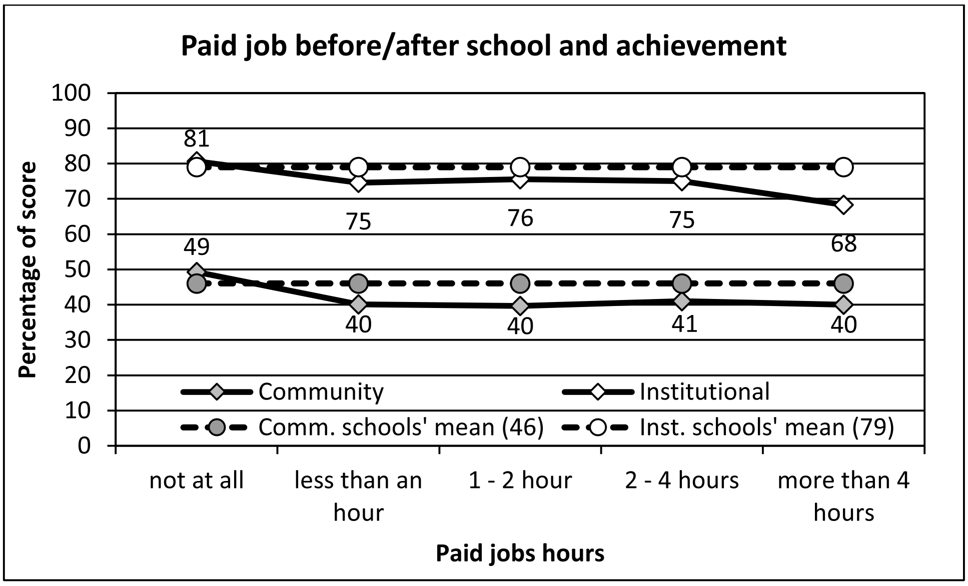

Figure 5.26 Relationship between achievements and paid job beyond school time

When it comes to the unpaid participation in the household work, it is usual and a supportive practice in families that the children take part in household chores at home which is a part of the socializing process of the children. The DTA shows that when the child spends some time (less than two hours) for the household chores, the results are statistically higher (57–58%) than those who spend not at all (52%) or more than 4 hours per day (50%). The effect of not participating in the household chores is seen to be more in the community schools than in the institutional schools (figure 5.27). Actually, it is seen that in the institutional schools, it does not make any difference whether the children work for two hours or less or do not work at all; the effect comes if children are spending four or more hours in chores. In community schools, the learning achievements are significantly lower for those not participating in the chores. Differences are significant ($p < 0.001$) though the effect size is small or moderate ($f = 0.16$ in community schools and $f = 0.12$ in institutional schools). It is somehow interesting that more than 10% of the students (n = 1,164) have reported that they spend more than 4 hours per day doing household work. In the rural area, it is somehow obligatory for them to support in cattle raising when the cattle are kept far away from home. It is self understood that, in these cases, there is not much energy to indulge in study and to exercise their school homework.

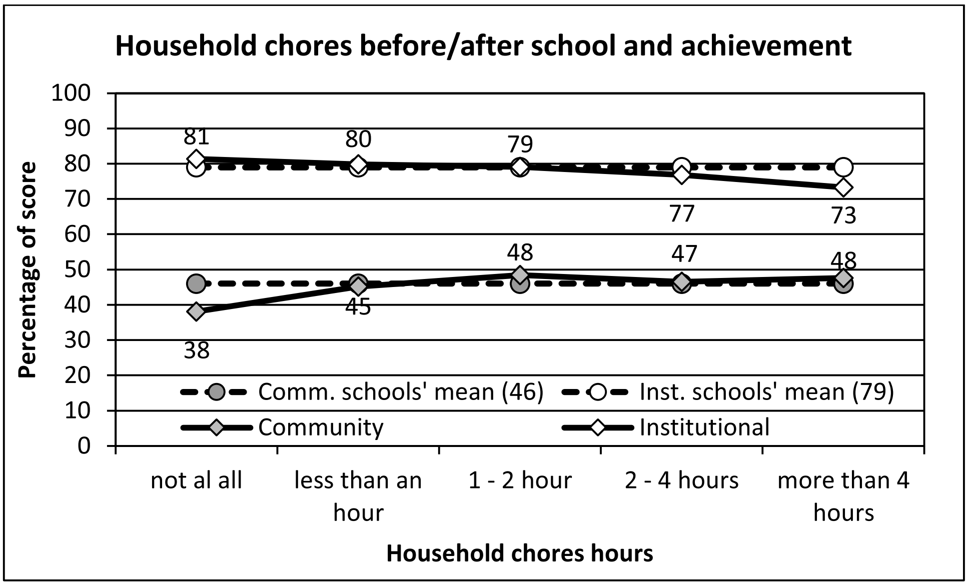

Figure 5.27 Relation between household work and achievement

The dataset reveals that either working for a paid job or for four hours per day in an unpaid household work outside school lowers down the school achievement of the student. However, a decent amount of household work up to two hours per day does not hamper the learning for the students in English. As many as 31.3% of the students worked in the paid capacity and 22.6% spent more than 2 hours in the unpaid household chores.

Attitude and achievement

In the context of assessment of the English language achievement, attitude tells us what the students think about English and its usefulness in their daily life and future. There is a more or less firm relationship between the attitude of the students and achievement. Though the relation is not always clear, the correlation between achievement and attitude towards the subject as well as self-efficacy in the subject is widely studied (see in Mathematics, for example, Metsämuuronen 2012a; 2012b; House & Telese, 2008; Shen & Tam, 2008; Kadijevich, 2006; 2008). Some researchers have noticed remarkable differences in correlation between countries (e.g., House & Telese, 2008; Kadijevich, 2006; 2008; Wilkins, 2004; Shen, 2002; Papanastasious, 2000; 2002; Stevenson, 1998). In some countries, the correlation between attitude and achievement has been found near zero, like in Macedonia (Kadijevich, 2008), in the Philippines (Wilkins, 2004), in Indonesia (Shen, 2002) and in Moldova (Shen, 2002), whereas in some other countries, it has been found as high as 0.60, e.g., in Korea (Shen, 2002). In NASA 2011, it was noticed that the grade 8 students were not consistent in the attitude test and the reliability of the international test stayed low (see ERO, 2013, table 2.11).

In NASA 2012, technically speaking, the same shortened version of Fennema–Sherman Attitude Scales (Fennema & Sherman, 1976), as is used in several international comparisons like in TIMSS and PISA studies, was used. The original scales included nine dimensions but in these international comparisons only three are used with four items on each dimension and two negative items on each of the first two dimensions (see in detail in chapter 2). The names of the original factors were "Liking English", "Self-Efficacy in

English”, and “Experiencing utility in English” (compare naming in, e.g., Kadijevich, 2006; 2008). Because of students’ inconsistent manner in answering the attitude scale in NASA 2011, only the dimension of “Experiencing utility in English” was taken into the measurement instrument of grade 5 students. Reliability of the score of five items is sufficient (α = 0.74). The relation between the attitudes (divided into seven groups with somehow an equal number of the students, that is, septiles[67]) and achievement score is shown in figure 5.28.

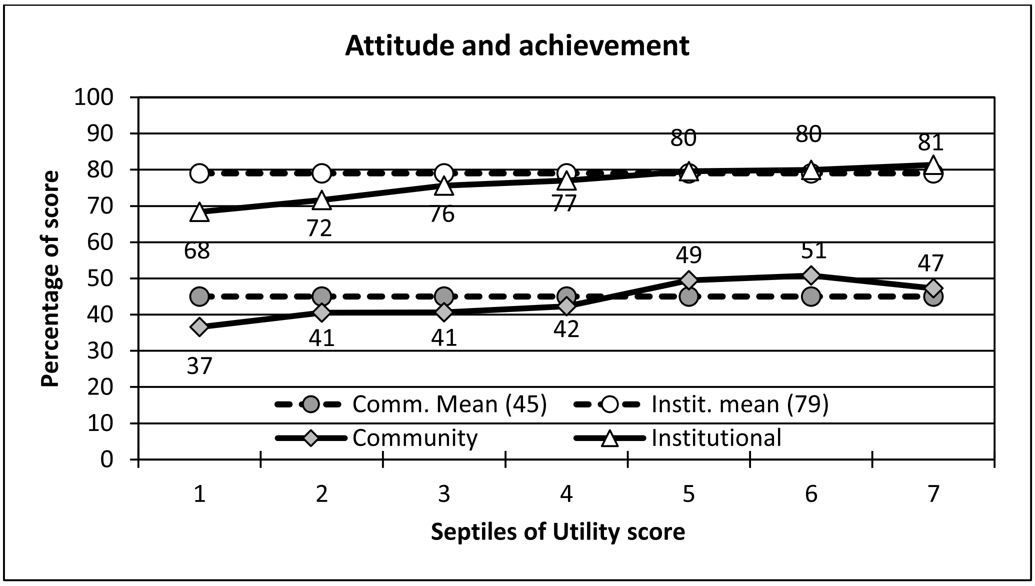

Figure 5.28 Relation between attitude and achievement by school type

There is a clear positive correlation between attitude and English achievement in the whole dataset (r = 0.27). The connection is moderately high (d = 0.65) indicating the fact that difference between the means of the lowest attitude group (40%) and highest one (59%) is remarkable. In the whole dataset, the division of attitudes to seven groups explains the achievement level somehow 9% (η^2 = 0.087). Connection is higher in the institutional than community schools. The reason is that, in community schools, the highest attitude group has not responded the question logically; it is possible that, within the highest attitude group, there are many students who have either fooled in the test or did not understand the questions (see the same kind of phenomenon in the SES analysis above). The difference between the lowest and highest attitude group is 10 percent in community schools (f = 0.24) and 13 percent in institutional schools (f = 0.27).

The connection of the sense of utility in English and achievement is clear though it is not known whether the positive attitude is a consequence of high achievement or the other way round. From the statistical point of view, on the basis of simple ANOVA GLM procedure, attitude explains the achievement 8.7% while achievement explains attitude 7.7%. Hence,

[67] The original score is short (maximum was 15 points) and quite many students (36%) gave the maximum score. Hence it was not possible to form more precise classification such as deciles. Seven classes (septiles) was the most precise alternative with the given dataset.

it will be more probable that, in Nepal in grade 5 English subject, the better achievement is a consequence of more positive attitude than other way round.

The dataset shows that positive attitude towards the subject correlate with positive achievement in English. The better achievement is more probable a consequence of more positive attitude rather than the other way round.

Age and student achievement

In the Nepalese context, the age of the students attending to grade 5 varies widely. Some students have reported their age as below nine years and some above 16. All the ages of the students below 10 were encoded as 'up to 9 years', and all students above 14 were encoded as '15 years or above'. The descriptive statistics of the mean in each year are given in tables 5.22 and 5.23 and depicted in figure 5.29.

Table 5.22 Descriptive statistics of the students' achievement in different age groups

Age	N	Mean	SD	CV
Up to 9 years	372	45.3	22.3	49.3
10 years	1,796	52.1	24.3	46.7
11 years	3,589	58.0	24.9	42.9
12 years	4,431	56.3	24.1	42.8
13 years	1,804	49.9	23.0	46.1
14 years	750	45.1	20.6	45.7
15 years or above	423	40.6	21.9	54.0
Total	**13,165**	**53.9**	**24.3**	**45.2**

Table 5.23 Students' achievement in different age groups by the type of school

Age	Community				Institutional			
	N	Mean	SD	CV	N	Mean	SD	CV
Up to 9 years	348	43.0	21.0	48.9	24	78.7	11.0	14.0
10 years	1,434	45.3	21.4	47.3	359	79.8	13.1	16.4
11 years	2,357	45.8	20.9	45.7	1,231	81.2	11.7	14.4
12 years	2,946	45.3	20.2	44.6	1,477	78.6	13.5	17.1
13 years	1,419	43.0	19.8	45.9	384	75.1	14.9	19.8
14 years	658	41.6	18.5	44.4	92	70.5	17.3	24.5
15 years or above	382	37.7	19.8	52.7	41	68.4	21.3	31.1
Total	**9,544**	**44.4**	**20.5**	**46.1**	**3,608**	**78.9**	**13.5**	**17.1**

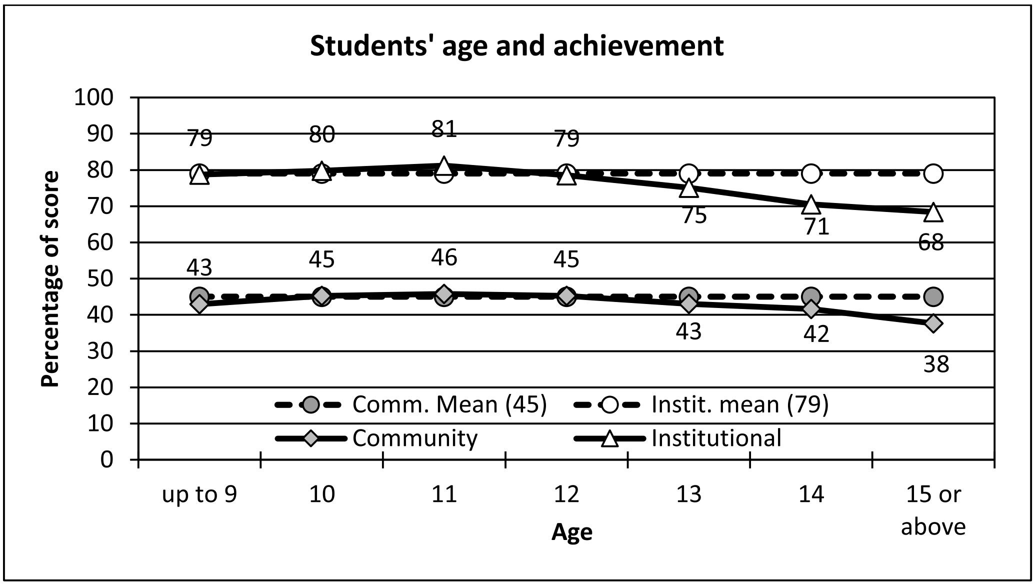

Figure 5.29 Relation between age and achievement in English

It is evident that the best achievers are those students who are at their proper age for grade 5 studies (10 to 12 years old), scoring 45–46% in community schools and 79–81% in institutional schools. The higher the age – meaning that the students have either started their schooling much later than they should have, or they have repeated the classes early or the same many years – the lower the results. The achievement level is remarkably lower than the average when the students are at the age 14 or above (38–42% in community schools and 68–71% in institutional schools). Correlation between the age and achievement in institutional schools is $r = –0.16$ ($p < 0.001$) – indicating moderate effect size ($d = 0.35$), whereas in community schools the correlation is low ($r = –0.06$, $d = 0.13$). The simple ANOVA with GLM procedure indicates that the age (that is, the delayed studies) explains more probably the lower achievement level (3.5%) than the achievement level due to the prolonged studies (1.1%). Another side of the matter is that it is good that these “over-aged” students are retained at school to learn though they should have been identified at a much earlier age for extra tuition or support.

Dataset suggests that the highest performance in English is found among the students studying at their proper age years, that is, at the age of 10 to 12 years. Otherwise, the achievement lowers down as the age increases. Around 25.4% of the students fell aside of 10–12 years.

Support for the study and student achievement

The relation between the support for studies and achievement was analysed based on the following question: "Who helps you when you do not understand what you have read?”. In the question, only one option was selected, though in many cases, there might be several helpers, which has not been detected now. The descriptive statistics of the supports received are given in tables 5.24 and 5.25.

Table 5.24 Descriptive statistics of supports for students and achievement level

Support received	N	Mean	SD	CV
Tuition	1,053	60.5	24.2	40.0
Mother	790	56.8	25.6	45.1
Brother/Sister	5,740	56.2	23.5	41.8
No one	293	52.7	25.2	47.8
Father	2,696	52.1	24.5	47.1
Teacher	2,223	48.5	23.6	48.7
Total	**13,182**	**53.9**	**24.3**	**45.2**

Table 5.25 Support to the student by the type of school

Support received	Community				Institutional			
	N	Mean	SD	CV	N	Mean	SD	CV
Brother/Sister	4,102	47.2	20.2	42.9	1,626	79.3	12.9	16.3
Tution	597	45.7	20.5	44.8	456	79.9	11.8	14.8
Mother	518	44.0	20.8	47.4	272	81.3	12.9	15.8
Father	2,070	43.8	20.9	47.6	626	79.4	13.4	16.8
No one	220	43.4	21.4	49.3	73	80.9	10.6	13.1
Teacher	1,696	40.2	19.1	47.5	526	75.3	15.5	20.5
Total	**9,545**	**44.4**	**20.5**	**46.1**	**3,624**	**78.9**	**13.5**	**17.1**

It is seen that an external support is, in many cases, necessary for the students to achieve better than average marks on the test. However, the reality is seen to be different in the community schools compared with the institutional schools. In the whole dataset, there is about 8 percent difference between those who do not receive any kind of support (53%) and those who receive (private) tuition (60%). It is possible that the children receiving the private tuition also spend more time on the homework which explains the high score. Mother's as well as brother's or sister's support at home raises the achievement level above the average (57–58%). Those who received support from their father or teacher obtained notably lower than the average – even lower than those receiving no tuition at all.

The students in the community schools are seen to have received most effective support from their brothers or sisters (47%) or (private) tuition (46%). In institutional schools, on contrary, the highest results are among those who receive support from mother (81%) or they have studied just by themselves (81%). The effect of the support received is, in any case, very low, where the effect size is $f = 0.15$ in community schools and $f = 0.14$ in institutional schools –indicating that the difference in mean is not notable.

The dataset reveals that the support received from mother, brother and sister raises the achievement level more than the support received from father or teacher. In the whole sample, the highest achieving group is the one who receives private tuition. However, the difference between the highest and lowest performing groups is not notable. It is possible that the group receiving private tuition also spends more time on their homework, which explains the higher score.

Availability of textbook and student achievement

The data shows that there are some students who do not have the proper textbook up to the end of academic year. Table 5.26 shows the descriptive statistics of availability of the English textbook and the achievement.

Table 5.26 Availability of textbook and the average achievement

Do you have a textbook of English?	N	Mean	SD	CV
Yes	12,215	54.8	24.1	44.0
No	532	46.4	23.5	50.6
Total	**12,747**	**54.5**	**24.2**	**44.4**

Out of 12,747 students who responded to the question, 4.3% (4.5% in community schools and 4.0% in institutional schools) did not have a textbook available at school even by the end of the academic year. The relation between the availability of textbook and achievement is significant ($p < 0.001$) though the effect size in the whole dataset is small ($f = 0.07$). The difference in achievement is 8 percent (7.6 in community schools and 6.7 in institutional schools).

The dataset shows that 4.3% students lack the proper textbook in English. The achievement level of these students is significantly lower than the level of those students those who have access to the textbook.

Homework assigned/checked and achievement

Homework is one way to enhance learning which can be used as drill, exercise, and as an evaluation tool. When homework is systematically assigned and checked, it tends to boost the level of achievement. Now, the related results in NASA 2012 are based on students' reports in which slightly deviant responses have also been found that the same classroom, some students have reported to have received and got it checked whereas others have not got it. However, in the wide scope, the results do make sense. Data on homework assigned and checked is presented in tables 5.27 and 5.28 and depicted in figure 5.30.

Table 5.27 Homework given and checked and the achievement

Status of homework	N	Mean	SD	CV
Given Everyday-Checked Someday	1,019	58.6	24.9	42.5
Given Someday-Checked Eachtime	1,824	55.2	24.0	43.4
Given Everyday-Checked Everyday	945	54.3	25.9	47.6
Given Someday-Checked Someday	8,577	54.3	23.9	44.0
Given Someday-Not checked	129	46.9	23.9	51.0
Not given	44	44.3	23.5	53.1
Given Everyday - Not checked	204	40.8	21.9	53.6
Total	**12,742**	**54.4**	**24.2**	**44.5**

Table 5.28 Homework given and checked by the type of school

Status of homework	Community				Institutional			
	N	Mean	SD	CV	N	Mean	SD	CV
Given Everyday-Checked Everyday	610	40.0	19.7	49.1	335	80.3	11.6	14.4
Given Someday-Checked Someday	6,510	46.2	20.5	44.4	2,056	79.9	13.3	16.6
Given Someday-Not checked	106	40.2	20.1	50.2	23	78.0	12.7	16.2
Given Every day-Checked Someday	562	42.8	20.8	48.6	457	78.0	13.1	16.8
Given Some day-Checked Everyday	1,172	43.2	19.6	45.3	652	76.7	13.9	18.2
Given Everyday - Not checked	167	33.7	16.1	47.6	37	72.9	14.9	20.4
Not given	31	33.3	15.7	47.1	13	70.7	17.4	24.7
Total	**9,158**	**44.9**	**20.4**	**45.5**	**3,573**	**79.0**	**13.3**	**16.9**

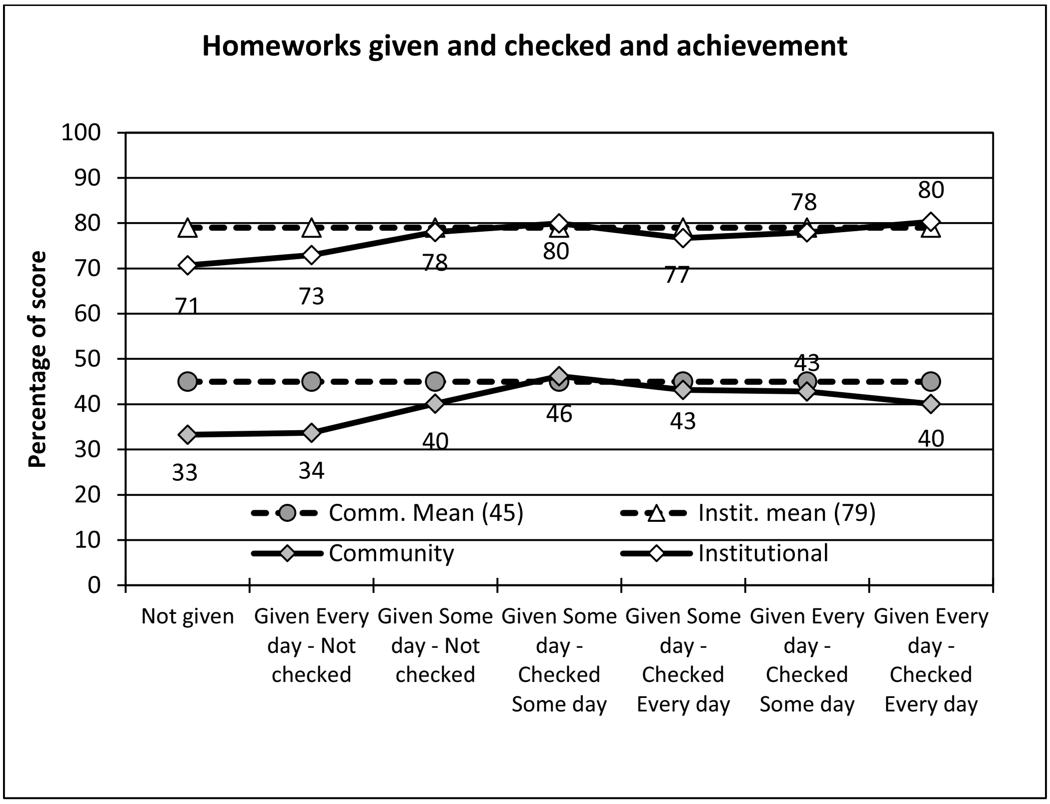

Figure 5.30 Relation between homework given/checked and achievement

It is evident that the students achievement is notably lower (33–34% in community schools and 71–73% in institutional schools) for those who report that the teachers does not assign and check homework regularly compared for those who report that the teacher assigns and checks them – even if not regularly (46% in community schools and 80% in institutional schools). The differences are statistically significant ($p < 0.001$). However, those groups not receiving homework assignments or not getting it checked are very small and hence, the effect size is small ($f = 0.13$ for community schools and $f = 0.12$ for institutional schools). Such grouping explains only 1–2% of the variance in the data ($\eta^2 = 0.016$ for community schools and $\eta^2 = 0.014$ for institutional schools).

The dataset strongly suggests that if the teacher assigns and checks the homework systematically, the achievement is higher than teaching without assigning homework and checking it. Regular assignment of homework with its checking, though not every day, is found to have raised the scores up to 13 percent. Around 3.0% of the students either did not get homework or did not get it checked.

Future aspiration of the student and achievement

The future aspirations or goal setting of the students can encourage in studies or, in some cases, when knowing that the future plan does not require long studies, the motivation for hard work in school tends to decline. The students' future aspirations were asked in eight categories. Those were (1) farming, (2) business, (3) teaching, (4) government officer, (5) going abroad, (6) engineer, (7) doctor, and (8) other. Future plan of the students is seen to have been connected strictly with the student achievement, which can be seen in tables 5.29 and 5.30 and figure 5.31.

Table 5.29 Students' future aspiration and achievement

Future aspiration	N	Mean	SD	CV
Farmer	458	38.2	21.5	56.4
Bussines	515	45.6	23.0	50.4
Teacher	3,614	46.7	21.6	46.2
Government Officer	1,113	49.5	23.5	47.4
Engineer	2,086	56.9	23.8	41.9
Going Aboard	879	58.2	24.7	42.4
Doctor	3,747	60.9	23.3	38.3
Other	493	75.7	20.1	26.6
Total	**13,182**	**53.9**	**24.3**	**45.2**

Table 5.30 Students' aspiration and achievement by the type of school

Future aspiration	Community				Institutional			
	N	Mean	SD	CV	N	Mean	SD	CV
Farmer	427	36.1	20.2	55.9	31	66.4	19.7	29.7
Bussines	428	40.1	19.7	49.3	87	72.8	17.9	24.6
Going Aboard	515	43.2	19.8	45.8	364	79.4	12.2	15.3
Government Officer	898	43.5	20.9	48.1	215	74.5	15.8	21.2
Teacher	3,229	43.5	19.8	45.4	378	73.9	16.2	21.9
Engineer	1,437	46.9	20.7	44.1	649	79.1	12.7	16.0
Doctor	2,252	48.3	20.0	41.4	1,490	80.1	12.1	15.1
Other	131	51.4	22.9	44.5	362	84.5	8.5	10.0
Total	**9,545**	**44,4**	**20,5**	**46,1**	**3,624**	**78,9**	**13,5**	**17,1**

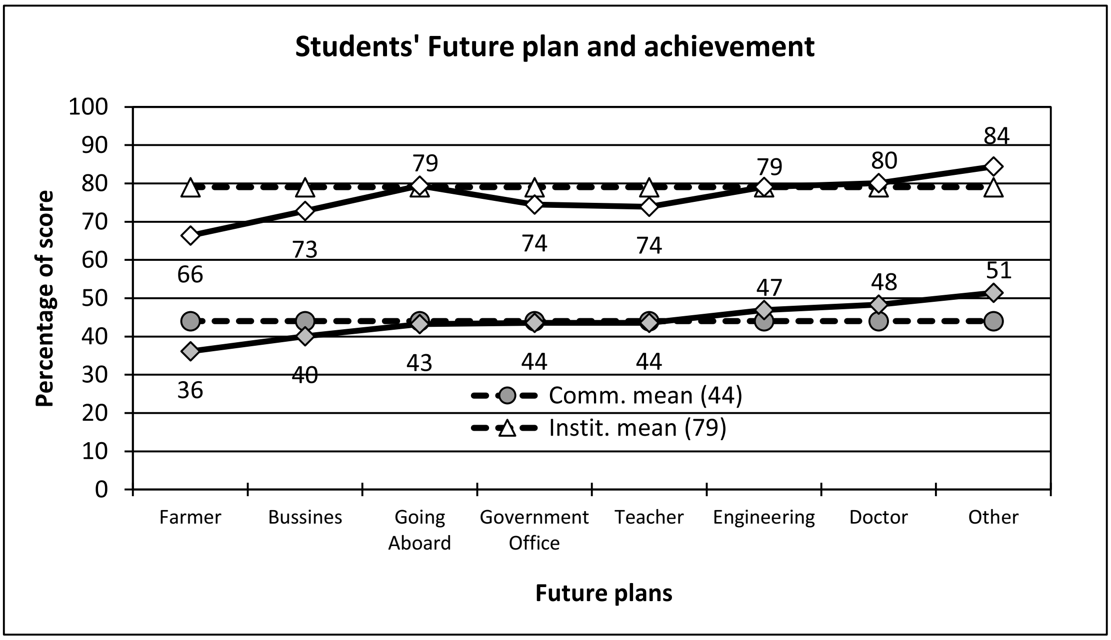

Figure 5.31 Relation between students' aspiration and achievement

Some of the occupations, like Engineers, Medical Doctors, and Teachers, are favoured in the society, being associated with probable guaranted and higher economic prospects. It is found that even the weakest students in the dataset (scoring less than 20%) are aspiring to these occupations. Of these low-level students (*n* = 1,325), 12% are aspiring to be Engineer, 16% to be doctor, and 33% to be teacher. On the basis of their achievement level, this dream, most probably, will turn to reality for a very few of them. Another reality is that, according to the fathers' occupation, 35% students come from the agricultural background but, in the whole dataset, only 3.5% students are aspiring to follow agricultural occupation–farming. When the student knows that s/he will be continuing the family occupation–farming, the learning outcomes in English are remarkably lower than the average (36% in community school and 66% in institutional school).

Because the professionals such as Engineers, Doctors and Teacher are demanded much in the job market, the competition for the study places is also very tough. Hence, the higher the goals the higher should be the achievement level in order to make the dream of the future profession come true. From this point of view, the students' future plan is seen logical while comparing it to the mean achievement level. Students who are aspiring to be Engineer (56%) or Doctor (61%) score remarkably higher than those aiming to be in agriculture (38%) or Business (46%). In the whole dataset, those students aspiring to be going abroad also score remarkably high (58%) which is higher than the score of those who are aspiring to be Engineer. The future aspiration explains 11% of the achievement level ($\eta^2 = 0.115$); the effect size is high ($f = 0.36$) indicating that the difference between the lowest and highest group is remarkable.

The dataset shows a strong connection between the future aspiration of students and their achievement. As the student aspires to hold the professional career of a doctor, an engineer, or going abroad, their achievement in English is higher than the average. The number of students aspiring to be a teacher, an engineer, or a doctor is remarkably high (72%).

Activities in the school and student achievement

Various activities of the students and teachers determine the learning environment of the school. Bullying, for example, is one of the hindering activities of the students in the school that tend to affect learning. In the student background information questionnaire, several student and school related activities were asked, some of which are positive and some are negative. Here, bullying is handled as one of the negative indicators and students' positive impressions towards schools' and teachers' activities are taken as the examples of positive indicators.

Negative activities - Bullying

Bullying is one of the problems in schools that worsens the learning environment for the students. International Studies like TIMSS and PISA emphasize specifically to study such phenomena which are seen in their background questionnaires. In the background questionnaire for students in NASA 2012, five questions consisted the varieties of bullying that tend to happen in the school. All the questions were stemmed by the phrase *"Which of the following activities happened in your school in the last month?"* The students' responses are presented in tables 5.31 and 5.32 and depicted in figure 5.32. 'No (%)' indicates the percentage of the students' response which tells that no such activity happened in the school and 'Yes (%)' indicates the percentage of the students who reported that the particular type of bullying happened in a month. About 26% of the students mentioned that, during a month, something of their own was stolen which is an alarming sign of the system.

Table 5.31 Frequencies of encountered bullying

Type of Bullying	No (%)	Yes (%)
I was made fun of or called by names	71.9	28.1
Something of mine was stolen	74.0	26.0
I was hit or hurt by other student(s)	79.2	20.8
Fellow students kept outside without involving me in activities	79.4	20.6
I was made to do things I didn't want to do by other students	84.9	15.1

Table 5.32 Bullying and the achievement by the type of school

Intensity of bullying	Community				Institutional			
	N	Mean	SD	CV	N	Mean	SD	CV
No bullying	4,093	48.1	20.6	42.9	1,639	80.1	13.0	16.2
20% bullying	2,298	44.4	19.7	44.3	876	79.3	12.6	15.9
40% bullying	1,498	42.0	19.6	46.8	556	78.1	13.6	17.4
60% bullying	766	39.7	19.5	49.0	374	77.4	13.6	17.6
80% bullying	288	39.1	18.9	48.4	94	75.5	14.5	19.2
100% bullying	270	33.8	19.5	57.5	26	67.8	20.8	30.7
Total	**9,213**	**44.8**	**20.4**	**45.5**	**3,565**	**79.1**	**13.2**	**16.7**

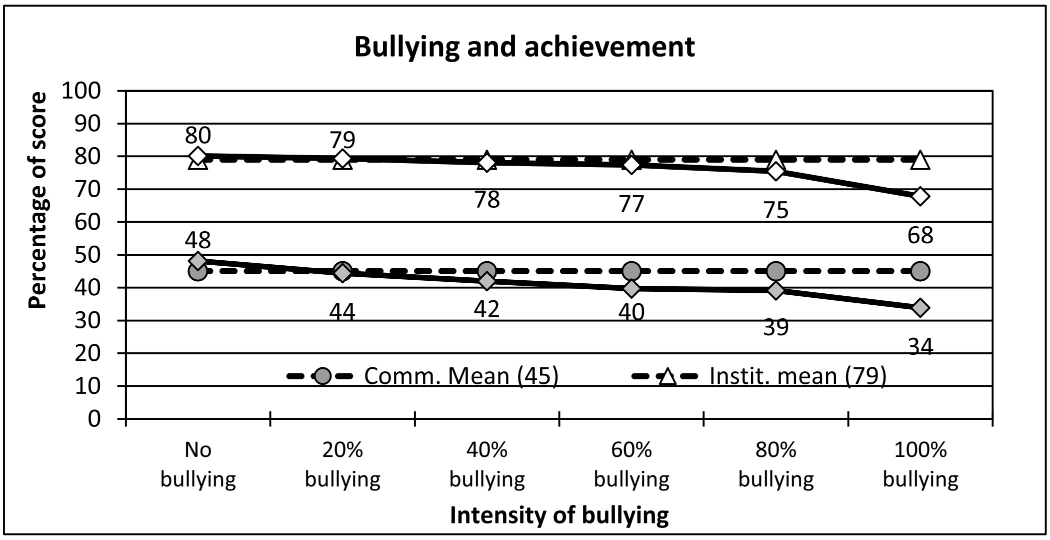

Figure 5.32 Relation between bullying and achievement

The sum of all five items is taken as an indicator of bullying. Figure 5.32 shows the achievement of the students in each category of bullying. If only one activity of bullying is reported, it is categorized as 20% bullying, and if all five activities are reported it is categorized as 100% bullying. Knowing that 44–46% of the students did not encounter any bullying during a month, one can infer that the remaining 54–56% did encounter at least one type of bullying. This is a remarkable number of students. As many as 5.3% students – 6.1% in community schools and 3.4% in institutional schools – are experiencing a severe kind of bullying (the sum of 80% and 100% bullying). This means, in practice, that more than 38,000 grade 5 students in Nepal are encountering physical, psychological, and social bullying in every month.[68] The number is too much even if it would not have any effect in learning outcomes. However, it is found that learning outcomes are notably lower than the average with the 14% students who have encountered more than two different types of bullying (34–40% in community schools and 68–77% in institutional schools). Students who do not experience bullying and those who encountered extreme bullying of four or five kinds have, at most 14 percent points' achievement gap, though a small number of students reported this kind of bullying (n = 678). The difference is statistically significant (p = 0.001) though the effect size is small or medium (f = 0.18 in community schools and f = 0.11 in institutional schools). Though extreme cases of severe bullying are rare, bullying tends to be quite common in schools. This negative phenomenon causes needless harm to young children and needs to be rooted out from schools.

Positive activities in school

The activities that can boost the learning and achievement of the students are categorized as positive activities. The students were asked about such positive activities that happen in the school in two sets of questions collected in table 5.32. The table shows the responses of the students in all four categories which are in the 4-point rating scale anchored to fully disagree (0) and fully agree (3). Generally speaking, the 5th graders express their satisfaction with the school and student related activities in school. However, remarkably high number of students (9%) expressed that they feel that the teacher is not treating them fairly. The same phenomenon was seen also in 2011 datasets with grade 8 students: 11% students in Mathematics , 12% in Nepali, and 13% in Social Studies felt unfair behaviour of teachers (see, ERO, 2013).

Table 5.33 Students' response towards teacher and school-related activities

Teacher and Students activities[1]	Respondents in % (valid %)			
	Fully agree	**Partially agree**	**Partially disagree**	**Fully disagree**
q28a: I like to come and stay in school	91.4	5.4	1.1	2.0
q27b:Most teachers are interested in student's well-being	86.9	8.1	2.4	2.6
q27a: Students get along well with most teachers	86.0	10.3	1.4	2.2
q28c: Teachers in the school care about the students	86.0	9.3	2.2	2.5
q27d: If I need extra help, I will receive it from my teacher	85.2	9.8	2.3	2.8
q27c: Most of the teachers really listen to what I have to say	81.7	11.9	3.1	3.3
q28b: Students in my school like me	79.5	15.7	2.6	2.2
q27e: Most of my teachers treat me fairly	70.6	15.6	5.0	8.8
Average	**83.4**	**10.8**	**2.5**	**3.3**

1)The activities are ordered on the basis of percentage in "Fully agree"

Further analysis is carried out by recoding the variables into two categories (2–3 = 1, that is, "agree" and 0–1 = 0, that is, "disagree"). Furthermore, the sum of eight indicators is converted into the percentage to analyse the level of positive activities and its relation to achievement.

[68] According to the "Primary level total enrollment in all types of schools by district, Flash I_2012–2013", there were 731,573 grade 5 students. 5.3% of these is 38,777 students.

DTA finds four attitude groups in the indicator. These boundaries and descriptive statistics are seen in tables 5.34 and 5.35 and illustrated in figure 5.33. 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 positive activities and achievement is positive (r = 0.22), statistically significant (p < 0.001) and moderately high (d = 0.50).

Table 5.34 Teacher- and school-related activities and the achievement

Percentage of positive actions	N	Mean	SD	CV
62.5% or lower	1141	39	20.6	54.9
75.0%	784	44	20.1	50.5
87.5%	1949	48	18.6	45.1
100%	9063	50.5	18.8	41.7
Total[1]	13794	53.6	24.2	45.2

1) Total includes also the cases without giving their opinion (missing n = 857)

Table 5.35 Teacher and school-related activities and the achievement by the type of school

Percentage of positive actions	Community				Institutional			
	N	Mean	SD	CV	N	Mean	SD	CV
62.5% or lower	1,012	34.5	18.2	52.7	129	66.8	18.9	28.3
75.0%	622	38.1	18.6	48.7	161	73.8	14.7	19.9
87.5%	1,278	42.1	19.5	46.2	671	78.7	13.1	16.6
100%	6,433	47.4	20.4	43.0	2,618	80.0	12.6	15.8
Total[1]	**10,107**	**44.4**	**20.4**	**45.9**	**3,674**	**79.0**	**13.4**	**17.0**

1) Total includes also the cases without giving their opinion (missing n = 857)

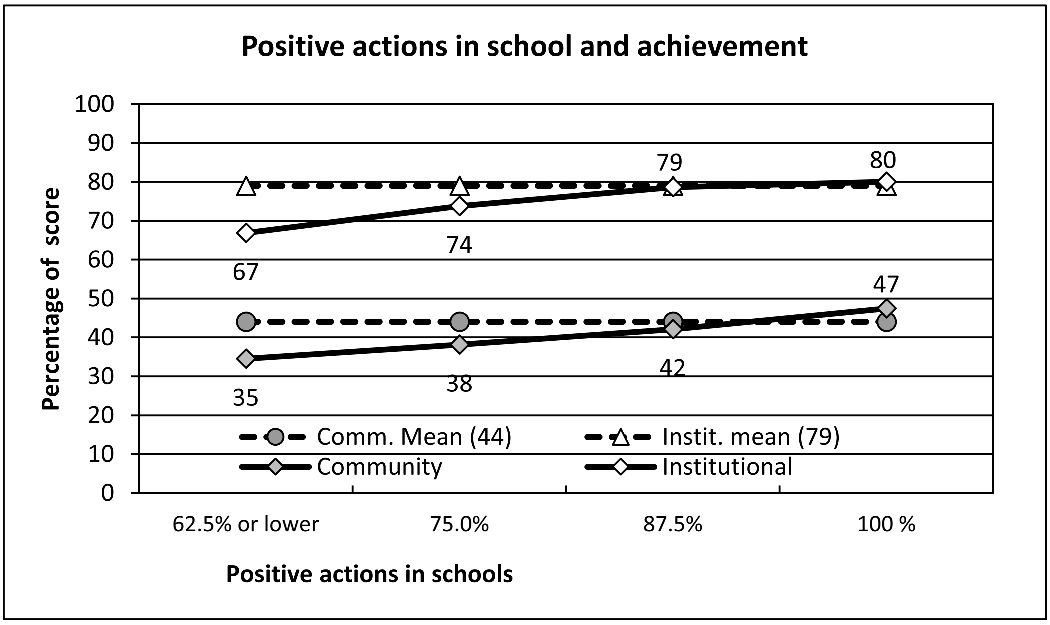

Figure 5.33 Relation between positive actions in school and achievement

The data shows that there is a positive relation between the students' feeling of the teachers and schools related activities and achievement. The increase in achievement is directly proportional to the increase in the intensity of such activities. After dividing the indicator into four groups on the basis of DTA, the differences between the groups are statistically significant (p < 0.001), however, the effect size is moderate (f = 0.22 in community schools and f = 0.21 in institutional schools) and the difference between the most positive group and the most negative group is notable (12–13 percent). Only when the students are extremely positive towards school and teachers' behaviour, the learning achievement is higher than the average. Students with a negative feeling on five or more of the eight indicators (62.5% of the total) are in great danger of achieving much lower than the average in English.

The dataset indicates that an alarmingly high number of the students (54–56%) have encountered bullying in school in a month of which 5.3% of students are experiencing a severe kind of bullying. The phenomenon tends to be affecting the learning outcomes in almost all the groups of the students who felt bullying that urge for all possible efforts to be put to root the phenomenon out from schools.

The dataset also hints that, when the students feel that the actions of the teachers and the schools are ultimately good, the English results are better than average (47% in community schools and 80% in institutional schools). At the other extreme of feeling that such actions are ultimately negative, the results are far below the average (35% in community schools and 67% in institutional schools). Almost 13.8% of the students feel that their teacher does not treat them fairly.

5.2 Synthesis of the Analysis

Above, several individual student and geographically related factors have been detected which individually explain the difference in achievement between the students. These factors are collected in table 5.36. It is notable that, except for gender, all the factors showed statistically significant difference between the groups when analysed individually.

Table 5.36 Individual variables handled within the text and their effect in one-way ANOVA

Variable and values[1]	Leverage[2]	Eta squared[3]	Effect size[4]
Ecological zone (1 = Mountain, 2 = Hill, 3 = Tarai, 4 = Valley)	+32.53	0.233	0.55
Development region(1=Eastern … 5=Far-Western, 6=Valley)	+40.21	0.289	0.64
School location (0=Rural, 1=Urban)	+25.10	0.191	0.49
School type (0=Community,1= Institutional)	+34.58	0.399	0.81
Gender (0 = girls, 1 = boys)	+1.80	0.001	0.03
Caste (1= Janjati, 2= Dalit,3=Madhesi,4=Brahman, 5 = Chhetri)	+19.01	0.051	0.23
Language at home (1 = Nepali, …, 12 = Other)	+18.15	0.081	0.30
Mother's Education (1 = Illiterate, …, 8 = Above MA)	+37.04	0.096	0.33
Father's Education (1 = Illiterate, …, 8 = Above MA)	+34.88	0.110	0.35
Mother's Occupation (1=working abroad,.,8=working at other home)	+26.27	0.085	0.30
Father's Occupation (1=working abroad,., 8=working at other home)	+26.10	0.166	0.45
Home possessions (sum; max 11)	+33.05	0.110	0.35
Home accessories (sum; max 3)	+18.36	0.081	0.30
SES (sum max 7)	+39.24	0.311	0.67
I do jobs at home (=not at all, …, 4= more than 4 hours)	+8.02	0.013	0.11
I work at a paid job (1 = not at all, …, 4 = more than 4 hours)	+16.85	0.079	0.29
Attitude utility in mathematics (sum max 15)	+26.42	0.091	0.32
Age	+12.81	0.036	0.19
Who helps you …? (1 = Father, …, 6 = Teacher	+12.05	0.031	0.18
Do you have textbook of English subject (0 = no, 1 = Yes)	+8.39	0.005	0.07
Homework (0=not given, …, 6= Given every day,checked every day)	+17.79	0.009	0.10
Bullying (sum; max 5)	+20.43	0.023	0.15
Positive Activities in school (sum; max 8)	+23.85	0.055	0.24

1) *The order of the variables is the same as handled in the Sections above*
2) *Difference between the lowest and highest group-mean*
3) *On the basis of one-way ANOVA 4)Cohen's **f***

On the basis of univariate ANOVA, school type, followed by the socio-economic status and development region, are seen to be the most effective single factors in affecting the achievement level of the student, where the effect sizes are $f = 0.81$, $f = 0.67$, and $f = 0.64$ respectively. Some of these variables in table 5.35 are seen to be strongly related to each other and hence not adding value in explaining why some students are performing much better than others. In what follows, the synthesis of the analysis is presented in two ways: all the variables are presented as a result of Multilevel Modelling in table 5.36 and in table 5.37 the statistically best factors are collected by using the Regression modelling. For the analysis, grouping factors are changed to be so called Dummy variables when needed; for example, the Ecological zone is transformed into three variables: variables indicative for Mountain, for Hill, and for Tarai.

Modelling the overall achievement by Multilevel Modelling

The data sets collected from schools are always clustered, that is, the students in the school are more alike to each other in comparison with the case that the same number of students would have been sampled totally from the population. Multilevel modelling is used to acquire the correct test values while taking into account the clustering effect of the school.

Table 5.37 shows the corrected estimates for the variables and in table 5.38 the hidden commonalities of the factors are revealed while modelling the phenomenon in a multivariate manner, by using the multivariate ANOVA .

Table 5.37 Individual variables and their effect in multilevel analysis

Source[1]	df_1	df_2	F	Sig.
Intercept	1	1268.6	267.4	<0.001
Ecol zone Mountain Dummy (Mountain = 1, other =0)	1	522.6	6.04	0,014
Ecol zone Tarai Dummy (Tarai = 1, other =0)	1	516.9	0.07	0.786
Dev region Eastern Dummy (Eastern = 1, other =0)	1	518.0	11.45	0.001
Dev region Western Dummy (Western = 1, other =0)	1	496.1	0.49	0.484
Dev region Mid-Western Dummy (Mid-Western = 1, other =0)	1	496.2	7.84	0.005
Dev region Far-Western Dummy (Far-Western = 1, other =0)	1	509.1	0.33	0.566
Dev region Valley Dummy (Valley = 1, other =0)	1	483.4	22.47	<0.001
School location (0 = Rural, 1 = Urban)	1	478.7	2.92	0.088
School type (0 = Community, 1 = Institutional)	1	487.0	172.4	<0.001
Gender (0 = girls, 1 = boys)	1	7935.8	0.61	0.435
Caste Brahman& Chhetri Dummy (Br. & Ch. = 1, other =0)	1	8035.3	15.70	<0.001
Caste Janjati Dummy (Janjati = 1, other =0)	1	8050.7	9.99	0,002
Caste Madhesi Dummy (Madhesi = 1, other =0)	1	8098.5	15.69	<0.001
Caste Dalit Dummy (Dalit = 1, other =0)	1	8028.6	3.71	0.054
Caste Other Dummy (Other = 1, other =0)	1	8033.5	6.66	0.01
Language Dummy (Nepali = 1, other = 0)	1	8313.3	4.63	0.032
Homework Dummy 1or 2h (1 – 2 hours = 1, other = 0)	1	8029.7	12.15	<0.001
Paid work Dummy (0 hours = 1, other = 0)	1	8212.7	80.80	<0.001
Attitude ”Utility in Mathematics” (Sum, max 15)	15	7994.4	4.6	<0.001
Age Dummy 11 to 12y (11 – 12 years = 1, other = 0)	1	7929.0	15.20	<0.001
Help by Father Dummy (Father = 1, other = 0)	1	7925.3	1.47	0.225
Help by Brother & Sister Dummy (Br./Sis. = 1, other = 0)	1	7913.3	0.93	0.336
Help by Tuition Dummy (Tuition = 1, other = 0)	1	7968.8	1.41	0.236
Help by No One Dummy (No one = 1, other = 0)	1	7923.3	3.50	0.061
Help by Teacher Dummy (teacher = 1, other = 0)	1	7995.0	6.77	0.009
Do you have a textbook of Mathematics (Yes = 1, No = 0)	1	7977.9	19.64	<0.001
Homeworks Not Given Dummy (Not given =1, other =0)	1	7905.1	5.80	0.016
Bullying (Sum, max 5)	5	7997.5	18.57	<0.001
Positive Activities in school (Sum, max 8)	8	7970.6	9.12	<0.001
SES2 (Sum, max 6)	6	7962.4	14.26	<0.001

1) *In the variables Ecological zone, Development region, and Help given, one of the classes needs to be omitted in the analysis because of singularity reasons. Hill zone, Central region, and Mothers' help are omitted; these dummies showed no statistical significance in the Regression analysis.*

2) *Shortened SES; School type is taken away; this enables estimating the parameters for School Type.*

When taking into account the clustered structure in the dataset and the conjoint effect of the factors, quite many of the factors do not show main effect in achievement in English. Such variables are living in the Tarai zone and Western or Far-Western region, Gender, and support provided by Father, Brother or Sister, or Tuition master.[69]

[69]Hill zone, Central region, and Mother's help showed also non-significant effect

Statistically the best factors by using Regression modelling

Traditional Linear regression analysis with Stepwise regression is used to explain the total score by the same variables as are above (see table 5.36). Table 5.37 shows the results.

Table 5.38 Statistically the best model of Linear regression analysis explaining the average of student achievement (Method: Stepwise)

Model	Coefficients				
	Unstandardized Coefficients		Standardized Coefficients		Sig.
	B	Std. Error	Beta	T	
(Constant)	23,91	1.539		15.54	<0.001
School type(0=Community,1= Institutional)	21.06	0.522	0.43	40.33	<0.001
Dev region Valley Dummy(Valley= 1,other =0)	8.78	0.529	0.16	16.59	<0.001
Paid work Dummy (0 hours = 1, other = 0)	4.83	0.414	0.09	11.67	<0.001
Bullying (Sum, max 5)	-1.77	0.148	-0,09	-11.93	<0.001
Attitude"utility of Mathematics" (Sum, max 15)	0.49	0.058	0.07	8.40	<0.001
SES2 (Sum, max 6)	1.02	0.140	0.07	7.28	<0.001
Positive Activities in school (Sum, max 8)	0.11	0.013	0.07	8.65	<0.001
Dev region Eastern Dummy(Eastern=1,other =0)	-6.33	0.634	-0.08	-9.98	<0.001
Dev region Mid-Western Dummy	-5.41	0.615	-0.07	-8.79	<0.001
Caste Brahman & Chhetri Dummy	3.01	0.419	0.06	7.18	<0.001
Dev region Far Western Dummy	-3.22	0.55	-0.05	-5.86	<0.001
Help by Teacher Dummy(teacher=1,other = 0)	-2.83	0.493	-0.05	-5.74	<0.001
Ecozone Mountain Dummy(Mountain=1,other =0)	3.67	0.632	0.05	5.80	<0.001
Caste Madhesi Dummy(Madhesi =1,other =0)	4.01	0.803	0.04	4.99	<0.001
School location (0 = Rural, 1 = Urban)	2.16	0.491	0.04	4.41	<0.001
Do you have a textbook of English(es=1, No=0)	3.94	0.918	0.03	4.29	<0.001
Age Dummy 11 to 12y(11–12 years=1,other = 0)	1.50	0.375	0.03	4.01	<0.001
Homework Dummy1or2h(1–2hours=1,other = 0)	1.48	0.392	0.03	3.79	<0.001
Help by Father Dummy (Father =1,other = 0)	-1.38	0.460	-0.02	-3.00	0.003
Caste Other Dummy(Other=1,other=0)	-1.55	0.543	-0.02	-2.86	0.004

$R = 0.728$ $R^2 = 0.530$ $R^2_{adj} = 0.528$

The model in table 5.37 can be interpreted as follows: The average mean of the students in English is 23.9% of the maximum score considering that the student was in the lowest group in all the factors. If the school was institutional (value = 1), the student's score was, on average, + 21.1% percent higher (note the sign of the coefficient). Additionally, if the students came from the Valley, they gained + 8.8 percent more, if they did not work for paid job, they achieved 4.8 percent, and so on. If, on the other hand, the student came from the Eastern region, the expected achievement level was 6.3 percent lower than if coming from the other regions. Also, if they face bullying, with each step (of five) the achievement level dropped down to 1.8 percent; the difference between the lowest and highest group was 5 × 1.77 = 8.8 percent.

5.3 Summary of Findings

The major findings in NASA 2012 on in English Grade 5 are summarized under three sub headings: basic results, equity indicators and selected explanatory factors as follows:

5.3.1 Basic results

- The English proficiency in grade 5 is not normally distributed. There are two distinctive student populations: students from community schools and students from institutional schools.
- The students in the institutional schools perform well and those in the community schools form two groups of schools: high-performing and low-performing schools. The variations between the community schools is remarkable.
- The learning outcomes are the weakest in the content areas of Reading and Writing and the highest in Grammar and Vocabulary. The differences between the content areas are wider in the community schools than in the institutional schools. This can be caused by the ceiling effect as the test became too easy for the students in the institutional schools.
- Students' ability to solve complex problems is quite low as only 35% of the maximum scores in tasks requiring higher ability were achieved. Students are much better in the recalling type of questions (64%). Remarkable numbers of students (18%) were not able to solve any of the tasks requiring higher ability. The students in the institutional schools are seen to be more able to solve complex problems than their peers in the community schools.
- Students are performing well in the tasks recognizing the correct answers and in recalling simple facts from the texts, fundamental thinking, the basic interpretation of paragraph, table, chart, and a few steps of logical thinking. They are much weaker in producing fluent texts or letters, or preparing synthesis and abstracts from a text. In many cases, the students attempted to do the open ended tasks (like free writing, problem solving and analysis) but the skills were not high enough for achieving the highest marks.
- The dataset reveals that, comparing the 1998 and 1999 results, there is no change in difference in English proficiency between boys and girls. The students in the Mountain zone and Mid- and Far-Western region score.Similarly,the students in urban schools score remarkably higher in 2012 than 15 years earlier. Comparison with the datasets from 20 years back (1993) indicates that the performance in the community schools in Kathmandu Valley has raised remarkably in 2012. On the other hand, the results in Eastern region and Tarai are found to have been lower in 2012 than 15 years ago.
- The average reading proficiency in English among Nepalese students is much lower than the international average of PIRLS standard.
- The most typical 5th grader of English in the community school is at the CEFR level of A1.3. This means that the typical student can read familiar and some unfamiliar words, and can understand very short messages dealing with everyday life and routine events or giving simple instructions. In institutional schools, the most typical 5th grader of English is at the CEFR level of B1.1. This means that the typical student can read a few pages of a wide variety of texts about familiar topics,

following the main points, key words and important details even without preparation.

5.3.2 Equality indicators

- There is a wide difference between the districts when it comes to the equal opportunities of children to reach the pre-set goals in English. The results are very high in the districts where the proportion of institutional schools is high. It is obvious that when the medium of language for instruction in the school is English, it has remarkable positive effect in English language proficiency for the students.
- Except Kaski district (72%), the outperforming districts come from the Central region particularly in the Valley area: Kathmandu (80%), Lalitpur (77%), and Bhaktapur (76%). In all cases the number of private schools exceeds 70% of the schools. Student performance was very low in Khotang (39%), Saptari (37%), Dhankuta (39%), and Udayapur (39%) from the Eastern region; in Mahottari (37%) from the Central region; in Manang (37%) from the Western region; and Jumla (33) and Rolpa (40%) from the Mid-Western region where the number of private schools is small.
- There is a moderate difference between the student performances in four Ecological zones within both community and institutional schools. Students in the Kathmandu Valley outperform the other students. The achievement is the lowest in Tarai area.
- There is wide inequality in the Development regions regarding the children's opportunities to reach an adequate level of English. Especially the wide difference between the community schools in the Valley and in the rest of the country (29% percent points as the highest) is a clear sign of inequality of opportunities in learning English. There are also wide differences between the regions in institutional schools. The difference in student performance in institutional schools between the Valley and Far-Western region is the highest around 15 percent.
- On average, the students in institutional schools outperform the students in community schools. This deviance can be explained by the English as the medium of language for instruction in the schools, since English is used as the medium of instruction for some of the subjects in most institutional schools.
- The students in the urban community schools achieve 12 percent more than the students in the rural areas. Excluding the Valley schools, the difference lowers down to 6 percent. However, the difference is not a good sign towards achieving equality. In institutional schools, there is not wide difference between the rural and urban areas.
- There is an educational inequality among the language groups in possibilities of learning English. In community schools, the students from Magar (59%) and Tamang (55%) backgrounds perform very high in English while the students from Sherpa (24%) and Gurung (25%) communities perform very low. The differences between the language groups are remarkable, which is also not a good sign for equality.

- There are statistically significant though not necessarily remarkable differences between the ethnic/caste groups in English. Dalit (43%) and Janjati students (44%) as well as "Other" castes (40%) are performing significantly lower than the Brahman, Chhetri and Madhesi groups. Dalit students perform lower especially in the Far-Western and Eastern Development regions.
- The differences between boys and girls in English proficiency are very moderate. Though the differences in institutional schools are statistically significant in total score and Writing, the effect sizes are very small indicating that the differences are not at all remarkable, which is a positive sign from equality point of view. A tendency also is seen that the girls are slightly out-performing boys in institutional schools whereas boys are moderately outperforming girls in community schools.

5.3.3 Selected explanatory factors

- Parents' educational level strongly predicts the children's future achievement level in English. Especially harmful for the achievement level is a situation where the father or mother or both are illiterate.
- Either economic or intellectual capacity or both at home helps children to boost their English proficiency. If the father or mother or both are dependent on an agricultural or related occupation, the students' achievement in English is significantly lower than those with the other occupational groups.
- When children have very few home possessions – zero to three out of the 11 – the achievement level is remarkably lower than the national average (< 47%). With nine to ten possessions, the average score is very high (> 70%) compared with the national average. The same is true for home accessories: When none or only one accessory indicator out of three is met, the results are lower than average (45%–50%); when there are two or more are met, the results are remarkably higher (58%–64%).
- Socio-economic status plays a vital role in the English language achievement in Nepal. The difference between the lowest and highest SES groups in grade 5 English proficiency is remarkable (40 percent). This means that if the SES of the lowest performing students is raised into a decent level, that is, in practice, that the problem in parents' low educational level is solved, the results in these groups will raise remarkably. Especially challenging is the situation in the families where the father or both parents are illiterate or both of them work for agriculture.
- English as medium of language for instruction in schools practically determines the English proficiency level. As English is used as the medium of instruction in most of institutional schools, students from institutional schools always have the possibilities to have higher level proficiency in English. Such a situation is not very easily changed to be equal in English proficiency between community and institutional school.
- Either working for a paid job or for four hours per day in an unpaid household work outside school lowers the achievement of students in English. However, a decent

amount of household work up to two hours per day is not found to have affected the students' learning in English.

- Positive attitude towards the subject correlates with positive achievement in English. The better achievement is more probable a consequence of more positive attitude rather than the other way round.
- The highest performance in English is found among the students studying at their normal age group, that is, at the age of 10 and 12 years. Otherwise, the achievement lowers down as the age increases.
- The support provided by the mother, brother and sister raises the achievement level more than the support provided by the father or teacher. In the whole sample, the highest achieving group is the one that has received private tuition. However, the difference between the highest and lowest performing groups is not notable. It is possible that the group with private tuition also spends more time on their homework, which explains the higher score among them.
- As many as 4.3% of the students lack the proper textbook in English even by the end of academic year. The achievement level of these students is significantly lower than those who have access to the textbook.
- If the teacher assigns homework and checks it systematically, the achievement is higher than those without getting it. The regular assignment of homework with its check, even if not done every day, is found contributory for raising the scores up to 13 percent.
- There is a strong connection between the future aspiration of students and their achievement. As the student aspires to hold a professional career like a doctor, engineer, or aspires to go abroad, their achievement in English is higher than the average. The number of students who aspire to be a teacher, an engineer, or a doctor is remarkably higher (72%) than those aspiring to be a farmer or business person.
- An alarmingly high number of the students (54%–56%) have encountered bullying in school. The phenomenon tends to be affecting the learning outcomes in almost all the groups of the students who felt bullying calling for all possible efforts to put to root the phenomenon out from the schools.
- When the students feel that the actions of the teachers and the schools are ultimately good, the English results are better than average (47% in community school and 80% in institutional schools). At the other extreme of feeling that such actions are ultimately negative, the results are far below the average (35% in community schools and 67% in institutional schools).

Chapter 6: Conclusions and Implications

It is established worldwide that the subject specific knowledge and literacy and numeracy skills acquired through school education prepare youngsters for better adult life on which personal development of an individual, and prosperity of a nation largely depends. These knowledge and skills are also fundamental for both technological adoption and innovation to have direct impact on an individual's success in the job market, economic growth of a nation and promotion of equality in the society. Moreover, this being the age of knowledge economy, national development largely depends on the availability of highly skilled labour force, its capability for innovation, and the intellectual property it generates. With this understanding, nations across the globe urge 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 get decent job for making their lives better, to play a role in building more peaceful and equitable societies (Matsuura, 2004), and to be able to face future challenges. For this purpose, the practice of measuring and monitoring students learning achievement by means of large-scale assessment developed worldwide during the last decade of 20th century, which includes, among others, international level of assessments like PISA, PIRLS, TIMSS in industrialized countries; regional level assessments 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) that were institutionalized during the same period. Nepal is no exception in adopting national assessment, which introduced it since 1995 for measuring students' achievement in order for monitoring it and providing policy feedback to the system.

This chapter along with summarizing the context briefly highlights objectives, methods, basic principles and standards of the assessment, presents major findings of the assessment. Before concluding it also draws implications of the assessment.

6.1. Context

Various assessments of student achievement in Nepal were also conducted before 2011 for various grades (in 1995, 1997, 2001 for grade 3; in 1997 for grade 4; in 1998, 1999; 2003, 2008 for grade 5; in 1999 for grades 6 and 8 and again in 2008 for grade 8). Though small scale in nature, they all aimed to assess learning outcomes of students to determine the level of learning for the respective grades and to provide policy feedback to the system. Large-scale assessment for the first time came into practice since 2011 after the establishment of the ERO under the MOE. During the implementation of SSRP 2009-2015, four assessments, two rounds for grade 3 and two rounds for grade 8, were planned to accomplish. Accordingly, a large-scale assessment for grade 8 has already been accomplished in 2011. Next to it, this is the first round assessment for grade 3 and 5 conducted in 2012.

6.2. Objectives

Like other student assessments institutionalized at international, regional and national level, this assessment was designed to generate accurate, objective and comparable data on student learning in order to assess the health of primary education system in Nepal. At the same time, it also aims to describe national levels of learning achievement in key subjects (Mathematics and Languages) and to determine the extent to which primary graders have developed a fundamental understanding on them. More specifically, this assessment is motivated to:

- Determine the current national level of achievement of grade 3 students in Mathematics and Nepali, and of grade 5 students in Mathematics, Nepali and English languages against the goals set in the curricula;
- Analyse variations in student achievement by region, gender, location, types of schools, and language of instruction;
- Explore the factors that influence student achievement in primary education;
- Create reliable baseline data for the future;
- Generate recommendations for policy making to improve educational quality;
- Compare student learning achievement in the current study with that of the previous studies of Nepal and international ones like TIMSS,PIRLS and PISA;
- Generate evidence-based data for monitoring the trends in students' achievement for these subjects over the period of time.

6.3. Methods

This study was conducted in 1690 randomly selected schools from 28 sample districts covering all Ecological zones and Development regions to assess the learning outcomes of 3 graders in Mathematics and Nepali and 5 graders in Mathematics, Nepali and English. Altogether 80,232 students (38,753 in grade three and 41,479 in grade five from randomly stratified 1,690 sampled schools) participated in the assessment. In the sample for grade three, 17,256 students were boys and 17,166 were girls. Similarly, out of the total sample for grade five, 19,617 students were boys and 19,783 were girls. Out of the 75 districts of Nepal, the dataset represents a random selection of 28 districts covering all five Development regions (Eastern, Central, Western, Mid-Western, Far-Western), and the Kathmandu Valley, as well as all Ecological Zones (Mountain, Hill, Tarai). In addition, both rural and urban schools as well as community and institutional (private) schools are proportionally represented so that the results of the assessment can credibly be extended to the whole student and school populations of Nepal.

Final items were standardized after pre-testing in 2000 students in 240 schools from the districts representing different strata. Only those items having high discriminatory power were included by analysing difficulty level of each them. Based on the pre-test results of the items, the difficulty levels of the tests were set around 50 –60%. Some linking items from international tests like PIRLS and TIMSS were also used to make the test results comparable to the international standard. All the items were analysed and equated using IRT modelling.

Three versions of the items in each subject were administered and the final scores were equated by utilising the IRT modelling. Reliability of the tests was found high and the validity was assured by applying specification grids of the national curriculum developed by the Curriculum Development Centre (CDC). From methodological standpoint, the process and practices of the inquiry have successfully followed the procedures as used in some international level tests with some contextualization on them to reflect the reality of the Nepali context. Thus, this test is believed to fulfil the national and international ethical principles, criteria and standard to qualify it a credible assessment. The results were linked to the set of results from the 2008 assessments as well as to the international item banks of TIMSS and PIRLS.

The tests were administered uniformly 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 selected subjects for each grade. Thus, the students in a grade were required to participate in one of the selected subjects assigned to the school. The answer sheets were marked and achievement scores were tabulated using Optical Mark Reading (OMR) machine.

The results are reported mainly as percentages marks where 100 (%) represents all tasks solved and 0 (%) none. As a result of the pre-testing of items, the difficulty levels of the tests were set at 50—60%. For Nepali language assessment, the CEFR was also used to obtain the level of students from language proficiency point of view.

6.4. Basic Principles

Student assessment, particularly the large-scale one, needs to be grounded on some ethical principles. According to Race, Brown and Smith (2005), an ethically good assessment should be just and fair, valid and reliable, transparent, motivating, sufficiently demanding having possibility to show excellence. Based on these premises, this test has the following characteristics:

1. The tests are considered equal, just and fair to all the sample schools and pupils as the test setting in all schools makes it possible that no single school or student is favoured. The pupils who participate in the tests have studied, in principle, the same national curriculum thorough out the country. Hence, they all are supposed to learn the same contents. These common contents are measured with a common test.
2. An attempt is made to keep the validity and reliability of the indicators at a high level as much as possible. By using item analysis, it is particularly taken care that the individual questions and the whole tests are of high quality as much as possible.
3. The measurement process is transparent in the sense that there is no hidden objective or surprises or traps set in the test and in its procedures. In the test, only important matters, based on a public document of curriculum, are measured. The test items, however, are not public in order to guarantee the possibility of using

linking items in the years to come. This secrecy of the test items is an international practice.

4. When compiling the tests, an attempt is made to keep the test versions versatile and motivating. The low-stake role of the assessment and the textbook-independent testing facilitates the fact that the students do not study for the test.
5. Attempt has always been made to conduct the tests in a way that the students are able to show their excellence. On the other hand, the fairness implies that it is important to motivate the weakest students. Hence, some very easy items are always selected in the test batteries .

On these grounds, one can conclude that this test fulfils the ethical principles as far as possible in the large-scale assessment at national level.

6.5. Pragmatic Standards of Assessment

Though overlapping partly with the ethical principles suggested by Race, Brown and Smith (2005) above, Ivernizzi and colleagues (2005) maintain that a good assessment takes into account the technical and pragmatic standards for a good assessment as validity and reliability, standardized administration, scoring and reporting, comparability of the scales, norms and scores, fairness in testing, the test-takers' separate language and cultural background. On these grounds, this test makes an attempt to the extent possible meeting the technical and pragmatic standard as follows:

1. The trustworthiness of the measurement instruments – reliability and validity – is secured separately to every test version. The level of accuracy has been tried to rise as high as possible by pre-testing the items and by selecting only highly discriminating items for the final tests. Validity has been secured by wider content coverage and by following the content weights allotted in the curriculum.
2. The test arrangements, scoring and reporting have been administered and centralized in an equal manner for all schools. The centralized marking of the papers by experienced teachers makes the scoring objective and fair for all students and schools.
3. IRT modelling is used to calibrate the different test versions with each other and, finally, to equate the test score over the versions. In this manner, it is possible to reduce the bias coming from the slightly different difficulty level of the test versions. Hence, serious attempts are made to keep the scales, norms and scores comparable over the different versions as well as over the different years within the same subject.
4. Fairness of the testing is assured in four ways: First, the tests are comparable with each other (over the years). Second, only the central contents are measured in the tests. Third, the tests are based on the public document of national curriculum. Fourth, the test results are openly reported and interpreted by using the generally known statistical methods. The interpretations concerning the school or the student are never based on only one indicator.
5. In the test arrangements, attention is paid to the fact that two language versions of the test papers are in use (Nepali and English). To put it another way, the same item

is present in two languages in the same test paper. In language testing (like in Nepali and English), the test paper is in one language. The mathematics test papers were administered with two languages (Nepali and English). It may be possible that in the future also some other languages are used – especially in the lower grades because officially the government is supporting to use the indigenous languages at the lower grades. The language issue is challenging in Nepal, where over 120 languages exist.

One can conclude that this achievement test also fulfils the practical principles as far possible in the national level assessment. There are some issues – such as the language of the tests – which need to be considered critically in the assessments to come.

6.6. Major Findings

At the outset, learning achievements for both grades look satisfactory in terms of the mean score which is 63% at the highest and 60% at the lowest in Nepali and Mathematics for grade 3, whereas for grade 5 it lowers down to 53-54% in Mathematics and English. Grade 3 students are ahead of the 5th graders in both subjects. To look at the mean achievement from equality point of view, gender parity is maintained in both the grades in Nepali and Mathematics where girls are ahead of boys by 1-2% percent in Nepali. On the contrary, though the difference is not so wide, the 5th grader boys are ahead of girls by 1-2% percent in Mathematics and in English. This level of performance in terms of national average above than 60% for grade 3 with almost equal performance of boys and girls in both the grades is seen promising indicating a good functioning of the system. However, when seen from other perspectives going deep down into the results on different variables, low levels of learning with high discrepancies are also noticed. Having analysed the database for each subject as described in earlier sections, the following conclusions are drawn as the main results.

I. Divided students population into three distinct groups

When analysing the distribution of student population, it is seen to have been divided into two to three distinct groups in terms of the achievement score with remarkable variances: both low and high performing students in community schools and mainly high performing students in the institutional schools. In comparison of Mathematics and Nepali for both grades and English for grade 5, the 5th graders' population for Mathematics is more normally distributed. In community schools, 50% students belong to the low performing groups below the average forming a large plot, whereas in institutional schools more than 50% students lie at above than average achieving population forming its bigger plot with a small plot of low achieving group. Because of the large population of low achieving students in community schools, the entire system is seen to be shifted towards low performing one indicating that the system is not able to give sufficient support for those students who are lagging behind in the early grades.

II. Wide differences in achievement among students, schools, districts across all locations, areas and regions

A wide difference is found in the achievement levels among students and schools in Nepal. Some students (0.2 - 4%) in both grades were not found able to answer even single test item (0%) while the best students achieved more than 98%. Of the sample, the mean achievement for 4.35- 6.33 percent schools was below 30% whereas only 13-19 percent schools obtained above 80% in Nepali. From an equal opportunity point of view, it is not a good sign that 'low' and 'high' performing students are concentrated in certain schools. Particularly high numbers of low performing students are concentrated in community schools.

Not only the variation is great amongst students and schools, but it is also significant across the districts with 32-44 percent difference in Nepali; 34-38 in Mathematics and 47 in English. In the lowest performing districts the average performance ranges from 33 to 48 %, while in the highest performing districts it ranges from 60 to 81% (vary by subjects). The students in the Kathmandu Valley schools exceed the average students in numerous indicators across the regions other than Kathmandu Valley. The achievement level of the students for the Mid-Western (46-54% in Nepali, 44-46% in Mathematics and 44% in English); the Eastern (52-54% in Nepali, 45-52% in Mathematics and 39% in English), the Central and the Far-western (54-59% in Nepali, 51-58% in Mathematics and 42-49% in English) Development regions is far behind that of the students in the Kathmandu Valley (78-81% in Nepali, 72-79% in Mathematics and 76-80% in English). Once again, from the perspective of equal opportunities, this is not a positive sign.

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

Students' average achievement in institutional schools is found remarkably higher (for example, with 24- 27% gap in Nepali, 18- 21% in Mathematics and 13 % in English) than in the community schools. One of the reasons behind this may be the families having higher socio-economic status have been sending children to institutional schools as higher socio-economic 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 institutional 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 socio-economic 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.

IV. Unbalanced learning across all curricular contents

Against the expectation of curricula, the dataset is evident that certain contents of the curricula are learnt less effectively than others. For instance, in Mathematics, the achievement level in Algebra and Numeracy is remarkably lower than Arithmetic and

Geometry; in Nepali and English, the Reading and Writing skills are poorer compared to the achievement in vocabulary. Circumscribed with such unbalanced learning on some domains, the entire system is shifted towards low performing, making it less effective to yield better results.

V. Lower level of cognitive ability for the tasks requiring higher ability

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. Of the total, 4-20% students in Mathematics, 3-18% in Nepali, and 18% in English could not solve any of the problems requiring higher ability. By types of school, the students in institutional schools are seen to be more able to solve complex problems in comparison to their peers in community school with a wide variation (47.6-70.5% in grade 3, 37- 51% in grade 5 for Mathematics; 32-54% in grade 3, 42-66% for Nepali in grade 5; and 23- 68% for English). More discouraging fact is that the higher the grade, the more is the population with lesser higher ability tasks.

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 or problems requiring longer procedures of higher cognitive level.

In Nepali and English subject, 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. The students tended to attempt open-ended tasks but the skills were not high enough for obtaining the highest marks. In the case of applying the gained knowledge in new situation, they are found very poor compared to the higher level thinking ability.

VI. Significant effects of caste/ethnicity and home language background in determining the students' achievement

The data is evident that, in community schools, the students from “Other” ethnic/caste groups are performing the lowest (50%) in Mathematics, followed by Dalit (53%), Madhesi and Janjati students (54 and 56% respectively). Dalit students perform below the average also in institutional schools. The overall difference between the groups is statistically significant ($p < 0.001$) though the effect size is small ($f = 0.12$) in community schools; dividing the students according to their caste background explains just 1.4% of the student variation ($\eta^2 = 0.014$). In institutional schools, the effect size is also small ($\eta^2 = 0.014$); dividing the students according to their ethnic background explains just 1.5% of the student variation ($\eta^2 = 0.015$). A positive sign from the equality point of view is that Dalit students perform better than the national mean (53%) in the Eastern (66%), Central (61%), Western (88%), and Far-Western (63%) Mountain areas as well as in the Western Tarai (63%) (Table 3.1.20). However, it is seen that results are much lower than the average in the Eastern (45%) and Mid-Western Tarai (39%), and Mid-Western Hill (43%). Especially low performance is found in Central Mountain area (35%). Among the Dalit students, boys slightly outperform girls (4 percent). Otherwise the differences are not notable.

Brahmans are ahead of all other ethnic groups whereas Madhesis and Dalit students' Nepali language ability is seen to be poorer. The ethnicity-wise effect size of high performer and low performer is medium in both grades. However, the variance explains more in grade 5 (7%) than grade 3 (5%). Except in Madhesi group, girls' language ability is better than boys in all caste and ethnic groups.

Variances in achievement are also seen in terms of the language spoken at home indicating an inequality among the language groups in possibilities of learning. The dataset reveals that there is an educational inequality among the language groups in the possibilities of learning Mathematics. In the whole data set, the students in grade 3 from Magar (65-79%) and Tamang (62-73%) backgrounds perform very high while the students from Tharu (50-52%) and Gurung (38-51%) background perform much lower in Mathematics with a remarkable difference between the lowest and highest groups. In Nepali, similar level of variance is seen between them. In the data set, the students from Magar (76-81%) background perform very high while the students from Tharu community are the lowest performing ones (40- 56%). Quite low performing are also the students having Newari (54-55%), and Gurung (47%), Rai (51%), Sherpa (52%) backgrounds. With regard to the performance in English, a remarkable inequality is seen in possibilities of learning English. In community schools, the students from Magar (59%) and Tamang (55%) communities perform very high in English while the students from Sherpa (24%) and Gurung (25%) communities perform very low. The differences between the language groups are remarkable.

VII. Significant differences in student achievement in terms of school location

Urban students outperform the rural students in all subjects. For instance, urban students achieved 74-75% in Nepali, 65-71% in Mathematics and 73% in English, whereas rural students scored 56-59% in Nepali, 50-56% in Mathematics and 48% in English. The achievement gap between urban and rural students is wider in English by 25% and in Mathematics by 18% - the rural schools lagging far behind. The rural -urban difference is remarkable and is not justifiable on any ground from equality point of view and urges for a serious attention.

The difference between grade 3 and grade 5 students' achievement level is more prominent and the performance gap is wider in grade 5 than in grade 3. The difference is statistically significant in Mathematics at $p < 0.001$. The effect size is medium (Cohen's $d = -0.56$ in grade 3 and $d = -0.67$ in grade 5). Generally, institutional schools are located in the urban location, which is inflating the achievement score of the urban areas. Nearly 90 percent community schools both with grade 3 and grade 5 are found located in rural areas. However, among the community schools 75 percent with 3rd grade and 78 percent with grade 5 are located in rural areas. While excluding valley from the analysis, still the difference is significant ($p < 0.001$) and effect size is medium (Cohen's d = – 0.39 and d = – 46 respectively in grade 3 and 5).

VIII. Moderate but growing differences between the Ecological belts and rural/urban schools

Dataset shows that the students from the Mountain region are slightly better in learning achievement than other students. The lowest achievement is found in Tarai but the difference is not wide in comparison with the students from Hill when the students from the Valley are excluded. The Valley students outperform those from the other regions. When it comes to the school location, the urban community schools outperform the rural community schools by 6–12 percent; excluding the Valley, the difference is 0–6 percent. For one reason or another, there is not wide difference between the rural and urban institutional schools but the trend is serious. In comparison to the previous results, it is seen that the students from Tarai are performing lower. In Mountain, they are performing higher than about 15 years ago. The urban schools have raised their position remarkably in comparison to the rural schools. If these trends continue in the future, it will lead to a wider inequality in society between rural and urban areas as well as between the Ecological zones. Ultimately this will lead to uncontrolled urbanization if the families continue to send their children to big cities to study and move later themselves to seek a better life.

IX. Wider disparity in achievement between the districts and Development regions

To look at the results from equality point of view, inequality persists between the districts and regions across the country revealing wide differences between the districts to give equal opportunities in reaching the pre-set educational goals. The average achievement in the Kathmandu Valley is very high compared to the lowest performing districts in the sample. Differences in the mean scores between the lowest and highest scoring districts vary from 33% to 43% in Math and Nepali and 51% in English. In English, the difference is found to be connected with the proportion of institutional schools in the district as the medium of instruction in most of the institutional schools is English, which has helped to raise the achievement. In other subjects too, the result is seen to be the higher more when there are more institutional schools. Some districts showed very high achievement without any institutional schools in the sample. More crucial is that, in the lowest performing districts, the average achievement is absolutely very low in the districts like Bardiya, Rolpa, Jumla, and Udayanpur in Mathematics; Saptari, Achham, and Mahottari in Nepali; and Khotang, Jumla, Saptari, and Mahottari in English. These seven low-performing districts concentrate in two Development regions: Eastern (three out of seven) and Mid-Western (three out of seven); which are the lowest regions in each dataset. Among the Development regions, the Kathmandu Valley is 21–29 percent units ahead of the rest, prevailing a wider disparity among them.

X. Association of low socio-economic status with remarkably lower learning achievement

Socio-economic Status (SES) and its components are found to have been strongly associated with the learning achievement in Nepal. The difference in achievement between students from the lowest and highest SES groups is 23–40 percent. The widest difference is found in English subject at 40 percent whereas it is 23 -30% in Mathematics and at 30-

31% in Nepali. Depending on the subject, 11–26% students are at the lowest level of SES meeting none of the seven indicators.

Results show that especially low achievement is common among the children whose parents are illiterate. As reported, 34–42% students have illiterate mothers and 15–19% of them have illiterate fathers. Similarly, the lowest result is also common among families where either mother or father or both are engaged in the agricultural occupations. According to the datasets, 53–66% mothers and 35–41% fathers work in the agriculture or are involved only in household chores.

When the children have very few home possessions or none of the home accessories, the achievement level is remarkably lower than the national average. According to the datasets, 2–7% of the students did not have any of the eleven home possessions including table, dictionary, peaceful place for exposition and the like; and 26–45% possesses neither TV, mobile phone nor the computer at home. If the issue of parents' low educational level is solved, the result in the low-SES group is likely to improve.

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

The same result was found in grade 8 too (ERO, 2013), and hence it indicates that the issue is a structural problem in Nepal. Especially, structural problem is the high rate of mothers' illiteracy and their low educational level. Right now, practically all the girls are attending school and their children will have better chances in education after 15–25 years. However, without targeted intervention from the MOE, within the next 15–25 years, it is not foreseen that a remarkable portion of students will have a mother with even the minimum level of reading skills.

XI. Involvement in paid job or work for many hours a day impeding learning

The results in all subjects show that students, either working for a paid job or spending more than two hours per day for unpaid household chores before or after the school, are found to have lower achievement level. The dataset shows that 27–36% students work for the paid job and 15–23% of them spend more than 2 hours in household chores. Though most of the low-grader students do not usually work many hours per day for the paid job, the number is notable. The pertinent question is: Why they need or are willing to work daily? Most probably they need to earn for pocket money or for subsistence livelihood. The need for working in the paid job or need to participate more than 2 hours in the household chores is only one part of a complex knot of problems involved with the low SES affecting the learning results. Though child labour is prohibited by the law, something more is also required to prevent school children from working for a paid job. It is seen that, in community schools, some work in household chores up to daily 2 hours has not lowered the level of children's achievement.

XII. Effect of the unavailability of textbooks on achievement

Despite the concerted efforts to make textbooks available in the hands of students in time, the datasets reveal a distressing fact that a significant number of students (5.7-5.8-% in Mathematics, 3.67- 4.3-3% in Nepali and 4.17% in English) are found studying the subjects without textbooks even up to the end of academic session. This problem has resulted in lowering the achievement scores by 7.2-9.5-8% in Mathematics, 10.5-13.4% in Nepali and 8.4% in English compared to the students having textbooks in time. The difference in achievement between having and not having textbook is greater in upper grades than the lower one, which is 13.7% in Nepali and 9.5% in Mathematics for grade 5.

XIII. Negative effect of school bullying on learning students' potential

Though bullying at school goes unnoticed to many of the parents and school teachers and hence they remain reluctant to understand such incidences experienced by students in school, it has keept on worsening learning environment negatively impacting the learning potential of students and thus is taken as a negative activity in school. The obtained datasets reveal that 45- 46% of the students (44% in community schools and 52 % in institutional schools) did not encounter any bullying during the last month implying that the remaining 54- 56% did encounter at least one type of bullying. As many as 5.8- 9.7% students (6.6-11.7% in community schools and 3.2-3.9% in institutional schools) in grade 3 are experiencing a severe kind of bullying (the sum of 80% and 100% bullying). This means, in practice, that more than 83,000 students in these grades in Nepal have been encountering physical, psychological, and social bullying every month. It is seen that learning outcomes are notably lower than the average with 15- 19% of the students who have encountered more than two different types of bullying (36–52% in community schools and 63–68% in institutional schools). Students who do not experience bullying and those who encountered extreme bullying of five kinds have 19–23 percent achievement gap; though there is a small number of students who reported this kind of bullying (n =351- 875). The difference is statistically significant though the effect size is small or medium (f = 0.18-0.19 in community schools and f = 0.11-0.04 in institutional schools). Though extreme cases of severe bullying are rare, bullying is found quite rampant in schools which has been causing needless constraints to young children and has to be rooted out from the schools.

XIV. Boosting students' achievement through homework

Properly assigned and checked homework which is an integral part of teaching-learning process boosts up learning achievement as it provides ample opportunity for self-learning and engages students in problem solving. However, the available datasets of student assessment evidently shows that nearly 3.2-3.5% students in Mathematics, 2.3-2.9% in Nepali and 3% in English are found never assigned or checked their home works. The students getting homework with its checking are found to have scored 55-61.4% in

Mathematics, 61.9-65% in Nepali and 46% in English, whereas students with no homework (or getting no checking) in those subjects score below 11-13.9% in Mathematics, 15- 16 % in Nepali and 13% in English. So promoting the practice of assigning homework in some forms (drill, exercise on solving problems, and some kind of writing) along with its regular checking is seen essential not only to boost achievement but also to promote study habit of children.

XV. Association of schooling at over age with lower achievement

Higher performance is found with the students studying at their proper age with the peers in normal age group, that is, at the age of 8–10 years in grade 3 and 10–12 years in grade 5. Their achievement lowers down as the age increases. For instance, the mean achievement ranges from 61 to 64% in Nepali, 56 to 61% in Mathematics for the students studying at their proper ages, whereas it lowers to 57% in Nepali and to 55% in Mathematics for students who are over-aged. According to the dataset, 25–30% students are over or under aged for the grade. The same phenomenon was also observed in grade 8 (see ERO, 2013) too indicating that delayed schooling or non-systematic entrance in schools is a structural problem for the educational system.

XVI. Noticeable changes in leaning outcomes over the past decades

The changes in learning achievement results have been remarkable in some areas while in others nothing has changed. For example, in Nepali and English datasets of grade 5, there is not much difference in achievement in comparison to the previous datasets between genders whereas in Mathematics datasets the gap between the genders has reduced which is a positive sign. On the other hand, the English and Nepali datasets hint that the students in the urban schools have gained remarkably higher during the years. This may have been caused by urbanization, concentration of educated families and due to the influence of private and boarding schools. It is also seen that the students in the Mountain zone have gained remarkably higher compared to the Tarai students. When it comes to Development regions, the Eastern region is seen to have lowered down while the Far-Western has made remarkable progress both in ranking and in absolute terms. For the Far-Western region the change is positive, but for the Eastern region it is naturally not a good sign.

XVII. Low level of performance of Nepalese students in comparison to international standards

When Mathematics dataset is compared with the international standard, Nepalese students are, on average, one year behind the international average, and the 5th graders are somehow at the level of grade 4 students. In Nepali and English, the average reading proficiency of grade 5 students is much lower than the international average of grade 4 in PIRLS standards. The datasets in Nepali hint that reading proficiency is lower than one year behind the international level and the estimated level of grade 4 students is around 1.5 standard units lower than the international mean. In Nepali and English, the grade 5 students are far below the grade 4 international average (-1.4 standards units in Nepali and -1.2 in English). The same kind of result was found also in grade 8 (ERO, 2013). However, within the dataset, there are several very highly performing students but their number in total is so low that they do not raise the national standard.

In comparison to the PIRLS standard, the achievement level of an average grade 3 student in the community schools ($\theta = -1.78$) is very low compared with an average international student of grade 4. The achievement level of an average student in the private schools ($\theta = -1.08$) is higher than his/her peer in the community school but still far behind the international average. In terms of the language proficiency level, the typical grade 3 reader of Nepali in community school is at the CEFR level of A2.1. This means that the typical student can understand simple texts containing the most common vocabulary, the main points and some details of a few paragraphs of text. In institutional schools, the most typical grade 3 reader is at the level B1.1, that is, s/he can understand the main points, some details of messages consisting of a few paragraphs in fairly demanding everyday contexts, factual texts and acquire easily predictable new information about familiar topics from a few paragraphs of clearly structured text.

Regarding the proficiency of 5th graders, the data shows that the average reading proficiency of students in Nepal is remarkably lower than that of grade 4 students at the international level. The students in Nepal are remarkably lower in all the content areas of language subject compared to the international average-the PIRLS. The average achievement level of grade 5 students in the community schools ($\theta = -1.58$) is very low while compared to the average international students of grade 4. The achievement level of an average student in the institutional schools ($\theta = -0.80$) is also remarkably lower than the international mean. Similarly, the dataset shows that the most typical 5th grader student of Nepali community school is at the CEFR level of A2.1. This means that the typical student can read and understand simple everyday texts and factual information and interpret them at a slow pace. In the institutional schools, the most typical 5th grader student of Nepali is at the CEFR level of B1.1. This means that the typical student can read a few pages of a wide variety of texts about familiar topics, following the main points, key words and important details even without preparation.

In English, reading proficiency of Nepalese students is generally below the international average against the PIRLS standard. This indicates that the students in Nepal score remarkably lower than their international peers. The achievement level of an average grade 5 student in community schools ($\theta = -1.49$) is very low compared with an average international student of grade 4. The achievement level of an average student in institutional schools ($\theta = -0.39$) is much higher than his/her peer in community school. Against the CEFR standard, the most typical 5th grader student of English in community school is at the level of A1.3. This means that the typical student can read familiar and some unfamiliar words, can understand very short messages dealing with everyday life and routine events or giving simple instructions. In institutional schools, the typical 5th grader of English is at the CEFR level of B1.1. This means that the typical student can read a few pages of a wide variety of texts about familiar topics, following the main points, key words and important details even without preparation.

Given context raises the question: 'What would be the fastest and most feasible way to increase the national achievement level in Nepal?' One lesson which could be learnt from Finland of high results and low variability between the schools (see e.g.Schleicher, 2006; Metsämuuronen, Kuosa, & Laukkanen, 2013) is the strong emphasis on support for the students at the early grades.

6.7. Implications

The low level of learning compounded with a wide gap in achievement among and between rural-urban, community-institutional schools, caste/ethnicity, social groups, across all Development regions, Ecological belts and districts are neither a good sign for equality and nor the positive indications for the system. Similarly, achievement variations of students in 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.

I. Reducing inequality in achievement

As the result confirms a wide inequality in achievement, which continues to persist between students from rural and urban locations, among various language speaking and ethnic/castes groups, and among districts, regions and ecological belts; it has been an imperative for the policy makers, curriculum planners, 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 one attends. Although a lot of efforts have been put into the system to reduce 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 inequality and improving the 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 additional support to the students attending institutional schools. If the rigorous teaching-learning practice and caring environment available at institutional 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 reduce 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.

II. **Improving reading ability**

As confirmed by the dataset, 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 the 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, comprehension exercises, introducing varieties of texts with different genres by spelling out specifically the standards on fluency, accuracy, performance and problem solving in curriculum and textbooks across the subjects. Another measure for fostering reading ability would be determining ceiling 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 be sensitized, oriented and trained in such instructional design. Allocating additional time for reading activity in school hours through curricula provision would also be the another 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 the skills of listening, speaking, oral presentation, reading fluency and accuracy. So, teachers 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.

III. **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, the fundamental goal of education is to equip students to think critically, solve complex problems and succeed in the society and economy of the 21st century (Fin et al, 2014). The subject-wise datasets reveal 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 in one context into the new one. Provided situation leads to conclude that either our teaching-learning activity deviates from organizing activity to engage students in the tasks requiring higher ability, or reading materials including textbooks lack adequate exercises that foster given ability.

Our education system needs to be well aware of the reasons behind why schooling is not promoting the desired higher cognitive skills among lower graders even after completing five years of 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 students engaged in tasks demanding higher cognitive ability. Similarly, teacher educators and training module designers also need to look for the possible measures to incorporate the 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 re-oriented 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.

IV. 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 not to have answered 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 possible to assess both the process and product. 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 the desired capacity in designing and using various assessment tools including standardized test items against the learning objectives set in the curricula. Schools should be re-oriented on the various assessment tools and processes 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 developed based on the curriculum objectives.

V. Rooting out the incidences of school bullying

The phenomenon of school bullying, as recognized in other parts of the world, is not so much familiar to parents and teachers, so it has gone unnoticed in schools; but it is found rampant in Nepalese schools in some forms negatively affecting the learning potential of those children who experience it. 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. Given situation alarms the concerned teachers, education managers and even the parents to be aware of the phenomenon and to expedite the possible preventive measures that help schools at least minimising such incidences.

One of the measures for this would be making teachers aware of the issue and its possible consequences among 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.

VI. Raising parents' educational level

Low level of parents’ education is found to have been directly associated with the under-achievement of their children at school which further has been the main source of perpetuating social inequality and disparities. Despite the 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 the adult population [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 to be the major determinant of their children's achievement level at school. So, raising parents’ literary level especially targeting the mothers is seen an urgent need not only for achieving higher achievement level at school but also for achieving social equality and a just society. One of the probable measures for raising parents’ literacy level would be devising a policy and program that encourages schools to make those illiterate parents literate residing within the catchment area and whose children are attending the particular school. For this, strict enforcement of 'Student Mobilization Guidelines for Literacy 2070' developed by the CDC and endorsed by the MOE would be 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 parents, improve the achievement level of their children, and develop strong school community relations that ultimately increases parental involvement in school in development endeavours.

VII. Achieving balanced learning in all content areas across the subjects

Besides the problems of low achievement level, Nepalese students are not found to have developed similar levels of knowledge, skills and competencies equally over all content areas within and across subjects. Little learning in one area of subject impedes to acquire expected level of proficiency 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 and English, Algebra and Numeracy in Mathematics are poorly learned areas in comparison to others.

Given the context of such unequal level of learning, curriculum planners of the respective subject first need to seriously consider 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, the 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 identified areas.

VIII. Ensuring timely availability of textbooks to all students

Despite the government's efforts for the delivery of textbooks 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 the availability of textbooks to all students in time. One of the measures towards this would be enforcing the provision to earmark certain amount of fund in 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 distribution of textbooks. In this context, the existing mechanism of textbook 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 ensuring the availability of textbooks with good quality in time is to introduce multiple textbooks system. One of the crucial factors regarding the use of textbooks is enhancing teachers' capacity to use multiple materials including printed, electronic and online materials in order to deliver the curricular competencies effectively.

IX. Catching up with international standards

Having compared the datasets of English, Mathematics and Nepali with the TIMSS and PIRLS standards, Nepalese students are seen behind their international counterparts of the same grade level which will likely handicap them to be competitive in the world of job market. Taking the issue into consideration, further researches and studies need to concentrate in exploring the reasons behind it as regard whether it is due to curricula or classroom practice. Similarly, the MOE also needs to initiate dialogues and discussions with academia, curriculum planners and policy makers on how to raise the performance of Nepalese students to the international level in order to make them competitive in the global world.

6.8. Conclusion

This study has not only determined the achievement level of primary grades in each of the three subjects, but also has analysed variations in student achievement in terms of Development regions and ecological belts, gender, location and type of school, language of instruction, caste or ethnicity of students, their socio-economic background and activity at schools. While doing so, it has also compared the achievement of Nepalese students to international tests like TIMSS and PIRLS.

Having analysed the results, the dataset shows some gaps within the system. First, there is a wide inequality in learning opportunity for children between regions, types of schools and their location, regions and ecological belts, language groups, caste and ethnicity. Second, inequality is also seen in learning of contents and domains across the curricula. Third, there is low level of proficiency in higher order skills in all the subjects. Moreover, it has also confirmed that the achievement of Nepalese students is lower than international standard lagging behind their international colleagues while comparing to the TIMSS and PIRLS results. Given the context, further reform strategies and programs are required to concentrate towards narrowing the inequality or gaps between regions, districts, locations, social and linguistic groups; enhancing higher level cognitive skill; and raising the standard of Nepalese education to the international level.

The study has generated rich data on different variables. This does not only serve the purpose of benchmarking for future but also provides numerous insights for policy planning, program designing and identifying the areas for reform. As it has generated micro level data of each individual school, specific support and intervention can easily be identified and targeted to the poor performing schools in order to support them raise the standard. More research is needed not only on the processes and practice of schools but also on how the achievement levels have changed over time. Moreover, the dataset can also be used as a basis for generating knowledge on the educative process of the best and poor performing schools for comparison.

References

Adams, R. & Wu, M. (2002). *PISA 2000 Technical Report.* Paris: OECD Publishing.

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

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. (1995). *Basic and Primary Education Project Achievements Study*. Kathmandu: Author .

----- (1997). *The Effect of New Curriculum on the Achievement of Grade 4 Student*. Kathmandu: Author.

----- (1998).*The Effect of New Curriculum on the Achievement of Grade 5 Student*. Kathmandu: Author.

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.

Central Bureau of Statistics (CBS (2012). *National Population and Housing Census 2011.* Kathmandu: Author.

----- (2013).*National Population and Housing Census 2011. (National Report)*. Kathmandu: Author.

CERID. (1993). *Performance Levels of Public and Private Primary Schools. A Comparative Study.* Kathmandu:Author.

----- (1998). *Performance Levels of Grade V Students.* Kathmandu:Author.

----- (1999). *Assessment of Learning Achievement of Lower Secondary Children: Grades 6 and 8.* Kathmandu: CERID and MOES.

CERSOD (2001). *District Level Assessment, 2001.* Sanothimi: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.

EDSC. (1997). *National Assessment of Grade 3 Students.* Kathmandu:BPEP.

----- (1999). *National Assessment of Grade 5 Students.* Kathmandu: BPEP.

----- (2001). *National Assessment of Grade 3 Students.* Sanothimi: DOE.

----- (2003). *National Assessment of Grade 5 Students.*Sanothimi: DOE.

----- (2008). *National Assessment of Grade 8 Students.* Sanothimi*:DOE.*

Education Review Office(ERO). (2013). *Where are We Now? Student Achievement in Mathematics, Nepali and Social Studies in 2011.*Kathmandu: Author

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.

Fink, A. (1995). *The survey handbook*. Thousand Oaks, CA: Sage Publications.

Finn, A.S.; Kraft, M.A.; West, M.R.; Leonard, J.A.; Bish, C.E.; Martin, R.E.; Sheridan, M.A.; Gabrieli, C.F.O & Gabrieli, J.D.E. (2014). Cognitive Skills, student achievement tests, and schools. *Psychological Science, 25(3):*736-744.

The Finnish National Board of Education (FNBE). (2004). *National Core Curriculum for Basic Education 2004*. Helsinki: Yliopistopaino.

Foy, P. & Kennedy, A.M. (2006).*Sample Passages, Questions, and Scoring Guides. TIMSS 2006 User Guide for the International Database.* Chestnut Hill,MA*:* TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

Full Bright Consultancy. (2008). *A Final Report on National Assessment of Grade-V Students.* Sanothimi*:* DOE.

Garandeau, C. F., Poskiparta, E., & Salmivalli, C. (2014). Tackling acute cases of school bullying in the KiVa antibullying program: Comparison of two approaches. *Journal Abnormal Child Psychology*, 2014 March 12 [Epub ahead of printing] at http://www.ncbi.nlm.nih.gov/ pubmed/24615005. PMID: 24615005.

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., & Black. W. C. (1998). *Multivariate Data Analysis,* (5thEd.). 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.

Invernizzi, M. A., Landrum, T. J., Howell, J. L. & Warley, H. P. (2005). Toward the peaceful coexistence of test developers, policymakers, and teachers in an era of accountability. *The Reading Teacher, 58(7)*: 610–618.

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.

----- (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.

----- (2009). Standard setting reference supplement. In S. Takala (Ed), *Relating Language Examinations to the Common European Framework of Reference for Languages(CEFR):Learning, Teaching, Assessment. A Manual.* Strasbourg: Language Policy Division.

Kärnä, A., Voeten, M., Little, T., Alanen, E., Poskiparta, E., & Salmivalli, C. (2011). Going to scale: The effectiveness of KiVa antibullying program during the first year of nationwide implementation. *Journal of Consulting and Clinical Psychology.* doi: 10.1037/a0025740.

Kärnä, A., Voeten, M., Little, T., Poskiparta, E., Kaljonen, A., & Salmivalli, C. (2011). A large-scale evaluation of the KiVa anti-bullying program; Grades 4–6. *Child Development*, *82*: 311–330.doi: 10.1111/j.1467-8624.2010.01557.x.

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. *Annual Review of Public Health, 18*: 341–78.

Liberatos, P., Link, B. G., Kelsey, J. I. (1988). The measurement of social class in epidemiology. *Epidemiology Review,10*: 87–121.

Lord, F.M. (1980). The relationship of the reliability of multiple-choice test to the distribution of item difficulties. *Psychometrika,18*: 181–194.

Martin, M. O., Mullis, I.V.S, & Kennedy, A.M. (2006). *Pirls 2006 Technical Report.*Chestnut Hill,MA*:* TIMSS& PIRLS International Study Center: Lynch School of Education, Boston College.

Matsuura, K. (2004).Foreword. In *UNESCO: EFA Global Monitoring Report 2005. Education for All. The Quality Imperative.* Paris:UNESCO.

McLoyd, V. C. (1997). The impact of poverty and low socioeconomic status on the socio-emotional functioning of African-American children and adolescents: mediating effects. In R. D. Taylor& M. Wang (Eds.), *Social and Emotional Adjustment and Family Relations in Ethnic Minority Families, 7-34*. Mahwah, NJ: Erlbaum.

Mehrens, W. A. & Lehmann, I. J. (1991). *Measurement and Evaluation in Education and Psychology.* (4thEd). Fort Worth: Brace College Publishers.

Metfesser, 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. (2009). *Methods Assisting the Assessment*. Oppimistulosten arviointi 1/2009. The Finnish National Board of Education. Helsinki: Yliopistopaino. [In Finnish].

----- (2012a). Comparison of mental structures of eighth-graders in different countries on the basis of Fennema-Sherman test. *International Journal of Psychological Studies,4(4).*

----- (2012b). Challenges of the Fennema-Sherman test in the international comparisons. *International Journal of Psychological Studies, 4(3).*

----- (2013a). *Handbook of Research Methods in Human Sciences.* International Methelp Oy. Available as E-book.

----- (2013b). A new method to setting standard for the wide range of language proficiency levels. *International Education Research* 1(1).Retrieved from http://www.todayscience.org/IER/v1/Jari%20Metsamuuronen.pdf.

Metsämuuronen, J., Kuosa, T. & Laukkanen, R. (2013). Sustainable leadership and future-oriented decision making in the educational governance–A Finnish case. *International Journal of Educational Management,27(4)*. http://www.emeraldinsight.com/journals.htm? articleid=17083274.

Mitzel, H. C., Lewis, D. M., Patz, R. J., & Green, D. R. (2001). The bookmark procedure: Psychological perspectives. In G. J. Cizeck (Ed.), *Setting Performance Standards: Concepts, Methods, and Perspectives, 249–281*. Mahwah, NJ: Lawrence Erlbaum.

Ministry of Education (MOE). (1971). *Education Act 2071.* Kathmandu: Author.

----- (2008). *School Sector Reform: Core Document.* Kathmandu: Author.

Mullis, I. V. S., Martin, M. O., & Foy, P. (with Olson, J. F., Preuschoff, C., Erberber, E., Arora, A. & Galia, J.) (2008). *TIMSS 2007 International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at fourth and Eighth Grade.* Chestnut Hill, MA: TIMSS& PIRLS International Study Center, Boston College.Retrieved from http://pirls.bc.edu/timss2007/mathreport.html.

Nunnally, J. C. (1978). *Psychometric theory (2nd ed.)*. New York: McGraw-Hill.

Oaks M. (2011). Measuring Socio-economic status. In e-Source. *Behavioral & Social Science Research*. Retrieved from http://www.esourceresearch.org/Portals/0/Uploads/Documents/Public/Oakes_FullChapter.pdf.

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.* Kathmandu. Author.

Phyak. P. (2012). *Language Issues in Educational Policies and Practices in Nepal: A Critical Review.* Kathmandu: Asian Development Bank.

PISA. (2003a). *PISA Student Questionnaire.* Retrieved from http://pisa2003.acer.edu.au/downloads.php.

----- (2003b). *PISA School Questionnaire.* Retrieved from http://pisa2003.acer.edu.au/downloads.php.

----- (2006a). *PISA student Questionnaire--Reading.* Retrieved from http://pisa2006.acer.edu.au/downloads.php.

----- (2006b). *PISA School Questionnaire.* Retrieved from http://pisa2006.acer.edu.au/downloads.php.

----- (2006c). *PISA Released Items-Reading* . OECD Publishing.

----- (2009). *Take the Test. Sample Questions from OECD's PISA Assessments.* OECD Publishing.

Race, P., Brown, S. & Smith, B. (2005). *500 Tips on Assessment*, *2nd Edition*. London: Routledge Falmer.

Salkind, N. J. (1997). *Exploring Research* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Salmivalli, C., Garandeau, C., & Veenstra, R. (2012). KiVa anti-bullying program: implications for school Adjustment. In G. Ladd & A. Ryan (Eds.), *Peer Relationships and Adjustment at School*. Charlotte, NC: Information Age Publishing.

Salmivalli, C., Kärnä, A., & Poskiparta, E. (2011). Counteracting bullying in Finland: The KiVa program and its effects on different forms of being bullied. *International Journal of Behavioral Development*, *35*: 405–411.

Schleicher A. (2006). *The Economics of Knowledge: Why Education is Key for Europe's Success. Policy Brief.* Brussels: Lisbon Council. Retrieved from http://www.oecd.org/dataoecd/43/11/ 36278531.pdf

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.

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.

Stevenson, H. W. (1998). A study of three cultures: Germany, Japan and the United States - an overview of the TIMSS case study project. *Phi Delta Kappan,79(7)*:524-529.

Takala, S. (Ed.) (2009). *Relating Language Examinations to the Common European Frame work of Reference for Languages(CEFR): Learning, Teaching, Assessment.A Manual.* Strasbourg: Language Policy Division.

TIMSS. (2003). *TIMSS Contextual Background Questionnaires.* Retrieved from: http://timss.bc.edu/timss2003i/context.html.

----- (2006). *Questionnaires.* Retrieved from: http://nces.ed.gov/timss/ questionnaire.asp.

----- (2007). *TIMSS 2003 Science Items. Released Set, Eight grade.* Chestnut Hill,MA: IEA, TIMSS & Pirls International Study Center, Lynch School of Education, Boston College.

----- (2009a). *TIMSS 2007 User Guide for the international Database. Released Items, Mathematics – Eight grade.* Chestnut Hill, MA: IEA, TIMSS & Pirls International Study Center,Lynch School of Education, Boston College.

----- (2009b). *TIMSS 2007 User Guide for the International Database. Released Items, Science – Eight grade.*Chestnut Hill, MA: IEA, TIMSS & Pirls International Study Center, Lynch School of Education, Boston College.

UNESCO. (2005). *EFA Global Monitoring Report 2006. Education for All. Literacy for Life*. Paris: Author.

-----(2009). *EFA Global Monitoring Report 2009.Overcoming Inequality:Why Governance Matters*.Paris: Oxford University Press & UNESCO Publishing.

Van der Schoot, F. (2009). Cito variation of the bookmark method. In S. Takala (ed.), *Relating Language Examinations to the Common European Framework of Reference for Languages(CEFR): Learning, Teaching, Assessment . A Manual.* Strasbourg: Language Policy Division.

Verhelst, N.D., Glas, C.A.W., Verstralen, H.H.F.M. (1995). *One-Parameter Logistic Model (OPLM)*. Arnhem: Cito.

White, K. R. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletein, 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.

Williford, A., Noland, B., Little, T., Kärnä, A., & Salmivalli, C. (2011). Effects of the KiVa anti-bullying program on adolescents' perception of peers, depression, and anxiety. *Journal of Abnormal Psychology*. doi: 10.1007/s 10802-011-9551-1.

Wright, B. D. (1968). Sample-free test calibration and person measurement. *Proceedings of the 1967 Invitational Conference on testing Problems.* Princeton NJ: Educational Testing Service.